

## **Rakkaalla lapsella on monta nimeä**

Englannin *manage*-verbin suomen- ja saksankieliset vastineet Euroopan komission teksteissä

Katariina Franssila  
Tampereen yliopisto  
Kieli- ja käännöstieteiden laitos  
Käännöstiede (saksa)  
Pro gradu -tutkielma  
Kesäkuu 2010

Tampereen yliopisto  
Kieli- ja käännöstieteiden laitos  
Käännöstiede (saksa)

FRANSSILA, KATARIINA: Rakkaalla lapsella on monta nimeä – Englannin *manage*-verbin suomen- ja saksankieliset vastineet Euroopan komission teksteissä

Pro gradu -tutkielma 63 sivua + liitteet 9 sivua + saksankielinen lyhennelmä 13 sivua  
Kesäkuu 2010

---

Tämän tutkielman tarkoituksena on selvittää, mitä suomen- ja saksankielisiä vastineita englannin *manage*-verbi saa Euroopan komission teksteissä. Lisäksi tavoitteena on selvittää *manage*-verbin kollokaattien merkityksen mahdollinen yhteys siihen, mitä suomen- ja saksankielistä vastinetta teksteissä on käytetty.

Aineistona on kaksi sähköistä tutkimuskorpusta, jotka on muodostettu Euroopan komission käännöstoimen pääosaston vuonna 2007 julkaisemasta säädöstekstejä sisältävästä käännösmuistista. Toinen tutkimuskorpus sisältää käännössegmenttejä englanniksi ja suomeksi, toinen puolestaan englanniksi ja saksaksi. Tutkielmassa hyödynnetään korpustyökaluja ja kontekstualisoivan semantiikan piiriin kuuluvaa semanttisen preferenssin käsitettä. *Manage*-verbin vastineet selvitettiin hyödyntämällä kaksikielistä konkordanssitoimintoa, jonka avulla tutkimuskorpuksista etsittiin sellaiset käännösyksiköt, joiden englanninkielisessä segmentissä *manage*-verbi esiintyy. Tämän jälkeen suomen- ja saksankielisistä segmenteistä poimittiin kyseisen verbin vastineet ja kaikki vastineet muutettiin verbimuotoon vertailun helpottamiseksi. Kollokaattien merkityksen vaikutusta tutkittiin valitsemalla tarkasteltaviksi kollokaateiksi *manage*-verbin saamat objektit. Nämä objektit luokiteltiin semanttisen preferenssin tutkimista varten korpusvetoisesti luotuihin merkitysryhmiin.

*Manage*-verbi sai sekä suomeksi että saksaksi runsaasti erilaisia vastineita, jotka jakautuivat kuitenkin hyvin eri tavalla. Saksankielisistä vastineista yleisin oli *verwalten*, joka muodosti yli 40 prosenttia kaikista esiintymistä. Muut saksankieliset vastineet muodostivat hyvin pirstaloituneen joukon. Suomenkielisten vastineiden osalta merkittävimiksi nousi neljä verbiä: *hallinnoida*, *hoitaa*, *hallita* ja *johtaa*, jotka yhdessä muodostivat yli 80 prosenttia kaikista suomenkielisistä vastineista. Luokittelemalla englanninkieliset objektit merkitysryhmiin päästiin kiertotietä käsiksi myös suomen- ja saksankielisten vastineiden yhteydessä esiintyvien sanojen merkitysryhmiin. Näitä tarkastelemalla pystyttiin erottamaan edellä mainituille aineistojen yleisimmille verbeille suuntaa-antavia semanttisia preferenssejä, joiden avulla voitiin vertailla vastineiden mahdollisia käyttöyhteyksien eroja. Saksan verbiä *verwalten* esiintyi monipuolisesti eri merkitysryhmien kanssa, kun taas suomen yleisimmät verbivastineet suosivat hiukan erikoistuneemmin tiettyjä merkitysryhmiä, etenkin verbien *hallita* ja *johtaa* osalta. Tuloksista voidaan suuntaa antavasti päätellä, että *manage*-verbiä suomennettaessa on suoritettava leksikaalinen valinta usean vaihtoehdon välillä, kun taas saksaan päin käännettäessä tilanne on hieman yksinkertaisempi.

Avainsanat: Euroopan komissio, korpustutkimus, kollokaatio

# Sisällysluettelo

1. Johdanto.....	1
2. Tutkimusongelma ja sen ratkaiseminen .....	5
2.1 Polysemia.....	5
2.2 Manage-verbi sanakirjoissa .....	7
2.3 Kontekstualisoiva tutkimusote.....	8
3. Toimintakontekstin kuvaus.....	11
3.1 Euroopan unionin kielipolitiikka.....	11
3.2 Kääntäminen Euroopan komissiossa.....	11
3.2.1 Kääntäjät ja tekstit .....	11
3.2.2 Kääntämisen apuvälineet .....	14
4. Korpustutkimus menetelmänä .....	16
4.1 Katsaus korpustutkimuksen historiaan .....	16
4.1.1 Varhainen korpuslingvistiikka .....	16
4.1.2 Chomskyn kritiikki ja vastauksia siihen .....	17
4.1.3 Moderni korpuslingvistiikka .....	18
4.2 Korpuksen määritelmiä.....	19
4.3 Korpuslingvistinen käänöstutkimus .....	23
4.3.1 Korpuslingvistisen käänöstutkimuksen korpustyytit .....	25
4.3.2 Korpuslingvistinen käänöstutkimus Suomessa .....	28
5. Tutkimuskorpusten kuvaus.....	30
5.1 Lähdeaineiston kuvaus .....	30
5.2 Tutkimuskorpusten kokoaminen .....	31
5.3 Tilastotietoja tutkimuskorpuksesta.....	32
5.4 Aineiston sopivuuden arviointia .....	32
6. <i>Manage</i> -verbi suomeksi ja saksaksi.....	36
6.1 Analyysin valmistelu .....	36
6.2 Analyysi .....	42
6.3 Suomen- ja saksankielisten vastineiden vertailua .....	56
7. Loppupäätelmät.....	59
Lähteet.....	64
Liitteet .....	68
Deutsche Kurzfassung.....	

# 1. Johdanto

Pro gradu -tutkielmani nimi, ”Rakkaalla lapsella on monta nimeä”, on ehkä hieman sarkastinen, sillä viitataan rakkaalla lapsella itselleni käännöstyössä kovastikin päänvaivaa aiheuttaneeseen englanninkieliseen *manage*-verbiin. Toisaalta eniten meille huolta kykenevät aiheuttamaan usein juuri ne, jotka ovat meille kaikista rakkaimpia, joten nimen voi lukea myös tosikkomaisesti. Joka tapauksessa tutkielmani aihe kumpuaa käännöshaasteista, joita kohtasin ollessani viiden kuukauden ajan harjoittelijana Euroopan komission käännöstoimen pääosastossa Brysselissä keväällä 2009. Yksikköni oli FI 02, jonka aihealueisiin kuuluvat tutkimus, talous, luonnonvarat ja tekniikka. Kääntäessäni erilaisia komission tekstejä huomasin nopeasti, että pysähdyin jatkuvasti tiettyjen sanojen kohdalla ihmettelemään, miten ne pitäisi kääntää. Tässä ei sinänsä ole mitään kummallista, olinhan vielä opiskelija ja uudessa työpaikassa. Lisäksi kääntäminen on aina hämmästyttävää ja asioiden selvittelyä. Nämä sanat tuntuivat minusta kuitenkin näennäisen helpoilta eivätkä olleet varsinaisia termejä, joihin pysähtyminen olisi tuntunut luonnollisemmalta. Aloin kerätä listaa tällaisista sanoista, jolloin huomasin, että yhä uudestaan, melkein jokaisen tekstin kohdalla, selvittelin sopivaa käännoästä muiden muassa englanninkieliselle verbille *to manage*. Tuntui, ettei suomeksi löytynyt yhtä napakkaa ja kaikissa tilanteissa ja konteksteissa toimivaa käännoästä tälle lähdeteksteissä tiuhaan käytetylle sanalle. Englanninkielinen sana saattoi tarkoittaa monta, hyvinkin erilaisen merkityksen saavaa asiaa suomeksi, jolloin minun piti kääntäjänä valita yksi monista vaihtoehtoista. Oikean päätöksen tekemiseksi asiaan piti perehtyä joskus syvällisestikin, ja tähän selvitystyöhön kului tietysti aikaa. Ajan tuhraantumista oli joskus vaikea hyväksyä, kun teksteissä oli monimutkaisia teknisiä termejäkin, joiden sisältö oli myös selvitettävänä.

Kun purnasin tätä ongelmaa muista jäsenvaltioista kotoisin oleville kääntäjäharjoittelijoille, sain vaihtelevaa empatiaa osakseni. Esimerkiksi ruotsalaisella kollegallani tuntui olevan samankaltaisia ongelmia: tuntui että englanninkielisessä tekstissä oli pystytty sanomaan asioita epämääräisemmin, kun taas suomeksi ja ruotsiksi täytyi usein tietää tarkemmin, mistä on kysymys, ennen kuin sanan voi kääntää. Saksan yksikössä työskennellyt itävaltalainen kollegani ei niinkään tunnistanut ongelmaani omakseen, vaan kertoi heillä olevan yksiselitteiset vastineet mainitsemilleni sanoille. Mieltäni alkoi askarruttaa kysymys: onko näiden suomalaiselle päänvaivaa aiheuttavien sanojen kääntäminen helpompaa joillekin

englantia lähempänä oleville kielille? Onko tässä nyt kysymys siitä perustavanlaatuisesta kääntämiseen liittyvästä ongelmasta, että eri kielten järjestelmät ovat erilaisia? Kyseessä taitaa olla oppikirjaesimerkki siitä, ettei kullekin vieraan kielen sanalle ole olemassa yhtä ja ainoaa suomenkielistä vastinetta ja toisaalta toisin päin. Tuntui kiehtovalta huomata käytännössä, miten tämä teoriassa tuttu ilmiö tulee esiin, kun samaa tekstiä käännetään useammalle kielelle (komission ollessa kyseessä alkutekstistä muille 22<sup>1</sup> viralliselle kielelle).

Hetken asiaa pohdittuani hyväksyin, ettei ole mahdollista löytää yhtä kaikenkattavaa suomennosta *manage*-verbille. Kerran eräs kokenut komission kääntäjä kuitenkin tarkasti käännettäni ja huomautti puolihuolimattomasti, että eräs toinen verbi sopisi *manage*-verbin vastineeksi paremmin englanninkielisen verbin objektin vuoksi. Tuolloin minua jäi askarruttamaan kontekstin vaikutus siihen, mitä suomenkielistä vastinetta minun tulisi käyttää. Saksan kääntämisen opiskelijana pohdin myös, miten suomi ja saksa eroavat sen suhteen, kuinka ne ilmaisevat *manage*-verbin sisällön. Näistä aineksista alkoi muodostua graduseminaarin edetessä tutkimuskysymykseni, jonka esittelen tarkemmin seuraavassa luvussa. Koska kiinnostukseni aiheeseen oli syntynyt käytännön käänntilanteessa, minulle oli alusta asti selvää, että tulisin tekemään empiiristä tutkimusta, siis tutkimaan tositalanteessa syntynyttä aineistoa. Aineiston valintaan vaikutti suuresti myös työkokemukseni komissiossa: halusin toisaalta tutkia nimenomaan niitä tekstejä, joiden parissa olin kuukausia työskennellyt ja toisaalta minua kiinnosti tutkia englanninkielisen aineiston lisäksi sekä suomea että saksaa. Onneksi kääntämisen pääosasto oli vuonna 2007 julkaissut osan käänntömuististaan internetissä. Tämän käänntömuistin englannin-, saksan- ja suomenkielisistä osioista koottiin kaksi tutkimuskorpusta, joita saatoinkin käsitellä korpustyökaluin.

Tutkielmani etenee siten, että toisessa luvussa muotoilen tutkimuskysymykseni sekä yritän hahmottaa käytännön tilanteessa syntyneitä havaintojani *manage*-verbin monimuotoisesta käytöstä teoreettisin välinein. Pohdin sitä, liittyvätkö verbin monet vastinevaihtoehdot sen monimerkityksisyyteen eli polyseemisyyteen. Mahdollista polyseemisyyttä tarkastelen toisaalta yleisesti semanttisen teorian ongelmana ja toisaalta tietokonekielintieteen näkökulmasta. Lisäksi lähestyn sitä leksikografiselta kannalta tarkastelemalla, onko *manage*-verbi käsitetty sanakirja-artikkeleissa yhden vai useamman merkityksen omaavaksi. Luvun lopussa kerron käyttämästäni kontekstualisoivasta tutkimusotteesta, jota on

---

<sup>1</sup> Euroopan komissio, kääntämisen pääosasto, [http://ec.europa.eu/dgs/translation/translating/index\\_de.htm](http://ec.europa.eu/dgs/translation/translating/index_de.htm)

korpuslingvistiksessä käännöstutkimuksessa hyödyntänyt muun muassa Jarmo Harri Jantunen (2004) tutkiessaan samamerkityksisyyden eli synonymian kontekstuaalisuutta. Tutkimusotteeseen liittyen esittelen semanttisen preferenssin käsitteen. Kolmannessa luvussa esittelen sen toimintakontekstin, jossa tutkimusaineistoni tekstit ovat syntyneet. Euroopan komission käännöstoimen pääosaston olemassaolo perustuu Euroopan unionin kielipolitiikkaan, joka pyrkii takaamaan kaikille unionin kansalaisille tasa-arvoiset mahdollisuudet viestiä tämän ylikansallisen hallintoelimen kanssa. Tästä perusperiaatteesta johtuen minunkin on nyt mahdollista tutkia hyvin pitkälle samojen tekstien englannin-, suomen- ja saksankielisiä versioita. Tutkimusaineistoni luonteen ymmärtämisen kannalta on myös oleellista esitellä hieman sitä käännösorganisaatiota, jossa tekstit ovat syntyneet. Vaikka varsinaisessa analyysiosiossa käsitteinkin tekstejä lähinnä näytteinä niiden edustamista kielistä, kiinnitän tarkemmin huomiota tekstien syntymekanismien ja kirjoittajien vaikutukseen loppupäätelmissä.

Neljännessä luvussa esittelen korpustutkimusta menetelmänä, jotta osaisin arvioida oman aineistoni soveltuvuutta ja luotettavuutta käyttämäni metodin kannalta. Pyrin antamaan kuvan korpusmenetelmien kehityksestä ensin yleisesti (pääasiassa kielitieteessä käytettynä menetelmänä), mutta pureutumaan sen jälkeen nimenomaan korpuslingvistiseen käännöstutkimukseen eli käännöstieteen käsityksiin korpustutkimuksesta. Suomalaisen korpuslingvistisen käännöstutkimuksen yhteydessä kerron tyypillisestä käännöstieteellisestä korpustutkimuksen kohteesta eli käännösuniversaaleista, joiden tutkimiseen oma tutkimukseni ei kuitenkaan liity. Erilaisten korpuksen määritelmien esittelyllä haluan tuoda ilmi niiden moninaisen luonteen ja auttaa oman tutkimusaineistoni määrittelyssä, johon keskityn tutkielmani viidennessä luvussa. Kyseisen luvun alussa kerron käytännönläheisesti, miten käyttämäni korpuksat ovat syntyneet. Lisäksi pyrin kriittisestikin arvioimaan, miten hyvin aineistoni soveltuu tutkimuskysymykseeni vastaamiseen. Kuudes luku käsittelee analyysin eri vaiheita, jotka jaottelen analyysin valmisteluun ja varsinaiseen analyysiin. Valmisteluvaiheessa tein molempiin tutkimuskorpuksiini kaksikielisen konkordanssihaun, jonka tulokset tallensin Excel-tiedostoon, missä niiden käsittely jatkui. Varsinaiseksi analyysiksi käsitän sen vaiheen, jossa tulkitseen aineistoista saamiani, laskennallisessa muodossa olevia tuloksia, jotka koskevat *manage*-verbin vastineiden määrää ja muotoa sekä niiden semanttisia preferenssejä. Luvun lopuksi vertailen yleisimpien suomen- ja saksankielisten vastineiden semanttisia ympäristöjä keskenään ja pohdin erojen vaikutuksia suomen ja saksan välillä kääntämiseen. Loppupäätelmissä vedän tutkimukseni tuloksia

yhteen ja esitän tutkimuksen edessä saamiani ideoita mahdollisiksi jatkotutkimuksen aiheiksi. Lisäksi pohdin aineiston syntyyn vaikuttaneen toimintakontekstin mahdollisia vaikutuksia tutkimukseni tuloksiin.

Tutkimusprosessin aikana olen vähitellen löytänyt yhteyden omien prosessin alussa vielä hyvin jäsentymättömien ajatusteni ja käännöstieteellisen tutkimuksen mahdollisuuksien välille. Kielen ja kääntämisen monimutkaisuutta kuvaa hyvin se, että tutkielmani teon aikana olen pyytänyt mielipiteitä ja neuvoja monien kielen eri aspektien kanssa työskenteleviltä ammattilaisilta, joihin kuuluvat muassa komission suomen kielen koordinaattori Erkki Vuorinen ja korpuslingvistiseen käännöstutkimukseen perehtynyt yliopiston lehtori Mikhail Mikhailov. Korpustutkimuksesta olen saanut välineet aineistoni käsittelyyn ja arviointiin, kun taas semantiikan ja erityisesti kontekstuaalisen semantiikan puolelta löysin omiin käsityksiini sopivia teoreettisia käsitteitä, joiden avulla sain aineistostani irti vastineiden semanttiseen preferenssiin liittyviä tuloksia.

Koko tutkielman taustalla piilee ajatus kielten perustavanlaatuisista eroista, mutta samalla tarpeesta käsitellä näitä eroavaisuuksia niin, että viestintä voisi olla onnistunutta myös kielten ja sitä myöden kulttuurien välillä. Uskon tämän tutkimuksen syventäneen omaa ammatillista ymmärtämystäni kääntämisestä ja paljastaneen ainakin pienessä mittakaavassa jotakin olennaista kääntämisen perusolemuksesta. Luulen voivani hyödyntää tutkimukseni tuloksia ja niistä tekemiäni päätelmiä työssäni kääntäjänä.

## 2. Tutkimusongelma ja sen ratkaiseminen

Alustavista käännoistyössä syntyneistä havainnoista muotoutui vähitellen kaksi varsinaista tutkimuskysymystä. Ensinnäkin halusin selvittää, mitkä ovat ne tavat, joilla englannin *manage* ilmaistaan Euroopan komission teksteissä suomeksi ja saksaksi. Halusin myös tietää, vaikuttavatko englanninkielisen verbin saamat kollokaatit siihen, mikä vastine esiintyy suomen- tai saksankielisessä tekstissä. Tässä luvussa yritän jäsentää *manage*-verbin olemusta semantiikan käsittein ja toisaalta pohtia tuon olemuksen mahdollista vaikutusta siihen, että käytännön kokemuksen mukaan se voi saada runsaasti vastineita ainakin suomeksi. Lisäksi esittelen kontekstualisoivan tutkimusotteen, jota hyödynnän korpusmuodossa olevan aineistoni analyysissä sekä esittelen semanttisen preferenssin käsitteen, jonka avulla pyrin hahmottamaan kollokaattien vaikutusta verbin saamiin vastineisiin.

### 2.1 Polysemia

Mistä oikeastaan johtuu, että *manage*-verbille tuntuu olevan monta erilaista suomennosta ja saksannosta? Selitys liittyy ainakin kielten ja kääntämisen perusolemuksen eli siihen, että eri kielissä käsitteet ovat muodostuneet ja jakautuneet ainakin osittain oman logiikkansa mukaisesti. Näin ollen ei ole mitenkään poikkeuksellista, ettei täysin vastaavaa käsitettä löydy toisesta kielestä. En kuitenkaan tyydy tähän yleiseen toteamukseen vaan yritän seuraavassa hahmottaa tätä saksan- ja suomenkielisten vastineiden kirjoa tarkastelemalla *manage*-verbiä semantiikan eli merkitysoopin tarjoamin käsittein. Erityisesti kiinnitän huomiota *manage*-verbin mahdolliseen polyseemisyteen eli monimerkityksisyyteen ja pohdin, selittykö vastineiden moninaisuus ainakin osittain sillä.

Polysemia on semantiikan käsite, jolla siis viitataan sanan monimerkityksisyyteen. Polysemialle läheinen käsite homonymiakin viittaa ilmiöön, jossa samalla sanalla on monta eri merkitystä. Polysemia ja homonymia erotetaan kuitenkin toisistaan, ja se tapahtuu etymologisista perusteista: homonyymisilla sanoilla on vain sattumalta samanlainen kirjoitusasu (homografia), mutta ne eivät muuten liity toisiinsa, kun taas polyseemisillä sanoilla on etymologinen ja näin ollen semanttinen yhteys (Ravin & Leacock 2000, 2). Monet polyseemit ovat syntyneet sanojen metaforisesta käytöstä (emt. 2), kuten *lehti* merkityksissä 'puun lehti' ja 'sanomalehti'. Toinen polysemialle läheinen käsite on monitulkintaisuus, jolla viitataan kielen luonnolliseen joustavuuteen eli siihen, että sanat saavat tarkan merkityksensä vasta kontekstissa. Esimerkiksi sanan *lapsi* tarkoite voi sukupuoleltaan olla joko nais- tai



miespuolinen, jolloin sanan voi käsittää monta merkitystä omaavaksi eli polyseemiseksi tai sillä voi käsittää olevan yksi merkitys, jonka tietty merkityspiirre on määrittelemätön (*indeterminate*) (emt. 3). Polysemian ja monitulkintaisuuden erottaminen on semanttisen teorian ydinkysymyksiä, sillä sen myötä määritellään kielellisten ilmausten ja niiden kielenulkoisten tarkoitteiden suhde (emt. 3). Niiden erottamiseksi on kehitelty joukko erilaisia testejä, jotka voidaan luokitella loogisiin, kielitieteellisiin ja määrittelyyn perustuviin testeihin (emt. 3). Näiden testien tulokset eivät kuitenkaan ole johdonmukaisia tai yhtenäisiä, joten polysemian ja monitulkintaisuuden raja on häilyvä (emt. 4).

Yleisimmät sanat ovat usein kaikkein polyseemisimpiä, mutta polysemia ei Ravinin ja Leacockin mukaan silti aiheuta ongelmia ihmisten välisessä viestinnässä, sillä polyseemisten sanojen oikea merkitys osataan valikoida viestintätilanteessa vaivatta ja tiedostamatta hyödyntämällä kontekstin tarjoamia vihjeitä. He toteavat polysemian kuitenkin olevan ongelma semanttisessa teoriassa ja semanttisissa sovelluksissa, kuten kääntämisessä ja leksikografiassa. (Ravin & Leacock 2000, 1.) Leksikografit eivät ole yhtä mieltä sanojen merkitysten määrästä, näiden merkitysten semanttisesta sisällöstä ja niiden ryhmittelystä, ja semantikoillakin on näistä asioista eriäviä mielipiteitä (emt. 2). Oman tutkimukseni kannalta on tärkeää pohtia, johtuuko *manage*-verbin kääntämisen haasteellisuus nimenomaan polysemiasta. Semantikot ja leksikografit keskittyvät pohdinnoissaan hieman eri näkökulmiin kuin kääntäjä.

Tilannetta voi hahmottaa kääntäjän näkökulmasta pohtimalla konekäännösohjelmienkin kannalta olennaista polysemian laskennallista määritelmää: kyseessä on monimerkityksinen eli polyseeminen sana, jos kohdekieleen käännettäessä on tehtävä leksikaalinen valinta eli jos yksi ja sama sana voidaan kääntää monella eri sanalla (vrt. *manage*: mm. *hallita*, *johtaa*, *hallinnoida* (fi) ja *verwalten*, *führen*, *regeln* (de)). Konekääntämisen ongelmat ovat silti hyvin erityyppisiä kuin ammattilaisten käännoistyössä kohtaamat ongelmat, sillä ihmisille monet homonyymit (*bank* 'rantatörmänä' ja 'pankkina') ja syntaktinen monitulkintaisuus (*table* verbinä ja substantiivina) on helppo erotella, kun taas tietokoneohjelmille näidenkin erottelu voi olla haastavaa. (Ravin & Leacock 2000, 24.) Laskennallisen määritelmän mukaan *manage*-verbiä voidaan siis joka tapauksessa pitää polyseemisenä. Myös sanakirjavertailun tulosten pohjalta vaikuttaisi siltä, että leksikografiassa *manage*-verbi käsitetään polyseemiseksi, joskin myös leksikografien erimielisyydet merkitysten määrästä

käyvät vertailusta hyvin ilmi. Lisää vertailun tuloksista alaluvussa ”Manage-verbi sanakirjoissa”.

Korpuspohjaisia merkityksen erottelumenetelmiä on kehitetty sähköisten korpusten yleistyessä ja niissä on usein hyödynnetty kontekstia tai tarkemmin sanottuna polyseemisten sanojen kanssa esiintyviä toisia sanoja (Ravin & Leacock 2000, 25). Merkityksen tunnistamiseen tietokoneella sisältyy kuitenkin monenlaisia ongelmia, ja on muun muassa huomattu, että tietokoneen avulla homonyymien merkityksen tunnistaminen on polyseemien merkityksen tunnistamista helpompaa. Tämä on ymmärrettävää, sillä usein homonyymien konteksti koostuu hyvin erilaisista sanoista, kun taas polyseemien konteksti voi olla hyvinkin samankaltainen (vrt. käsillä olevan tutkimuksen aineiston muodostavat säädöstekstit). (emt. 26.)

Kuten jo aiemmin totesin, polysemia ei ole haaste ainoastaan konekäännösohjelmeille. Ilmauksen monimerkityksisyys on aina haaste, jos yhdelle lähdekielen sanalle on olemassa useampi kohdekielinen vastine. Kääntäjän on tällaisessa tilanteessa pääteltävä sanan merkitys kyseisessä kontekstissa ja eksplikoitava tulkintansa kohdekielisen sananvalinnan muodossa. Omana tutkimuksellisenä tavoitteenani ei olekaan palvella konekäännösohjelmiä kehittämällä algoritmeja, joiden avulla ohjelma voisi tunnistaa kontekstin perusteella polyseemisen sanan merkityksen laskennallisesti (vrt. Miller & Leacock 2000, 156), vaan saada näkyviin polyseemisen verbin merkitykseen liittyvän kontekstin mahdollisesti tarjoamia vihjeitä, joiden tiedostaminen voisi nopeuttaa ammattilaiskääntäjän työtä.

## **2.2 Manage-verbi sanakirjoissa**

Tietokonelingvistiikan laskennallisen määritelmän ohella *manage*-verbin käsittäminen polyseemiseksi sai tukea sanakirjaverailusta. Eri sanakirjojen artikkelit kuitenkin ilmensivät hyvin sitä, ettei leksikografeilla ole yhteneväistä käsitystä *manage*-verbin alamerkitysten määrästä. Yhdessä englanti–suomi-sanakirjan artikkelissa (*MOT Englanti 4.8*) *manage* ilmoitettiin sekä transitiivi- että intransitiiviverbiksi, minkä jälkeen sille annettiin viisi eri merkitystä. Käsillä olevan tutkimuksen kannalta relevantteja merkityksiä oli kuitenkin osittain niputettu yhteen (esimerkiksi *johtaa* ja *hoitaa* samana alamerkityksenä), eikä esimerkiksi yhtä tutkimusaineistoni yleisimmistä *manage*-verbin suomenkielisistä käännösvastaineista (*hallinnoida*) mainittu artikkelissa lainkaan. Toisessa tarkastelemassani sanakirjassa (*Suursanakirja englanti–suomi* 2003) *manage*-verbin transitiivimuodolle

annettiin kolme merkitystä, joista merkitykset 1 ja 3 vielä jakautuivat tarkempiin alamerkityksiin, ja intransitiivimuodolle yksi merkitys, jota tarkennettiin kahdella alamerkityksellä. Tässäkin sanakirjassa esimerkiksi *hoitaa* ja *johtaa* oli merkitty kuuluvaksi samaan merkitykseen.

Englanti–saksa-sanakirjassa (*The Concise Oxford German Dictionary* 2004) *manage*-verbin alamerkitykset jaoteltiin numeroin transitiivisen (1.) ja intransitiivisen (2.) verbin merkityksiin, joista transitiiviselle verbille annettiin kuusi alamerkitystä ja intransitiiviselle yksi alamerkitys. Tässäkin sanakirjassa osa käsillä olevan tutkimuksen kannalta oleellisista merkityksistä oli koottu yhden alamerkityksen alle (esim. *verwalten*, *leiten* ja *führen*). Englanti–saksa-taskusanakirjassa (*Langenscheidt Taschenwörterbuch Englisch–Deutsch* 2003) transitiiviverbi *manage* sai viisi alamerkitystä ja intransitiiviverbi kaksi. Kyseisessä artikkelissa esimerkiksi *leiten* ja *führen* oli annettu vastineiksi yhdestä ja samasta merkityksestä. Verbiä *verwalten* ei tässä sanakirjassa annettu verbin *manage* vastineeksi lainkaan, vaikka se aineistossani oli hyvinkin korostetusti esillä.

Englanninkielisistä sanakirjoista mainittakoon ääriesimerkkinä *Longman dictionary of contemporary English* (2007), joka luettelee *manage*-verbille peräti 13 merkitystä. Maltillisemmin verbin monimerkityksisyyttä käsittelee puolestaan *The Penguin English Dictionary* (2007), jossa verbille annetaan transitiivimuotoisena seitsemän alamerkitystä ja intransitiivimuotoisena yksi alamerkitys. Sanakirjatarkastelun perusteella voidaan tiivistetysti todeta, että *manage*-verbillä katsotaan yleisesti olevan enemmän kuin yksi merkitys, mutta alamerkitysten määrä vaihtelee. Lisäksi on merkille pantavaa, etteivät merkitysjaottelut vastaa tutkimuksessa esiin tulleita suomen- eivätkä saksankielisiä käännösvastineita. Käännettäessä on usein suoritettava leksikaalinen valinta, vaikka eri suomennokset tai saksannokset olisi sanakirjoissa katsottukin kuuluvaksi samaan alamerkitykseen. Näissä tapauksissa kääntäjä joutunee turvautumaan kontekstiin ja pääättelemään sopivan käännökseen siitä.

### **2.3 Kontekstualisoiva tutkimusote**

Yksi käännöstieteellisen korpustutkimuksen keskeisistä suomalaistutkijoista, Jarmo Harri Jantunen (2004), on väitöskirjassaan tarkastellut synonymian eli samamerkityksisyyden kontekstuaalisuutta ja käännöskielen leksikaalisia erityispiirteitä korpusaineistoa hyödyntäen. Hänen tutkimusotteensa on kontekstualisoiva eli sekä käännöskieltä että synonymiaa

tarkastellaan ”niitä ympäröivän ja niiden kanssa vuorovaikutuksessa olevan ympäristön avulla”. Synonymiaa Jantunen tutkii erityisesti lähikontekstin avulla katsomalla, miten ilmaukset erottuvat tekstiympäristönsä puolesta toisistaan. (Jantunen 2004, 4–5.) Hänen tutkimusotteensa kuvaa hyvin myös käsillä olevan tutkimuksen taustalla piilevää ajatusta tarkastella *manage*-verbin merkityksiä niiden lähikontekstissa, joka tässä tapauksessa muodostuu käänössegmentin sisältämästä tekstiaineksesta. Jantunen tarkastelee synonymiaa monelta kantilta ja pyrkii lopuksi kokoamaan saaduista tuloksista kolmen synonymimisena pidetyn sanan kontekstuaaliset profiilit. Näiden profiilien rakennusaineksina toimivat tulokset hän sai tutkittuaan sanojen kollokationaalista ja kolligationaalista valintaa, morfologista preferenssiä ja semanttista valintaa. Kollokationaalisella valinnalla tarkoitetaan hakusanan (*node*) taipumusta saada tiettyjä sanoja kollokaateikseen (emt. 15–21). Kolligationaalisella valinnalla viitataan puolestaan kieliopillisiin luokkiin perustuvaan valintaan (emt. 21–24) ja morfologisella preferenssillä lekseemin yksittäisten taivutusmuotojen (sananmuotojen) toisista saman lekseemin taivutusmuodoista mahdollisesti poikkeavaa kontekstuaalista valintaa (emt. 27). Omaan tutkimuskysymykseeni vastaamisessa näistä mielenkiintoisin on kuitenkin semanttiseen valintaan liittyvä semanttisen preferenssin käsite.

Semanttisen valinnan tutkimisella tarkoitetaan sitä, että tutkitaan, onko lekseemillä preferenssiä tietynlaisiin semanttisiin piirteisiin eli taipumusta saada kollokaateikseen tietynmerkityksisiä sanoja. Semanttisen valinnan voi jakaa semanttiseen prosodiaan ja semanttiseen preferenssiin, joista ensiksi mainittu on useammin käsitelty. Semanttisella prosodialla tarkoitetaan ilmausta ympäröivää, sen kontekstista ilmitulevaa ”merkitysauraa”, joka heijastaa kielenkäyttäjän asenteita, ja se jaotellaan usein luokkiin positiivinen–negatiivinen. John Sinclair on tutkinut ilmiötä hyödyntämällä kontekstuaalisen valinnan ajatusta ja esittänyt esimerkiksi, että englanninkielisen verbirakenteen *set in* kontekstin subjektit viittaavat tyypillisesti epämiellyttäviin asioihin. (Jantunen 2004, 25–26.) Semanttisen preferenssin on esitellyt ensimmäisen kerran samainen Sinclair, ja sillä tarkoitetaan tietyn lekseemin ja sen kollokaattien merkitysisältöjen suhdetta. Tämä vaikuttaa ilmiönä hyvin samantyyppiseltä kuin semanttinen prosodia, ja semanttiseksi preferenssiksi luettavaa valintaa ei usein olekaan erotettu semanttisesta prosodiasta. (Jantunen 2004, 24.) Semanttista preferenssiä tutkittaessa pyritään käsitykseni mukaan löytämään hieman hienojakoisempia tietyn sanan kanssa esiintyviä merkitysjoukkoja kuin semanttista prosodiaa tutkittaessa. Semanttista preferenssiä tutkitaan nimittäin luokittelemalla tietyn lekseemin

kollokaatit merkitysryhmiin, ja nämä merkitysjaottelut tehdään analyysissa yleensä sanaluokkajaot ylittävästi (emt. 133).

Jantunen huomauttaa, että semanttisen preferenssin analyysissa on tavoitteena löytää melko hienojakoisia merkityspiirteitä, mutta samalla on järkevää pyrkiä tarpeeksi pieneen määrään ryhmiä, jotta analyysi olisi hallittavissa. Hän toteaa luokittelun jäävän näin ollen pakostakin läheisten merkitysisältöjen yhteenniputtamiseksi. (emt. 135.) Lisäksi Jantunen huomauttaa, että kollokaattien merkitysisällöistä ilmeneviä semanttisia sisältöjä määrittäessä tutkijan intuitiolla on suuri vaikutus (emt. 24). Näistä varauksista huolimatta katson semanttisen preferenssin tutkimisen olevan sopiva tapa jäsentää englanninkielisen *manage*-verbin saamien kollokaattien mahdollista yhteyttä suomen- ja saksankielisten vastineiden valintaan. Aion keskittyä tietynlaajuisen kontekstin kaikkien lekseemien analysoinnin sijaan ainoastaan *manage*-verbillä ilmaistun tekemisen kohteiden (useimmiten objekti, tarkemmin tästä analyysiluvussa) semanttiseen luokitteluun ja katsoa, onko niiden merkitys jossakin havaittavassa suhteessa suomen- ja saksankielisiin vastineisiin. Tarkoituksena on siis tällä tavoin tutkia, vaikuttaako *manage*-verbin objekti siihen, miten se suomennetaan tai saksannetaan. Semanttista preferenssiä aiemmin hyödyntäneet tutkijat ovat luoneet monenlaisia merkitysluokkia (esim. Dixon on luokitellut adjektiiveja ulottuvuutta, ikää, väriä, arvoa, ja vauhtia kuvaaviin merkitysryhmiin, koottu esitys: Jantunen 2004, 136), joista en kuitenkaan pidä yhtäkään ryhmittelyä omaan tutkimukseeni soveltavana. Näin ollen muodostan analyysissa käytettävät luokat aineistoni pohjalta eli korpusvetoisesti. Aineistoa tarkastelemalla syntyneitä merkitysryhmiä ja niitä perusteita, joiden mukaan englanninkieliset objektit on niihin luokiteltu, kuvaan tarkemmin luvussa ”*Manage*-verbi suomeksi ja saksaksi”. Seuraavaksi siirryn käsittelemään sitä toimintakontekstia, jonka puitteissa tutkimukseni aineisto eli Euroopan komission teksteihin lukeutuvat säädöstekstit ovat syntyneet.

## **3. Toimintakontekstin kuvaus**

### **3.1 Euroopan unionin kielipolitiikka**

Yksi Euroopan unionin monikielisyyspolitiikan päätavoitteista on, että ”kansalaiset voivat perehtyä EU:n lainsäädäntöön, saada EU:lta tietoa ja osallistua EU:n menettelyihin omalla kielellään” (Euroopan komissio 2009a). Tästä peruseriaatteesta on sovittu jo EU:n ensimmäisessä asetuksessa (Asetus N:o 1 Euroopan talousyhteisössä käytettäviä kieliä koskevista järjestelyistä), joka on peräisin vuodelta 1958 ja jossa luetellaan viralliset kielet ja säädetään niiden käytöstä (Euroopan komissio 2009b). Uusien jäsenvaltioiden myötä mukaan on tullut uusia virallisia kieliä, ja nykyään virallisia kieliä on kaiken kaikkiaan 23<sup>2</sup> (Euroopan komissio 2009c). Myös Euroopan unionin toiminnasta tehdyssä sopimuksessa<sup>3</sup> (20 artiklan 2 kohdan d alakohta) todetaan, että EU:n kansalaisilla on oikeus viestiä EU:n toimielinten kanssa valitsemallaan virallisella kielellä ja saada vastaus samalla kielellä.

Vaikka kaikkea EU:n viestintää ei käännetäkään kaikille virallisille kielille, työsarkaa riittää silti. Euroopan komissiolla on oma käännöspalvelunsa, käännöstoimen pääosasto. Neuvostolla, parlamentilla, Euroopan unionin tuomioistuimella, Euroopan talous- ja sosiaalikomitealla ja alueiden komitealla sekä tilintarkastustuomioistuimella on omat käännösosastonsa. (Euroopan komissio 2009d.) Luxemburgissa toimiva Euroopan unionin elinten käännöskeskus tarjoaa puolestaan käännöspalveluja ympäri Eurooppaa sijaitseville EU:n erillisvirastoille. (Euroopan unioni 2009).

### **3.2 Kääntäminen Euroopan komissiossa**

#### **3.2.1 Kääntäjät ja tekstit**

Euroopan komission käännöstoimen pääosasto (Directorate-General for Translation, jatkossa DGT) on yksi maailman suurimmista käännösorganisaatioista (Euroopan komissio 2009e, 0). DGT:n toimipaikat ovat Brysselissä ja Luxemburgissa, ja sen palveluksessa työskentelee noin 1 750 kokoaikaista kääntäjää (emt. 5), jotka jakautuvat kielten mukaisesti kieliosastoihin ja edelleen eri aihealueiden yhdistelmiin erikoistuneisiin yksiköihin (emt. 2–3). Poikkeuksen

---

<sup>2</sup> bulgaria, tšekki, tanska, hollanti, englantti, viro, suomi, ranska, saksa, kreikka, unkari, iiri, italia, latvia, liettua, malta, puola, portugali, romania, slovakki, sloveeni, espanja, ruotsi

<sup>3</sup> Sopimuksen konsolidoitu toisinto: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:115:0047:0199:FI:PDF>

muodostaa komission internet-sivujen sisältöä kääntävä ja kehittävä verkkokäännösyksikkö, jossa erikieliset kääntäjätiimit tekevät yhteistyötä (emt. 11). Koska tutkimukseni käsittelee suomen-, saksan- ja englanninkielisiä tekstejä, esittelen kyseisten kielten osastot hieman muita tarkemmin. Suomen osastoon kuuluu kolme yksikköä (DGT.A.FI.01–03), joista yksiköt 1 ja 2 sijaitsevat Brysselissä ja yksikkö 3 Luxemburgissa. Yhteensä suomen osastossa on tällä hetkellä 57 kääntäjän virkaa. Saksan osasto jakautuu puolestaan kuuteen yksikköön (DGT.B.DE.01–06), joista yksiköt 1–4 sijaitsevat Brysselissä ja yksiköt 5–6 Luxemburgissa. Saksan osastossa on kääntäjän virkoja 124. Englannin osasto koostuu suomen osaston tavoin kolmesta yksiköstä (DGT.B.EN.01–03), joista yksiköt 1 ja 2 sijaitsevat Brysselissä ja yksikkö 3 Luxemburgissa. Englannin osastossa kääntäjän virkojen määrä on 108. (Virkojen lukumäärät haettu komission yhteystietoluettelosta, joka oli päivitetty kyseisten yksikköjen osalta 4.5.2010; Euroopan komissio 2010.)

Vakituisten ja määräaikaisten in-house-kääntäjien lisäksi DGT käyttää alihankkijoita, jotka voivat olla käännöstoimistoja tai yksittäisiä kääntäjiä. Freelance-kääntäjiä tarvitaan, koska käännöstarve vaihtelee poliittisista tekijöistä johtuen, eikä sitä ole mahdollista täydellisesti ennakoita. Alihankkijoiden käyttö on lisääntynyt viimeisten kymmenen vuoden aikana, ja vuonna 2008 freelancerit käänsivät jo yli neljänneksen DGT:n kokonaistuotannosta. (Euroopan komissio 2009e, 9–10.) Periaatteena kuitenkin on, että alihankkijat eivät käänne EU:n lainsäädäntöä, vaan että sen kääntäisivät kokonaisuudessaan komission in-house-kääntäjät (Neves 2010).

Ensimmäisenä komission teksteistä tulevat usein mieleen nimenomaan lainsäädäntöön liittyvät säädöstekstit, joista tämän tutkimuksen tutkimusaineistokin muodostuu. Käännöstoimen pääosastossa käännetään kuitenkin lisäksi paljon muutakin komission toimintaan liittyvää viestintää. Käännettäviin asiakirjoihin kuuluu säädöstekstien lisäksi ainakin

puheita, puhemuistioita, tilanneselostuksia ja lehdistötiedotteita, kansainvälisiä sopimuksia, poliittisia lausumia, vastauksia parlamentin jäsenten kirjallisiin ja suullisiin kysymyksiin, teknisiä selvityksiä, taloushallinnon kertomuksia, pöytäkirjoja, sisäisiä hallinnollisia asioita ja henkilöstötiedotteita, filmien ja muun tiedotusmateriaalin käsikirjoituksia ja kuvatekstejä, kirjeenvaihtoa

ministeriöiden, yritysten, eturyhmien ja yksittäisten henkilöiden kanssa, verkkosivuja sekä erilaisia ja -muotoisia mielipidevaikuttajille ja suurelle yleisölle suunnattuja julkaisuja lukuisista eri aiheista. (Euroopan komissio 2009e, 4.)

Sen lisäksi, että käännettävien tekstilajien kirjo on valtava, sivumäärätkin ovat hulppeita. Vuonna 2008 DGT:ssä käännettiin kaikkiaan 1 805 689 sivua<sup>4</sup>, joista 72,5 prosenttia oli laadittu englanniksi, 11,8 prosenttia ranskaksi, 2,7 prosenttia saksaksi ja 13 prosenttia muilla kielillä. (Euroopan komissio 2009e, 6.)

Käännöstoimen pääosastosta kertovassa esittelyssä todetaan, että käännöksen laatuun vaikuttaa kääntäjien ja kääntämisen apuvälineiden lisäksi se, miten alkuteksti on kirjoitettu. On erityisen tärkeää toimia korkeiden laatuvaatimusten mukaisesti, kun tekstien laatijat ovat usein muita kuin laadintakieltä äidinkielenään puhuvia. Käännöstoimen pääosastolle onkin perustettu oma editointiyksikkönsä, jonka tehtäväksi on annettu alkutekstien kielen korjaaminen ja hiominen sekä tekstin laatijoiden ja asiakirjoja käännettävien osastojen ohjeistaminen. (Euroopan komissio 2009e, 5.) Käännöstoimen pääosastossa ulkoisen tiedotuksen parissa työskentelevältä Tytti Granqvistilta (2010) saamani tietojen mukaan editointiyksikkö käsittelee ainoastaan englanniksi ja ranskaksi laadittuja alkutekstejä, joista englanninkieliset muodostavat noin 90 prosenttia. Vuonna 2009 editointiyksikkö käsitteli yhteensä noin 37 200 sivua tekstiä, ja luku on Granqvistin mukaan pysynyt suurin piirtein samana viimeiset kolme vuotta. Kun näiden tietojen perusteella laskee editoitujen englanninkielisten sivujen määrän (33 480 sivua) ja vertaa sitä aiemmin esitettyihin tietoihin vuonna 2008 alun perin englanniksi laadittujen käännettäviksi tulleiden asiakirjojen määrästä (noin 1 309 125 sivua), voidaan todeta, että editointiyksikkö vaikuttaa lopulta vain murtoosaan kaiken käännettäväksi tulevan englanninkielisen materiaalin laadusta. Komission tekstit ovatkin poikkeuksellisia säädöstekstejä siinä mielessä, etteivät ne pääsääntöisesti ole äidinkielisten puhujien kirjoittamia tai editoimia. Tutkimusaineistoni kannalta asetelma on näin ollen mielenkiintoinen: tilastotietojen valossa voidaan arvella, että valtaosa englanninkielisistä teksteistä on muiden kuin äidinkielenään englantia puhuvien kirjoittamia, kun taas saksan- ja suomenkieliset tekstit ovat oletettavasti pääosin äidinkielisten kielen ammattilaisten, kääntäjien, kirjoittamia.

---

<sup>4</sup> Sivun on DGT:ssä 1 500 merkkiä ilman välilyöntejä. (Granqvist 2010.)



### 3.2.2 Kääntämisen apuvälineet

Kuten tekstityyppien kirjostakin jo voi päätellä, kääntäjien on hallittava monta eri kielen rekisteriä sekä kyettävä ymmärtämään moninaisia ja monimutkaisia asiakokonaisuuksia (Euroopan komissio 2009e, 4). Kääntäjiltä vaaditaan vankkaa harkinta- ja arvostelukykä (emt. 4). Toisaalta DGT:n esittelyssä todetaan myös hieman yksinkertaistaen, että ”termistön johdonmukaisuus *taataan* käyttämällä käännösmuisteja ja EU:n perustermistön sisältäviä tietokantoja” [kursivointi kirjoittajan]. Totuudenmukaisempaa olisi kai sanoa, että termistön johdonmukaisuus *pyritään* näillä kääntämisen apuvälineillä *varmistamaan*. Käännösvolyymit ovat niin suuria, ettei termistön täydelliseen johdonmukaisuuteen varmasti koskaan voida päästä. Seuraavassa esittelen kuitenkin kaksi toisiinsa kytkeytyvää merkittävää apuvälinettä, joiden avulla johdonmukaisuuteen tähdätään.

*Translator’s Workbench*: SDL Tradosin *Translator’s Workbench* eli *TWB* on paikallinen, kunkin kääntäjän omalle tietokoneelle luotava käännösmuisti, johon voidaan tallentaa käännösyksiköitä ja josta niitä voidaan hakea. Laajempia asiakirjapaketteja käännettäessä käytetään useamman kääntäjän jakamaa käännösmuistia, joka sijaitsee palvelimella, jolle kaikilla projektiin osallistuvilla on pääsy. Yhteismuisti on reaaliaikainen, mikä tarkoittaa, että kaikki tallennetut käännössegmentit ovat välittömästi projektin kaikkien kääntäjien käytössä. (Vuorinen 2010.) Komissio käyttää Tradosin kaupallisesta sovelluksesta omiin tarpeisiinsa sovitettua versiota. Se on ollut DGT:n kääntäjien käytävissä vuodesta 1997 alkaen, ja se koetaan hyödylliseksi, koska ”suuri osa komissiossa laadituista valmisteluteksteistä perustuu aiempiin teksteihin tai voimassa olevaan lainsäädäntöön”. (Euroopan komissio 2009e, 13.)

*Euramis*: Euramis on lyhenne englanninkielisistä sanoista *European advanced multilingual information system*, ja se on komissiossa kehitetty järjestelmä. Sen avulla voi käyttää monia kielen käsittelyyn liittyviä palveluja, joista yksi tärkeimpiä on keskuskäännösmuisti. Kun kysynhallinnasta vastaava keskusyksikkö hyväksyy käännöstoimeksiannon, alkuperäinen asiakirja lähetetään Euramis-järjestelmään ja asiakirjaan liittyvät mahdollisesti jo käännetty tekstijaksot poimitaan keskusmuistista. Hakutulokset voidaan sitten ladata kääntäjän koneella olevaan paikalliseen käännösmuistiin. Kun käännös on valmis (käännetty ja tarkastettu), kääntäjä lähettää paikallisen muistin (*clean-upin* yhteydessä) Euramoksen keskuskäännösmuistiin. (Euroopan komissio 2009e, 13.) *Clean-up* eli tiedoston ”puhdistaminen” tarkoittaa sitä, että sekä lähde- että kohdekieliset segmentit sisältävästä tiedostosta tehdään yksikielinen tiedosto, jossa mukana on pelkkä käännösteksti. Osa näin

kootusta keskuskäännösmuistista julkistettiin internetissä marraskuussa 2007, minkä jälkeen se on ollut julkisesti hyödynnettävissä (DGT & JRC 2008). Käsillä olevan tutkimuksen aineistona toimivat tämän julkistetun käännösmuistin englannin-, saksan- ja suomenkieliset osiot niiltä osin, kun niistä on voitu muodostaa kaksikielisiä korpuksia kielipareissa englanti–suomi ja englanti–saksa. Lisää tutkimuksen aineistona käytettävistä korpuksista kappaleessa ”Tutkimuskorpusten kuvaus”. Ennen omien tutkimuskorpusteni kuvausta käsittelen kuitenkin korpustutkimusta sekä kielitieteellisenä että käännöstieteellisenä tutkimusmetodina.

## 4. Korpustutkimus menetelmänä

### 4.1 Katsaus korpustutkimuksen historiaan

Koska aineistoni koostuu kahdesta sähköisestä aineistosta ja koska hyödynnän tutkimuksessani niiden jäsentämiseen korpustyökaluja, on aiheellista selvittää, miten kyseinen kielen ja kääntämisen tutkimiseen sovellettava menetelmä on kehittynyt ja mitkä ovat sille tyypillisiä piirteitä. Korpustutkimusta käsittelevän luvun aluksi tuntuisi tietysti luonnolliselta määritellä, mikä korpus oikeastaan on. Korpuksen määritelmä on kuitenkin aikojen saatossa muuttunut melko rajusti, minkä vuoksi kannattaa luoda silmäys menneeseen: se auttaa tässäkin tapauksessa ymmärtämään nykyisyyttä. Korpuslingvistiikasta saatetaan joskus puhua, kuin se olisi kielitieteen osa-alue samalla tavalla kuin vaikka semantiikka tai sosiolingvistiikka, joissa pyritään kuvaamaan ja selittämään jotakin kielen aspektia. Sitä se ei kuitenkaan ole, vaan korpuslingvistiikka on *metodi*, ja sitä voidaan käyttää melkein millä tahansa kielitieteen alalla (McEnery & Wilson 2001, 2). Lisäksi korpustutkimusta hyödyntävät kielitieteen lisäksi monet muutkin tieteenalat, kuten vaikkapa käänntiede korpuslingvistisen käänntutkimuksen piirissä, johon omakin tutkimukseni sijoittuu. Kun ymmärtää tämän, tajuaa samalla, että on mahdotonta antaa kattavaa kuvausta korpustutkimuksesta. Jotkin perusasiat auttavat silti sijoittamaan käsillä olevan tutkimuksen laajempaan korpuslingvistisen käänntutkimuksen yhteyteen. Pikainen katsaus korpuslingvistiikan historiaan auttaa ymmärtämään, miten korpustutkimus on saavuttanut nykyisenkaltaisen suosionsa – tie ei nimittäin ole ollut suoraviivainen tai vailla kivikkoja, kuten seuraavasta voidaan huomata. Historiakatsauksen näkökulma on väistämättä hieman kielitiedepainotteinen, sillä kielitieteellisen tutkimuksen piirissä erilaisilla korpuksilla on paljon käänntiedettä pidempi historia, minkä selittää tieteenalojen eripituinen olemassaolo.

#### 4.1.1 Varhainen korpuslingvistiikka

1900-luvun alkupuolen kielentutkimuksessa korpus ja sen analysoiminen kuuluivat vahvasti alan perusmetodologiaan. Tuolloin korpus käsitettiin laajaksi kokoelmaksi jonkin kielen tallennetussa muodossa olevia ilmauksia, ja vielä nykyäänkin se voi laajasti ottaen merkitä empiiristä aineistoa ylipäänsä. Tuon ajan tutkijat eivät ehkä määritelleet itseään korpustutkijoiksi, mutta esimerkiksi kenttätutkimusta tehneiden kielitieteilijöiden ja strukturalistiseen perinteeseen kiinnittyneiden kielentutkijoiden perusmetodologian voidaan katsoa olleen korpuspohjainen. McEnery ja Wilson nimittävätkin tätä 1950-luvun loppupuolelle ulottuvaa kautta varhaisen korpuslingvistiikan kaudeksi. (McEnery & Wilson

2001, 2–3.) Varhaisen korpustutkimuksen haasteena oli kuitenkin hyvin käytännöllinen ongelma: laajoihin korpuksiin liittyvä tiedonkäsittely tapahtui suurilta osin ilman koneiden apua (McEnery & Wilson 2001, 12). Monen miljoonan sanan korpuksien käsittelyyn tarvittiin suuri joukko ihmisiä, joista jokainen oli mahdollinen virhelähde (emt. 12). Lisäksi tällainen tiedonkäsittely oli hyvin hidasta ja kallista (emt. 13).

#### **4.1.2 Chomskyn kritiikki ja vastauksia siihen**

Kielitieteilijä Noam Chomsky esitti 1950-luvun lopulla kärkeästä kritiikkiä korpustutkimusta kohtaan, ja tätä kritiikkiä voidaan pitää merkittävänä vedenjakajana korpustutkimuksen historiassa (McEnery & Wilson 2001, 4). Käsittelemällä lyhyesti Chomskyn korpuksia kohtaan esittämää kritiikkiä on mahdollista ymmärtää, miksi korpuksilla oli niin pitkään marginaalinen rooli kielitieteellisenä tutkimusaineistona (emt. 5). Chomsky oli aikoinaan muun muassa sitä mieltä, ettei korpus olisi hyödyllinen työkalu kielitieteilijälle, koska tämän tulisi mallintaa performanssin sijaan kompetenssia (eli tutkia mitä on mahdollista sanoa sen sijaan, että tutkittaisiin, mitä tosiasiasa sanotaan). Toisin sanoen Chomskysta kielitieteilijän tehtävänä oli selittää ja luonnehtia tietoaamme kielestä eikä kuvata todellista kielenkäyttöä, johon ulkoiset tekijät saattoivat vaikuttaa häiritsevästi (McEnery & Wilson 2001, 6).

Chomskyn toinen kritiikin aihe korpustutkimusta kohtaan olivat aineiston mahdolliset vääristymät. Tällä hän viittasi siihen, että rakenteet, jotka ovat yleisesti ottaen harvinaisia tutkittavassa kielen varieteetissa, saattavat sattumalta olla yliedustettuina korpuksessa ja yleiset rakenteet vastaavasti aliedustettuina. McEnery ja Wilson myöntävät kritiikin olleen paikallaan, mutta toteavat samalla, että ilmiön aiheuttamiin vaikutuksiin voidaan vastata ainakin osittain kiinnittämällä enemmän huomiota korpuksen edustavuuteen. (McEnery & Wilson 2001, 75.) Chomskyn kanta oli myös, että koska korpus koostuu näytteestä eikä kaikista mahdollisista tutkittavista lauseista (joita kielessä voi olla loputon määrä!), se on vääristynyt eikä kuvaa edustavasti koko perusjoukkoa (McEnery & Wilson 77). McEnery ja Wilson toteavat kritiikin kuitenkin pätevän mihin tahansa tieteelliseen tutkimukseen, joka perustuu näytteiden tutkimiseen sen sijaan, että tutkittaisiin tutkimuksen kohteena olevaa, rajallista perusjoukkoa kokonaisuudessaan (emt. 78).

Oikeastaan Chomsky herätti henkiin ikaikaisen väittelyn empiristien ja rationalistien välillä. Rationalistit kehittävät teoriaa kielestä tietoisien introspektion avulla, empiristit puolestaan tutkivat luonnollisesti esiintyvää tutkimusaineistoa, todellista kielenkäyttöä, tyypillisesti

korpuksia hyödyntäen (McEnery & Wilson 2001, 5). Käännöstieteen tutkija Sari Eskola selventää asiaa toteamalla, että korpustutkimuksessa korostuvat kielellisten ilmiöiden frekvenssit ja korrelaatiot, jotka saadaan tutkimusaineistosta systemaattisesti ja empiirisesti. Introspektiossa tutkija puolestaan analysoi itse valitsemiaan esimerkkitapauksia. Eskola kuitenkin huomauttaa, että vaikka korpustutkijan tavoitteena onkin analysoida todellista kielenkäyttöä, myös kielen systeemi on jatkuvasti läsnä, ja käyttö ja systeemi ovat vuorovaikutuksessa keskenään. (Eskola 2002, 24.) Lisäksi tuntuu selvältä, että myös korpustutkijan käsitys kielestä ohjaa tämän kysymyksenasettelua ja jäsentää tutkijan korpuksesta tekemiä havaintoja. Tämäkään dikotomia ei siis todellisuudessa ole yhtä selvärajainen, kuin miltä ensikatsomalta saattaa vaikuttaa. Kritiikistä on ollut ainakin se hyöty, että korpustutkijat ovat joutuneet täsmentämään ja kehittämään menetelmäänsä ja käsityksiään vastauksena kritiikkiin. Esimerkiksi huomion kiinnittäminen korpuksen tasapainoisuuteen (*balance*) ja edustavuuteen (*representativeness*) ovat suoria vastauksia Chomskyn kritiikkiin (McEnery & Wilson 2001, 5).

#### **4.1.3 Moderni korpuslingvistiikka**

Tietotekniikan kehityksen myötä korpustutkimus sai uuden elämän elektronisessa muodossa olevien korpusten ansiosta. Ensimmäisiä sähköisiä korpuksia alettiin kerätä 1960-luvulla, jolloin korpustutkimuksen pioneirit Francis ja Kučera kokosivat ensimmäisen englanninkielisen sähköisen tekstiaineiston, *Brown-korpuksen*. Se sisältää yhteensä noin miljoona sanaa kirjoitettua amerikanenglantia (Eskola 2002, 25). Tästä ja muistakin sähköisistä korpuksista kerron lisää seuraavassa korpuksen määritelmiä käsittelevässä alaluvussa. Erityisen nopeasti korpuslingvistiikka alkoi kehittyä 1980-luvulla (McEnery & Xiao 2008, 18), mikä osuu hyvin yhteen tietokoneiden yleistymisen kanssa. Nykyisin termi korpus onkin lähes synonyyminen tietokoneella luettavissa olevan korpuksen (*machine readable corpus*) kanssa (McEnery & Wilson 2001, 17).

Modernissa korpuslingvistiikassa kieltä tutkitaan empiirisesti ja järjestelmällisesti laajojen, elektronisesti luettavissa olevien tekstimassojen avulla. Kun korpuksia käytetään tutkimuksessa, perusidea on yksinkertainen: kootaan hankkeen tarpeita vastaava elektroninen korpus ja analysoidaan sitä jonkinlaisella haku- eli konkordanssiohjelmalla. (Kujamäki & Jääskeläinen 2003, 243.) Korpuksen kokoamisperiaatteet (*corpus design*) ja korpusanalyysiin käytettävät koneelliset menetelmät muodostavat korpuslingvistiikan metodologian. (Laviosa 2002, 6) Juuri metodologia on ollut keskeinen tekijä tieteenalan kehitykselle (emt. 7), kuten

voidaan havaita toisaalta siitä, että korpustutkimus on kehittynyt ja täsmentynyt muun muassa Chomskyn kritiikin pohjalta (huomion kiinnittyminen kokoamisperiaatteisiin) ja toisaalta tehnyt uuden tulemisen automaattisen tiedonkäsittelyn kehittymisen myötä (menetelmät).

Kun Noam Chomsky 1950-luvun lopulla esitti kritiikkiään korpustutkimusta kohtaan, hän keskittyi problematisoimaan harvinaisten ja yleisten kielellisten elementtien suhdetta aineistossa, mikä antaa olettaa hänen käsittäneen korpustutkimuksessa hyödynnettävän kvantitatiivista lähestymistapaa. Tämä onkin osittain totta. Mahdollisimman edustavan, rajallisen otoksen ansiosta on nimittäin mahdollista ilmaista tuloksia määrällisesti ja vertailla niitä muihin tuloksiin samaan tapaan, kuin minkä tahansa muun otantaan perustuvan tieteellisen tutkimuksen tuloksia. McEnery ja Wilson nostavat kuitenkin esille näkökulman, joka uhkaa unohtua korpustutkimuksesta puhuttaessa: korpustutkimusta voi tehdä myös laadullisella tutkimusotteella, ja monet tutkijat ovat niin tehneetkin. (McEnery & Wilson 2001, 75.) *Kvantitatiivinen korpusanalyysi* perustuu kielen piirteiden luokitteluun, laskemiseen ja mahdollisesti myös monimutkaisten tilastollisten mallien hyödyntämiseen havaintojen selittämisessä. *Kvalitatiivisessa korpusanalyysissä* ei edes pyritä laskemaan frekvenssejä aineiston kielellisille piirteille. Siinä korpuksen muodossa olevaa tutkimusaineistoa hyödynnetään tiettyyn ilmiöön liittyvien aitojen, tosielämästä peräisin olevien esimerkkien varastona, ja kielenkäytön piirteitä pyritään yksilöimään ja kuvailemaan. McEnery ja Wilson kehottavat tarkastelemaan kvalitatiivista ja kvantitatiivista korpusanalyysia kahtena erillisenä näkökulmana korpustutkimukseen. Olennaista on se, etteivät tutkimusotteet välttämättä sulje toisiaan pois. (McEnery & Wilson 2001, 76.)

## **4.2 Korpuksen määritelmiä**

Seuraavassa keskityn määrittelemään, miten korpus käsitetään tämänhetkisen korpustutkimuksen piirissä. Kielitoimiston sanakirja antaa korpukselle seuraavan kielitieteeseen liittyvän määritelmän: se on ”tutkimuksen aineistoksi valittu kokoelma kielenainesta” (MOT Kielitoimiston sanakirja 2.0). Määritelmä on lyhykäisyydessään toimiva, sillä siihen tiivistyy peräti kolme oleellista korpuksen piirrettä: korpus on jonkun 1) tietoisesti kokoama 2) kieltä sisältävä 3) tutkimusaineisto. Korpustutkijat toki syventävät ja tarkentavat määritelmää tieteellisessä kirjallisuudessa monilta osin. Yksinkertaisimmillaan korpus viittaa kuitenkin tekstimassaan. Kennedy laajentaa korpuksen käsitteeseen kirjoitettujen tekstien lisäksi myös litteroidusta puheesta koostuvaa tekstimassaa, jota voi käyttää kielen analysoimiseen ja kuvailemiseen (Kennedy 1998). Käännöstieteen piirissä

Kujamäki ja Jääskeläinen käyttävät artikkelissaan korpuksia ja elektronisia tekstiaineistoja suhteellisen synonyymisinä, mutta huomauttavat, että ”*elektroninen korpus* on – erotukseksi muista elektronisista tekstiaineistoista – huolellisesti suunniteltu ja jollain periaatteella koottu kokonaisuus” (2003, 251). Kujamäen ja Jääskeläisen kommentin myötä tulee esille yksi oleellinen nykyaikaisen korpuksen piirre: se on elektroninen ja sitä voi näin ollen käsitellä tietokoneen avulla.

*Käännöstieteellisen* korpustutkimuksen yksi keskeisimmistä hahmoista on Mona Baker, joka kirjoitti korpustutkimuksesta ja sen tarpeellisuudesta kääntämisen ominaisuuteen selvittämiseksi jo 1990-luvun alkupuolelta peräisin olevassa Euroopan komissiolle luovutetussa raportissa (mainitaan mm. artikkelissa Baker 1995, 223). Hän toteaa korpuksen alun perin tarkoittaneen mitä tahansa kirjoitusten kokoelmaa, joka on ollut käsitellyssä tai käsittelemättömässä muodossa (Baker 1995, 225). Lisäksi korpus oli Bakerin mukaan varhaisessa muodossaan usein ainoastaan yhden tietyn henkilön kirjoittama, kuten tietyn kirjailijan tuotannosta muodostettu. Bakerin mukaan korpuksen määritelmä oli jo vuoteen 1995 mennessä muuttunut kolmella merkittävällä tavalla (emt. 225):

1) Korpus tarkoittaa tietokoneella luettavissa olevaa tekstien kokoelmaa, joka on analysoitavissa automaattisesti tai puoliautomaattisesti monin eri tavoin.

2) Korpuksen käsitettä ei enää nykyään rajata tarkoittamaan ainoastaan kirjoituksia, vaan se voi sisältää kirjoitettujen tekstien sijaan myös alun perin puhuttuja tekstejä (vrt. edellä Kennedy 1998).

3) Korpus voi koostua useista eri lähteistä peräisin olevista, monien eri kirjoittajien tai puhujien tuottamista eri aihepiirejä koskevista teksteistä.

Erityisen tärkeäksi korpuksien muista tekstikokoelmista erottavaksi tekijäksi Baker mainitsee sen, että se on koottu tiettyä tarkoitusta varten selkeiden kriteerien perusteella (Baker 1995, 225). Eskola mainitsee merkittävimmäksi hyvin muodostetun korpuksen edellytykseksi edustavuuden (*representativeness*) (2002, 25). Se, että luonnollinen kieli on ääretön ilmiö, asettaa omat haasteensa edustavan otoksen muodostamiselle (Eskola 2002, 26). Kieli ei sovellu alistettavaksi tilastotieteessä yleisesti käytettäville otantamenetelmille, eikä korpuslingvistiikassa vallitse laajaa yksimielisyyttä validin ja edustavan korpuksen

täsmällisistä kriteereistä. Erilaisten keräämisperiaatteiden huomioimista pidetään kuitenkin ratkaisevan tärkeänä tulosten luotettavuuden kannalta. Vaikka muodostamiskriteerit voivatkin vaihdella tutkimuksen tavoitteiden mukaan, jokainen korpuksen laatija joutuu väistämättä valitsemaan ainakin korpustyyppin (eri tyytit esitellään myöhemmin), korpuksen koon (kuinka monta sanetta), tekstien otoskoon (koko teksti vai tietynpituisen osuus tekstistä), niiden lukumäärän, kirjoittajayksilöiden edustavuuden sekä aikakauden. Suunnitelmallisuus ja järjestelmällisyys ovat edellytys sille, että elektronisesta aineistosta muodostuu korpus eikä mikä tahansa lausekokoelma. (Eskola 2002, 26.) Tämä on siis käsitys tiukan määritelmän mukaisesta korpuksesta käännöstieteellistä korpustutkimusta tekevien tutkijoiden mielestä. Seuraavassa esittelen muutaman sähköisen korpuksen, joista osa on syntynyt kielitieteellisessä ja osa käännöstieteellisessä viitekehyksessä. Lyhyiden esittelyiden tarkoituksena on osoittaa, että sähköisiksi korpuksiksi kutsutaan nykyäänkin hyvin monenlaisia aineistoja.

Elektronisten korpusten jo edellä mainitut pioneerit Francis ja Kučera kehittivät 1960-luvulla laajan *Brown-korpuksen*, joka koostui senhetkisestä amerikanenglannista. Korpus sisältää ainoastaan vuonna 1961 julkaistuja proosatekstejä (esimerkiksi laajan kirjon erilaisia asiatekstejä ja kaunokirjallisuutta), jotka olivat äidinkielenään amerikanenglantia puhuvien kirjoittamia. Korpus koostuu 500:stä noin 2 000 sanan tekstinäytteestä, jotka eivät siis ole kokonaisia tekstejä. Kategoriat ja niiden sisältämät tekstimäärät päätettiin tutkijoiden toimesta, jonka jälkeen näytteet pyrittiin keräämään satunnaisesti. *Brown-korpus* on siis hyvin tarkkojen kriteereiden perusteella koostettu, ja näin on pyritty mahdollisimman hyvään edustavuuteen (vrt. Chomskyn kritiikki). (Brown Corpus Manual 1997.) *Brown-korpusta* uudempi ja laajempi brittiläinen korpus, *British National Corpus* eli *BNC*, pyrkii edustamaan nykyaikaista yleiskieltä. Sen vanhimmat tekstit ovat muutamaa poikkeusta lukuun ottamatta vuodelta 1975 ja sen kokoaminen aloitettiin 1991 (BNC 2009a ja 2009b). Brownista poiketen *BNC* sisältää myös litteroitua puhetta. Tässäkin korpuksessa on tekstejä monilta eri aihealueilta, jotka määriteltiin ennen kokoamisen aloittamista. Luokittelu oli kuitenkin hyvin erilainen kuin *Brown-korpusta* koottaessa. Tekstien moninaisesta luonteesta johtuen kolme eri tahoja oli vastuussa niiden keräämisestä (BNC 2009c).

Savonlinnassa koottua *Käännössuomen korpusta* käsittelen tarkemmin alaluvussa 4.3.2, mutta mainittakoon tässä yhteydessä sen verran, että se on tärkeä suomalaisessa korpuslingvistisessä käännöstutkimuksessa käytetty verrannollinen korpus, joka sisältää



seitsemän eri tekstilajin toisaalta alun perin suomeksi tuotettuja ja toisaalta eri kielistä suomennettuja tekstejä (Jantunen 2004, 40). Esimerkiksi *Brown-korpuksesta* poiketen *Käännössuomen korpus* sisältää ainoastaan kokonaisia tekstejä (Jantunen 2004, 41).

Varmaankin yksi löyhimmin korpukseksi määriteltävissä olevista korpuksista on *WebCorp Live*, joka on internet-pohjainen työkalukokoelma, jonka avulla internetiä voi käsitellä tekstikorpuksen tavoin. *WebCorp* on kehitetty Birmingham City Universityn yksikössä *The Research and Development Unit for English Studies* (Birmingham City University 2010). Projektin tavoitteena on kaikkein uusimpien ja harvinaisimpienkin kielen ominaisuuksien tuominen tutkijoiden, opettajien, opiskelijoiden ja muiden kielestä kiinnostuneiden ulottuville. Sen vahvuutena on tuoda laajasta aineistosta runsaasti sanojen ja fraasien käyttöesimerkkejä. Korpus edustaa myös luonnollista kielenkäyttöä. Mitään tiukkoja määrällisiä analyysejä korpuksen avulla tuskin on tarkoituksenmukaista tehdä, mutta se tarjoaa silti mielenkiintoisen näkökulman kieleen nimenomaan todellisten käyttöesimerkkien avulla. (WebCorp Live 2010.) Toinen internetiä lähteenä hyödyntävä korpusprojekti on Jörg Tiedemanin luotsaama *OPUS – an open source parallel corpus* (Tiedeman 2010), jossa kerätään ja kohdistetaan internetissä julkaistuja käännettyjä tekstejä. Tavoitteena on tuottaa julkisesti saatavilla olevia eriaiheisia rinnakkaiskorpuksia. Tällä hetkellä Euroopan unioniin liittyviä korpuksia on internet-sivun kautta saatavissa ainakin kaksi: *Euconst – The European constitution* ja *EUROPARL – European Parliament Proceedings*.

Korpuksiksi kutsutaan siis nykyäänkin monenlaisia tekstien kokoelmia. Yhtenä korpustutkimuksessa käytettävän korpuksen tärkeimmistä kriteereistä voikin nähdäkseni pitää sitä, että se soveltuu vastaamaan siihen tutkimuskysymykseen, jonka tutkija sille esittää. Laajat kansalliset korpuksset, kuten *BNC* ja *Brown* voivat kertoa tutkijalla, miten tiettyä kielen kansallista varieteettia on käytetty tietynä aikana. *Käännössuomen korpuksella* selvitetään käännösprosessiin liittyviä mahdollisia lainalaisuuksia, käännösuniversaaleja, joita voi kartoittaa esimerkiksi vertaamalla alun perin suomeksi kirjoitettujen tekstien piirteitä käännössuomen piirteisiin, ja *WebCorpin* etuna on puolestaan tekstimassan laajuus ja tuoreus. *OPUS* sekä muutkin rinnakkaiskorpuksset tarjoavat kääntämisen ja kielen tutkijoille monikielistä tutkimusaineistoa.

### **4.3 Korpuslingvistinen käänöstutkimus**

Kuten jo aiemmin totesin, koko korpustutkimuksen laajaan kirjoon olisi mahdotonta luoda edes pintapuolista katsausta. Niinpä keskityn seuraavassa nimenomaan käänöstieteen piirissä tehtyyn tutkimukseen, joka on nykyisellään jo itsessäänkin hyvin runsasta.

Korpusmenetelmät ovat vallanneet alaa käänöstieteellisessä tutkimuksessa toisaalta 1970-luvulta lähtien tapahtuneen kohdetekstin merkitystä korostavan deskriptiivisen käänöstutkimuksen yleistymisen ansiosta (Eskola 2002, 19) ja toisaalta yleisesti korpuslingvistiikassa tapahtuneen kehityksen vuoksi (emt. 28). Deskriptiivisessä tutkimuksessa ollaan kiinnostuneita siitä, mitä käännettäessä todella tapahtuu, joten todellisissa tilanteissa syntyneistä teksteistä muodostetut korpukset ovat tämän näkökulman kannalta arvokasta aineistoa. Jos ajatellaan tieteen kenttää laajemmin ottaen, niin yleinen kansainvälistyminen ja kiihtyvä maapalloistuminen (globalisaatio) ovat osaltaan vaikuttaneet siihen, että käänöstieteelliset ja kontrastiiviset tutkimukset ovat tällä hetkellä suosittuimpia kuin koskaan. Oma tutkimuksenikin kumpuaa kansainvälisen yhteistyön lisääntymisestä ja tiivistymisestä, joka on saanut konkreettisen muotonsa Euroopan unionissa. Korpusten ja erityisesti monikielisten korpusten merkitys on korostunut yllä mainitun tutkimuksen uudessa aallossa. Korpuslingvistiikka on kehittynyt kiihtyvällä vauhdilla 1980-luvulta lähtien, ja kenttää edelleen hallitsevien englanninkielisen korpusten rinnalle on syntynyt myös muunkielisiä korpuksia. (McEnery & Xiao 2008, 18.)

Ensimmäiset tutkimukset, jotka perustuivat käännettyjen tekstien korpuksiin, julkaistiin 1980-luvun loppupuoliskolla. Niissä pyrittiin arvioimaan käännetyn kielen laatua verrattuna alkuperäiskielen laatuun. Nämä tutkimukset syntyivät vertailevan kielitieteen piirissä. 1990-luvun alusta lähtien korpustutkimus on ollut kiinteä osa käänöstiedettä, ja se on yleistynyt nopeasti sekä vaikuttanut voimakkaasti kääntämisen opetukseen, opiskeluun ja käsitteenmuodostukseen. (Laviosa 2002, 1.) Väljästi määriteltynä kaikkea korpuslingvistiikan työkaluja ja menetelmiä hyödyntävää käänöstiedettä voidaan kutsua korpuslingvistiseksi käänöstutkimukseksi, vaikka alan tutkimus onkin keskittynyt pitkälti Mona Bakerin lanseeraamaan käänösuniversaalien kartoittamiseen (Eskola 2002, 30). Korpuslingvistisen käänöstutkimuksen keskeisimpiä uranuurtajia on ollut Bakerin lisäksi Sara Laviosa. Baker on antanut sysäyksen koko suuntaukselle ehdottamalla korpuslingvistiikkaa yhdeksi käänöstutkimuksessa sovellettavista menetelmistä, esittämällä ajatuksen sekä myöhemmin tarkemman rajauksen universaaleista hypoteeseista sekä käsittelemällä eri korpustyyppien soveltuvuutta käänöstutkimukseen (Eskola 2002, 28–29). Laviosa on puolestaan Bakerin

ohjauksessa tekemässä väitöskirjassaan ensimmäisenä tutkijana soveltanut Bakerin ajatuksia käännettyjen ja alun perin kohdekielellä kirjoitettujen tekstien vertailuun (emt. 29).

Korpusmenetelmiä hyödyntävässä tutkimuksessa voidaan menetelmällisessä mielessä erottaa induktio ja deduktio. Induktiivisessa eli korpusvetoisessa (*corpus-driven*) tutkimuksessa aloitetaan havainnoista, jotka voivat kummuta esimerkiksi pelkistä frekvenssilistoista, jonka jälkeen muodostetaan hypoteeseja ja sitä kautta siirrytään yleistyksiin ja teorianmuodostukseen. Tämä lähestymistapa on korpuslingvistiikassa yleinen: usein tavoitteena on analysoida tiettyjä kielen piirteitä aineiston pohjalta ilman etukäteisolettamuksia. Deduktiivisessa eli korpuspohjaisessa (*corpus-based*) tutkimuksessa puolestaan lähtötilanteessa on jo jokin olemassa oleva teoria, jota sitten testataan ja täydennetään tilastollisilla tuloksilla. Prosessissa siis pyritään joko verifioimaan tai falsifioimaan teoreettisella tasolla esitettyjä hypoteeseja. Korpuspohjaisesta lähestymistavasta on esimerkkinä käänösuniversaaleja kartoittava korpustutkimus. (Eskola 2002, 29.) Eskola kuitenkin huomauttaa, etteivät induktio ja deduktio ole suoranaissessa ristiriidassa keskenään ja että ne voivat olla keskenään vuorovaikutuksessa samankin tutkimuksen sisällä. (Eskola 2002, 30.)

Bakerin näkemyksen mukaan (Eskola 2002, 29–30) korpuslingvistisessä käänöstutkimuksessa päämääränä on testata käänöskieltä koskevia, lähdekielestä riippumattomia yleisiä oletuksia, kun taas korpuslingvistiikassa voidaan olla kiinnostuneita myös yksittäisistä kielen ilmentymistä, kuten tietyn sanan käytöstä. Eskola kuitenkin katsoo yksittäisten syntaktisten ja leksikaalisten piirteiden tutkimisen myös käänösteksteistä olevan tieteellisesti merkittävää ja kiinnostavaa kyseessä olevan kieliparin ja kohdekulttuurin kannalta. Hän nostaa universaalien piirteiden tutkimuksen ohella paikallisten lainalaisuuksien tutkimisen tärkeäksi näkökulmaksi (Eskola 2002, 30), mikä tukee omaa käsitystäni korpuslingvistisen käänöstutkimuksen mahdollisuuksista paljastaa yksittäisen kielellisten ilmentymien kautta jotain laajempaa lähde- ja kohdekielten välisistä eroista – joiden tunteminen ja ymmärtäminen puolestaan voivat edistää käänösprosessissa ilmenevien haasteiden ymmärtämistä ja helpottaa niiden ratkaisemista.

Laviosa katsoo korpustutkimuksella olevan mahdollisuuksia paljastaa käänösprosessista ja siinä syntyvästä tuotteesta jotain uutta, joka on johdonmukaista ja perustuu vankalle empiiriselle perustalle. (Laviosa 2002, 1.) Käänöstieteellisen korpustutkimuksen

teoreettisessa suuntauksessa korpuksia käytetäänkin pääasiassa käännösprosessin tutkimiseen. Niiden avulla voidaan katsoa, miten yhdellä kielellä esitetty käsite ilmaistaan toisella kielellä. Korpusten avulla voidaan myös vertailla kielellisiä piirteitä ja niiden frekvenssejä käännettyissä teksteissä ja verrannollisissa alkukielellä kirjoitetuissa teksteissä. (McEnery & Xiao 2008, 22.) Tällaisen vertailun avulla voidaan yksilöidä sellaisia käännösten kielen piirteitä, jotka eroavat alkukielisen tekstin piirteistä, mikä puolestaan antaa uutta tietoa käännösprosessista ja kääntämisen normeista (emt. 23). Uuden tiedon lisäksi kääntäjät ja kääntämisen opiskelijat voivat tällaisten tutkimusten avulla tiedostaa käännösten kieleen liittyviä ongelmia ja epätyypillisyyksiä (emt. 23–24). Nimenomaan käännöskielen erityispiirteiden selvittämisestä on kyse, kun tutkimuksessa kartoitetaan kääntämisen universaaleja piirteitä. Tämä on nähdäkseni hyödyllistä sen vuoksi, että ainoastaan tiedostamalla nämä haasteet käännöskielisyyden ongelmakohtiin voidaan puuttua ja tavallisimmat sudenkuopat jopa välttää. Soveltavassa suuntauksessa korpuksia hyödynnetään kääntäjien koulutuksessa ja erilaisten sovellusten kuten konekäännösohjelmien sekä tietokoneavusteiseen kääntämiseen (CAT) liittyvien järjestelmien kehittämisessä (emt. 22).

#### **4.3.1 Korpuslingvistisen käännöstutkimuksen korpustyytit**

Korpuksia eivät ole mikään kovin homogeeninen ryhmä, mikä tulee hyvin esiin Tony McEneryn ja Richard Xiaon (2008) artikkelista, jossa he määrittelevät, mihin eri kategorioihin korpuksia voidaan jakaa. Jaotteluperusteista riippuen korpuksia voidaan jakaa eri ryhmiin esimerkiksi kielten lukumäärän tai korpuksen sisällön perusteella. Kielipohjaista luokittelua käyttäen korpus voi olla joko yksi-, kaksi- tai monikielinen. Jos jaotteluperusteena puolestaan käytetään korpuksen sisältöä, sen voidaan sanoa olevan käännöskorpus (*translation/L2 corpus*) tai ei-käännöskorpus (*non-translation/L1 corpus*). (McEnery & Xiao 2008, 19.)

Tutkimuskysymyksestä riippuen korpus voi olla joko erikoiskorpus, jolloin se sisältää tietyn tekstityypin tai erikoisalan tekstejä tai yleiskielen korpus (eli yleiskorpus), jolloin se sisältää mahdollisimman sopusuhtaisesti monta eri tekstityyppiä monista eri lähteistä. (McEnery & Xiao 2008, 21; Eskola 2002, 26.) Kysymyksenasettelusta riippuu myös se, onko korpus yksi-, kaksi- vai monikielinen ja koostuuko se alkukielisistä vai käännettyistä teksteistä, vai molemmista (Eskola 2002, 32). Korpusten sisältö voi vaihdella senkin mukaan, sisältävätkö ne tekstiä sellaisenaan ilman mitään lingvististä analyysiä vai onko sisältö lingvistisesti annotoitu, jolloin tekstiin on joko automaattisesti, puoliautomaattisesti tai manuaalisesti

liitetty esimerkiksi morfosyntaktinen analyysi. Melko yleistä on, että korpuksat sisältävät metalingvististä tietoa eli esimerkiksi tiedon kirjoittajasta tai puhujasta, julkaisuajankohdasta ja tekstityypistä. Korpuksat voivat lisäksi olla staattisia tai muuttuvia. Muuttuvien korpusten sisältöä päivitetään jonkin näkökulman mukaisesti: esimerkiksi monitorikorpusten avulla pyritään tarkastelemaan tietyn ajankohdan kielenkäyttöä. (Arppe 2002.)

McEnery ja Xiao toteavat kaksi- ja monikielisten korpusten olevan verrattain uusi ilmiö ja niihin liittyvän tutkimuksen olevan pääosin peräisin 1990-luvulta. Tämä selittää heidän mukaansa sen, että näihin korpuksiin liittyvä terminologia on vielä osin vakiintumatonta. (2008, 18.) Eskolakin toteaa terminologian kirjavuuden johtuvan siitä, että korpustutkimus [modernissa muodossaan] on vielä hyvin nuori ala. Lisäksi hän huomauttaa, ettei korpuslingvistiikkaa ja korpuslingvististä käännöstutkimusta yhdistävää vakiintunutta käytäntöä vielä ole. (Eskola 2002, 32.) Terminologisesta horjumisesta johtuen esittelen seuraavaksi ainoastaan korpuslingvistisen käännöstutkimuksen kannalta oleellimmat korpustyyppit (Eskolan mukaan) nimenomaan käännöstutkimuksessa käytetyin termein.

Korpuslingvistisen käännöstutkimuksen kannalta kaikkein käyttökelpoisimpia korpustyyppit ovat:

- a) lähdeteksteistä ja niiden käännöksistä koostuvat korpuksat sekä
- b) samalla kielellä kirjoitetuista toisaalta käännetyistä ja toisaalta alun perin kohdekielellä tuotetuista teksteistä koostuvat korpuksat. (emt. 32)

Tyyppin a) korpuksista käytetään yleisesti sekä termiä käännöskorpus (*translation corpus*) että rinnakkaiskorpus (*parallel corpus*). Korpuslingvistiksessä käännöstutkimuksessa tyyppin b) korpusta kutsutaan verrannolliseksi korpukseksi (*comparable corpus*), kun korpuslingvistiikan puolella kyseistä termiä käytetään myös samoin kriteerein valittujen erikielisten tekstien muodostamasta korpuksesta. (Eskola 2002, 32.) Mona Bakerin käyttöön näistä termeistä ovat valikoituneet rinnakkaiskorpus ja verrannollinen korpus (Baker 1995, 230). Koska Baker on käännöstieteellisessä korpustutkimuksessa niin keskeinen tutkija, ja koska myös Suomessa urauurtavaa korpustutkimusta tehneet tutkijat (esim. Jantunen ja Eskola) ovat käyttäneet hänen terminologiaansa, minäkin jäsennän omaa tutkimustani näiden käsitteiden avulla.

*Rinnakkaiskorpus* on korpus, joka sisältää lähdetekstejä sekä niiden käännöksiä yhdelle tai useammalle kohdekielelle. Se voi olla yksi- tai kaksisuuntainen (esim. koostua saksankielisistä alkuteksteistä ja niiden suomennoksista sekä suomenkielisistä alkuteksteistä ja niiden saksankielisiä käännöksistä). (Eskola 2002, 33.) McEneryn ja Xiaon mukaan rinnakkaiskorpus voi olla jopa monisuuntainen, jolloin korpuksessa on monta eri kieliversiota samasta tekstistä. Yhdeksi esimerkiksi monisuuntaisesta korpuksesta he antavat EU:n asetukset, jotka laaditaan samanaikaisesti eri kielillä. (McEnery & Xiao 2008, 20.) Rinnakkaisuus on kuitenkin vain niiden virallinen status, todellisuudessa asetuksetkin syntyvät käsitykseni mukaan pääasiassa yhdellä tietyllä työkielellä, ja Euroopan komission käännöstoimen pääosastossa niitä käsitellään tavallisina käännösasiakirjoina.

Rinnakkaiskorpuksia käytetään pääasiassa käännöstieteellisessä tutkimuksessa (McEnery & Xiao 2008, 20). Niitä suunniteltaessa ja koottaessa kiinnitetään koontikriteerien (*sampling frame*) lisäksi erityistä huomiota alkutekstien ja niiden käännösten kohdistamiseen. Rinnakkaiskorpuksen hyödyllisyys riippuu paljon juuri kohdistamisesta eli siitä, että alkutekstin ja käännöksen välille on luotu yhteys lauseiden, sanojen tai muiden tekstin osien välillä. (emt. 20–21.) Rinnakkaiskorpuksia mahdollistavat kääntäjien todellisten ratkaisujen analysoimisen. Lisäksi niitä hyödynnetään kääntämisen eri sovelluksissa, kuten konekääntämisen kehittämisessä (Eskola 2002, 33; Baker 1995, 231) sekä kielten yhteneväisyyksiä ja eroja kartoittavassa kontrastiivisessa tutkimuksessa (Eskola 2002, 33), jossa käännöksillä tosin on kiistanalainen asema tutkimusaineistona.

*Verrannollinen korpus* on vähintään kahdesta tietyin kriteerein toisiinsa nähden verrannollisesta (yksikielisestä) tekstimateriaalista koostuva korpus. Verrannollisuuden ansiosta tekstimassoja voidaan vertailla toisiinsa samoista lähtökohdista. Eskola katsoo niiden olevan hedelmällisimmillään koostuessaan käännetyistä ja vastaavista alkukielisistä teksteistä eli niiden ollessa tyyppin b) mukaisia. (Eskola 2002, 34.) Jos verrannollinen korpus on huolellisesti koostettu, sen avulla voidaan arvioida käännösten ja alun perin kohdekielellä tuotettujen tekstien ominaispiirteitä. Havaittujen erojen perusteella voidaan sitten tehdä johtopäätöksiä jopa kääntämisen universaaleihin piirteisiin asti (emt. 35). Käännösuniversaalien tutkimuksesta kerron lisää seuraavassa suomalaista korpusmenetelmiä hyödyntävää käännöstutkimusta esittelevässä alaluvussa.

### 4.3.2 Korpuslingvistinen käännöstutkimus Suomessa

Savonlinnassa aiemmin sijainneessa Joensuun yliopiston alaisessa Kansainvälisen viestinnän laitoksessa on tehty urauurtavaa työtä suomalaisen korpustutkimuksen saralla (Eskola 2002, 36). Vuosina 1997–2001 siellä toimi Suomen Akatemian rahoittama projekti *Käännössuomi ja kääntämisen universaalit – tutkimus korpusaineistolla* (Jantunen 2004, 5). Professori Anna Maurasen luotsaaman projektin puitteissa koottiin *Käännössuomen korpuksen* nimeä kantava verrannollinen korpus, joka sisältää alun perin suomeksi kirjoitettuja tekstejä sekä näille verrannollisia käännösprosessin tuloksena syntyneitä suomennoksia. Tekstejä on seitsemästä eri tekstilajista ja käännöksiä kymmenestä eri kielestä (Jantunen 2004, 40). Jantunen toteaa *Käännössuomen korpuksen* olevan erikoislaatuinen lajissaan, sillä sen käännöskielen osuus on peräisin useasta kielestä, kun taas aiemmissa alkuperäis- ja käännöskieltä sisältävää tutkimusaineistoa hyödyntäneissä tutkimuksissa käännöstekstit oli tehty vain yhdestä lähdekielestä. Lisäksi se on harvinaisen yhteismitallinen, sillä sen molemmat osat on koottu samanaikaisesti ja samanlaisia koontiperiaatteita noudattaen, toisin kuin esimerkiksi ensimmäinen varsinainen elektroninen verrannollinen korpus, Laviosan koostama *English Comparable Corpus*, jossa on hyödynnetty osin myös aiemmin koottua korpusta. (Jantunen 2004, 41.)

Edellä mainitun projektin tavoitteet tiivistävät hyvin, millaisista aiheista korpuslingvistinen käännöstutkimus on ylipäänsä kiinnostunut. Jantunen mukaan projektin päämääränä oli kartoittaa sellaisia käännössuomelle ominaisia kielellisiä piirteitä, jotka erottavat sen alkuperäissuomesta. Toinen tavoite oli saada lisätietoa käännösuniversaalien tutkimusta varten sekä kehittää tutkimusmetodeja. (Jantunen 2004, 5.) Käännösuniversaaleilla tarkoitetaan käännöskielessä lähde- ja kohdekielestä riippumatta esiintyviä yleisiä ominaisuuksia (Jantunen 2004, 5), ja verrannollisia korpuksia onkin käytetty enimmäkseen niiden tutkimiseen (emt. 45).

Andrew Chesterman jaottelee käännösuniversaalihypoteesit S- ja T-universaaleihin. S-kirjain tulee englanninkielisestä sanasta *source*, ja S-universaaleilla hän viittaa sellaisiin hypoteeseihin, jotka liittyvät käännösten ja lähdetekstien välisiin universaaleihin eroihin, eli siihen, miten kääntäjät prosessoivat lähdetekstiä. T-kirjain on puolestaan peräisin sanasta *target*, ja T-universaaleiksi Chesterman nimittää hypoteeseja, jotka esittävät väitteitä käännösten ja alun perin kohdekielellä kirjoitettujen tekstien eroista, eli kääntäjän tavoista käyttää kohdekieltä. Näiden erilaisten hypoteesityyppien tutkimiseen käytetään myös

erityyppisiä korpuksia eli rinnakkaiskorpuksia ja verrannollisia korpuksia. (Chesterman 2004, 39.) Chesterman luettelee artikkelissaan mahdollisiksi S-universaaleiksi mm. Touryn esittämän interferenssin ja standardisoinnin ja mahdollisiksi T-universaaleiksi Bakerin konventionaalistumisen, Laviosa-Braithwaiten yksinkertaistumisen, Maurasen epätyypilliset leksikaaliset kombinaatiot ja Tirkkonen-Conditin kohdekielispesifisten uniikkiainesten aliedustumisen. Tutkimus etenee näiden yleisten väitteiden operationalisoimiseen ja testaamiseen erilaisilla aineistoilla. (Chesterman 2004, 40.) On tärkeää huomata, että käänösuniversaalin käsite on myös monella tapaa ongelmallinen. Hypoteesit muun muassa perustuvat yksinomaan nykykieltä käsitelleisiin tutkimuksiin ja hyvin niukkaan kielivalikoimaan (Jantunen 2004, 50). Chesterman huomauttaakin, että on tärkeää pohtia niiden todellista universaaliutta ja harkita, pätevätkö oletukset esimerkiksi vain tiettyyn käännösten osajoukkoon kaikkien käännösten sijasta (Chesterman 2004, 40).

Käsillä oleva tutkimus ei sijoitu universaalitutkimuksen piiriin, vaan hyödyntää korpustutkimusta metodina englannin-, suomen- ja saksankielisten Euroopan komission tekstien vertailevassa tutkimuksessa. Tutkimuskorpukseni ovat rinnakkaiskorpuksia, jonka tyyppisten korpusten tärkeimmäksi merkitykseksi käänöstieteelle Baker mainitsee sen, että ne tukevat muutosta preskriptiivisyydestä deskriptiivisyyteen (Baker 1995, 231). Suurin osa tutkimuskorpusten suomen- ja saksankielisistä teksteistä on syntynyt käänösprosessin välityksellä<sup>5</sup>, ja tutkimalla suomennoksia ja saksannoksia voidaan saada kiinnostavaa tietoa siitä, miten kääntäjät ovat ratkaisseet käännösongelmia käytännössä. Bakerkin mainitsee tämän yhdeksi merkittävimmistä tavoitteista rinnakkaiskorpuksilla tehtävälle käänöstutkimukselle, ja lisää sen mahdollistavan realististen mallien kehittämisen kääntämisen opiskelijoille (emt. 231). Seuraavassa luvussa esittelen tarkemmin tutkimukseni aineiston muodostavat korpukset.

---

<sup>5</sup> Ks. tarkempi selvitys luvusta "Tutkimuskorpusten kuvaus"



## 5. Tutkimuskorpusten kuvaus

### 5.1 Lähdeaineiston kuvaus

Tutkimuksessa käyttämäni korpuksat on muodostettu Euroopan komission käännöstoimen pääosaston (*Directorate-General for Translation*, jatkossa DGT) marraskuussa 2007 julkaiseman monikielisen käännösmuistin pohjalta. Sitä esittelevän verkkosivun otsikko kuvaa käännösmuistia hyvin: *The DGT Multilingual Translation Memory of the Acquis Communautaire: DGT-TM*. Kyseinen käännösmuisti on saatu *Euramixesta* (*European advanced multilingual information system*), jonka mainitsin jo toimintakontekstia esitelleessä luvussa ja joka on DGT:n yhteisesti käyttämä laaja käännöstietokanta. (Yhteinen tutkimuskeskus 2008.) Yleisön käyttöön annettu käännösmuisti pitää sisällään suurimman osan niin kutsutusta *acquis communautairesta* (Yhteinen tutkimuskeskus 2008) eli yhteisön säännöstöstä, joka on kaikkia Euroopan unionin jäsenvaltioita sitova oikeusperusta. Käännösmuistin julkistamishetkellä *acquis communautaireen* kuului

perussopimusten sisältö, periaatteet ja poliittiset tavoitteet; perussopimusten nojalla hyväksytyt lainsäädäntö ja EY:n tuomioistuimen oikeuskäytäntö; unionin toimialalla hyväksytyt julistukset ja päätöslauselmat; yhteiseen ulko- ja turvallisuuspolitiikkaan liittyvät asiakirjat; oikeus- ja sisäasioita koskevat asiakirjat sekä yhteisön tekemät kansainväliset sopimukset sekä jäsenvaltioiden unioniasioissa keskenään tekemät sopimukset. Varsinaisen yhteisön oikeuden lisäksi *acquis communautaire* sisältää kaikki unionin toista ja kolmatta pilaria koskevat säädökset ja perussopimuksissa vahvistetut yhteiset tavoitteet. (Europa-sanasto, 2010).

Aineistoni edustaa siis säädöskieltä. DGT:n julkaisemassa käännösmuistissa on tekstiä 22 kielellä. Kaikkien unionin jäsenyyttä hakevien valtioiden on hyväksyttävä *acquis communautaire* ennen liittymistään, siirrettävä se osaksi kansallista lainsäädäntöään ja sovellettava sitä heti liittymisestään alkaen. Tämän vuoksi *acquis communautaire* on myös käännetty kokonaisuudessaan kaikille EU:n virallisille kielille irlantia lukuun ottamatta. (Yhteinen tutkimuskeskus 2008.)

Ennen kuin EurLex-asiakirja päätyi kohdistuksessa käytettyyn ohjelmaan nimeltä *Euramis Alignment*, sitä oli prosessoitu etukäteen. Tämän esiprosessoinnin yhteydessä asiakirjoista

poistettiin muun muassa turhia välilyöntejä ja rivinvaihtoja.<sup>6</sup> Julkaistun käännösmuistin kohdistuksia on osin korjattu käsin hyödyntämällä ohjelmaa *Euramis alignment editor*, mikä parantaa käännösmuistin kohdistusten tarkkuutta. (Yhteinen tutkimuskeskus 2008.)

## 5.2 Tutkimuskorpusten kokoaminen

Korpusten kokoamisen teknisestä toteutuksesta vastasi Tampereen yliopiston monikielisen viestinnän ja käännöstieteen lehtori, tohtori Mikhail Mikhailov.<sup>7</sup> Laajasta monikielisestä kohdistettujen lauseiden (tarkemmin sanottuna käännössegmenttien) kokoelmasta on mahdollista muodostaa kaksikielisiä rinnakkaiskorpuksia. DGT:n käännösmuistissa teksti on tallennettuna nimenomaan segmentteinä, joten kokonaisten tekstien palauttaminen näkyviin ei onnistu. DGT:n edustajan Anabela Nevesin mukaan DGT ei takaa, että tekstit olisivat muistissa kokonaisuudessaan. Hän huomauttaa, että segmenttejä on voitu poistaa monista erisyistä. DGT ei myöskään takaa, että segmenttien keskinäinen järjestys olisi sama kuin teksteissä, joihin ne liittyvät. (Neves 2010.) Oman silmäilevän tarkastelun perusteella vaikuttaa kuitenkin siltä, että segmenttien järjestys on karkeasti ottaen säilynyt, joten TACT-käyttöliittymän (*Tampere University Corpus Tools*) avulla on mahdollista laajentaa kontekstia useamman lauseen pituiseksi. Kukin segmentti sisältää lisäksi alkuperäisen asiakirjan EurLex-numeron, jolloin asiakirja on löydettävissä kokonaisuudessaankin pienellä vaivannäöllä esimerkiksi internetissä olevasta EurLex-tietokannasta. (Yhteinen tutkimuskeskus 2008.) Tämä osoittautui hyödylliseksi ominaisuudeksi aineiston analyysin vaiheessa.

Varsinainen tutkimuskorpusten kokoaminen tapahtui seuraavasti. Ensin Yhteisen tutkimuskeskuksen internet-sivulta<sup>8</sup> ladattiin 12 zip-tiedostoa, joista kukin oli kooltaan noin 100 megatavua. Jokainen niistä sisälsi kymmeniä tmx-tiedostoja, jotka oli nimetty EurLex-numeron mukaan. Lisäksi oli ladattava *TMXtract*-purkuohjelma ja kopioitava se samaan kansioon zip-tiedostojen kanssa. Seuraavaksi *TMXtractilla* valittiin kieliparit en–fi ja en–de, jonka jälkeen ohjelma haki kyseisillä kielipareilla saatavissa olevat käännösyksiköt zip-tiedostoista ja tallensi ne tmx-tiedostoon. Lopuksi näin syntyneen tmx-tiedoston sisältämät elementit siirrettiin php-skriptillä Tampereen yliopiston kieli- ja käännöstieteiden laitoksen palvelin *mustikkaan*, joka sijaitsee osoitteessa <https://mustikka.uta.fi/corpora/>. Näihin

<sup>6</sup> ks. tarkemmin [http://langtech.jrc.it/Documents/DGT-TM\\_EUR-LEX-preprocessing.pdf](http://langtech.jrc.it/Documents/DGT-TM_EUR-LEX-preprocessing.pdf)

<sup>7</sup> Haluan tässä yhteydessä kiittää häntä sydämellisesti korpuksiin liittyneestä avusta ja ohjauksesta.

<sup>8</sup> [http://wt.jrc.it/lt/Acquis/DGT\\_TU\\_1.0/data/](http://wt.jrc.it/lt/Acquis/DGT_TU_1.0/data/)

PostgreSQL-tietokannan muodossa oleviin tutkimuskorpuksiin pääsee käyttäjätunnuksella internetin kautta TACT-käyttöliittymän avulla (ohjelmoitu PHP-kielillä). Englanti–suomi-korpuksen nimi on *DGT TM Acquis (en–fi)* ja englanti–saksa-korpuksen *DGT TM Acquis (en–de)*.

### **5.3 Tilastotietoja tutkimuskorpuksesta**

Seuraavaksi esittelen tutkimuskorpuksiani koskevia tilastotietoja, jotka on saatu TACT-käyttöliittymästä. Käytän alla käyttöliittymän omia termejä, joista osa kaipaa hieman selvennystä. Irtotekstillä tarkoitetaan erillistä asiakirjaa, jolla on oma EurLex-numeronsa ja joiden määrä siis kertoo sen, kuinka monesta eri asiakirjasta kunkin korpuksen segmentit ovat peräisin. Sanalla tarkoitetaan tässä yhteydessä sanetta.

#### **DGT TM Acquis Corpus (en–de)**

-3 117 irtotekstiä kummallakin kielellä

-sanojen määrä: yhteensä 19 976 277, en: 10 620 764, de: 9 355 513

-kirjainten määrä: yhteensä 118 991 778, en: 56 267 591, de: 62 724 187

#### **DGT TM Acquis Corpus (en–fi)**

-1 764 irtotekstiä kummallakin kielellä

-sanojen määrä: yhteensä 16 782 105, en: 9 981 904, fi: 6 800 201

-kirjainten määrä: yhteensä: 108 841 959, en: 52 880 406, fi: 55 961 553

Näistä tilastotiedoista näkyy suomen- ja saksankielisten sanojen verrattainen pituus englanninkielisiin sanoihin nähden, sillä molemmissa tutkimuskorpuksissa englanninkielisiä sanoja on lukumääräisesti enemmän, mutta suomen- ja saksankielisissä osioissa on kuitenkin enemmän kirjaimia. Segmenttien määrän tutkimuskorpuksissani voi päätellä seuraavasti: DGT:n käännosmuistissa on kokonaisuudessaan 2 187 504 englanninkielistä segmenttiä, 514 868 suomenkielistä segmenttiä ja 532 668 saksankielistä segmenttiä (Yhteinen tutkimuskeskus 2008.). Jos oletetaan, että kaikki suomen- ja saksankieliset segmentit löytyvät myös englanniksi, segmenttien määräksi voi arvioida molempien tutkimuskorpusten osalta yli 500 000.

### **5.4 Aineiston sopivuuden arviointia**

Omat tutkimuskorpukseni ovat tarkasti ottaen käännosmuistikokoelmia eivätkä varsinaisia korpuksia korpuksen tiukimman tieteellisen määritelmän mukaan. Niitä voi myös nimittää luonnollisiksi korpuksiksi, sillä ne ovat syntyneet tosielämän tilanteessa muuta kuin

tutkimustarkoitusta varten. Vaikka korpuksen määritelmää onkin pyritty rajaamaan sen alkuaikojen määritelmästä (”mikä tahansa tekstien kokoelma”), määritelmät ja kokoamisperiaatteet ovat kuitenkin edelleen epäyhtenäisiä, mikä kävi ilmi jo erilaisten korpustyyppien ja korpusten esittelyn yhteydessä.

Nyt onkin tärkeää pohtia, miten hyvin tutkimusaineistoni soveltuu juuri omiin tutkimuskysymyksiini vastaamiseen. Tutkimuksessani olen kiinnostunut selvittämään, missä merkityksessä englanninkielistä *manage*-verbiä käytetään saksan- ja suomenkielisissä EU-teksteissä. Lisäksi haluan tietää, näyttäisivätkö *manage*-verbin vastineet määräytyvän sillä ilmaistun tekemisen kohteen mukaan. Koska olen kiinnostunut nimenomaan EU-teksteistä enkä yleiskielestä, tutkimusaineistoni on tekstilajin suhteen ihanteellisesti rajattu. Komission lisäksi EU-tekstejä syntyy toki muuallakin, esimerkiksi parlamentissa, mutta komission käännöstoimen pääosasto on maailman suurin kielipalvelu (Ulkoasiainministeriö 2010), joten sen tuottamien käännösten voidaan ajatella edustavan suhteellisen hyvin EU-kieltä ainakin säädöskielen osalta. Lisäksi DGT:n julkaisema monikielinen käännösmuisti soveltuu hyvin tutkimukseni aineiston raaka-aineeksi, sillä se tarjoaa komission tekstejä kaikissa niissä kielissä, joiden tutkimisesta olen kiinnostunut. Käännösmuistista muodostetun korpuksen suuri kokokin puoltaa sen valintaa tutkimusaineistokseni. Tutkimuskysymykseeni vastaamisen kannalta on nimittäin olennaisen tärkeää, että saan mahdollisimman paljon konkreettisia käyttöesimerkkejä. Näin laajan korpuksen kerääminen manuaalisesti olisi kuitenkin ollut liian suuri urakka gradun puitteissa tehtäväksi: pienimmässäkin osakorpuksessa (suomenkieliset segmentit) sanoja on reilusti yli kuusi miljoonaa. Käännössegmentit ovat myös huolellisesti kohdistettuja, sillä ne ovat joko syntyneet käännösprosessin tuloksena tai kohdistuksia on DGT:n antamien tietojen mukaan kohdistusohjelman käytön jälkeen myös manuaalisesti korjattu. Rinnakkaiskorpusten arvo määräytyykin pitkälti niiden kohdistusten laadun mukaan (McEnery & Xiao 2008, 20–21). Lisäksi tutkimuskysymyksen kannalta on hyödyllistä, että kirjoittajia on useita, jolloin yhden ihmisen kielikäsitys ei täysin sanele tuloksia. Toisaalta säädösteksteistä koostuvassa aineistossa kirjoittajien joukko on myös tiedossa – lainsäädäntöä kääntävät nimittäin periaatteessa ainoastaan komission in-house-kääntäjät (Neves 2000). Joukon rajallisuuden mahdolliset vaikutukset otan vielä pohdintaan loppupäätelmissä. Asiakirjat sisältävät tiedon niiden julkaisuvuodesta, mikä auttaa hahmottamaan tekstien syntyajankohtaa. Tosin osa teksteistä on julkaistu esimerkiksi jo ennen Suomen liittymistä Euroopan unioniin, jolloin

kääntämisaikajaksoksi on hankalampi saada tarkasti selville, sillä ne on voitu kääntää vasta huomattavan monta vuotta niiden alkutekstien julkaisemisen jälkeen.

Käyttämääni tutkimusaineistoon sisältyy kuitenkin myös muutamia varauksia. Tekstien alkukieli ei ole tiedossa, joten esimerkiksi englanninkielisessä aineistossa voi olla sekä alun perin (euro)englanniksi kirjoitettuja asiakirjoja että jostakin muusta unionin työkielestä tai virallisesta kielestä käännettyjä asiakirjoja. Sama pätee tietysti myös suomeen ja saksaan. Segmenttien kohdistus ei siis ole tapahtunut lähde- ja kohdekielen välillä, vaan eri kieliversioiden välillä. Yhteisestä tutkimuskeskuksesta kuitenkin todetaan, että monet asiakirjoista on alun perin kirjoitettu englanniksi ja sen jälkeen käännetty muille kielille (Yhteinen tutkimuskeskus 2008). Tämä on totta etenkin uusimpien asiakirjojen osalta, sillä Englanti dominoi lähdekielten jakaumaa yhä selvemmin. Vielä vuonna 1997 alkuteksteistä laadittiin 45,4 prosenttia englanniksi ja 40,4 prosenttia ranskaksi, 5,4 prosenttia saksaksi ja 8,8 prosenttia muilla kielillä. Vuonna 2008 vastaavat prosenttiluvut olivat jo: 72,5 prosenttia englanniksi, 11,8 prosenttia ranskaksi, 2,7 prosenttia saksaksi ja 13 prosenttia muilla kielillä. (Euroopan komissio 2009e, 6.) En siis voi suoranaisesti tutkia englanninkielisten sanojen suomennoksia ja saksannoksia. Euroopan komission tekstien tulee kuitenkin olla rinnakkaisversioita samoista teksteistä, joten niiden tavoitteena on ilmaista sama asia kaikilla kielillä. Näin ollen voin tutkia, miten englanniksi *manage*-verbillä ilmaistu asia on ilmaistu suomeksi ja saksaksi. Nimitänkin analyysivaiheessa näitä suomen- ja saksankielisiä sanoja suomennosten ja saksannosten sijaan suomen- ja saksankielisiksi vastineiksi. Analyysin yhteydessä on myös syytä pitää mielessä, että tutkin kansainvälisessä yhteisössä syntyneitä englanninkielisiä tekstejä, jotka eivät aina ole äidinkielenään Englantia puhuvien kirjoittamia (tästä enemmän luvuissa ”Toimintakontekstin kuvaus” sekä ”Loppupäätelmät”).

Käännösmuistin perusluonteeseen liittyvä ominaisuus on se, että täysin samanlaiset segmentit esiintyvät muistissa ainoastaan kerran eli yhtenä segmenttinä<sup>9</sup>. DGT:n käännösmuistissa on kuitenkin jonkin verran toistoa (esim. ”Article 1” ja sitä vastaavat suomen- ja saksankieliset segmentit toistuvat useaan kertaan). Tämän ilmiön syntymekanismiin liittyy se, että tietokannassa on ilmeisesti myös jälkikäteen kohdistettuja asiakirjoja, jotka eivät siis ole päätyneet tietokantaan käännösmuistiohjelman kautta. Lisäksi tiedän omasta kokemuksestani komission harjoittelijana, että kaikkia käännettävän asiakirjan osia ei välttämättä käännetä

---

<sup>9</sup> Perustietoja Tradoksesta: <http://www.trados.com/en/language-service-providers/translation-memory.asp#anchor2>

muistin avulla, mikä aiheuttaa omalta osaltaan sen, etteivät käännösmuistissa olevat asiakirjat välttämättä ole täydellisiä. Nämä aineiston ominaisuudet eivät kuitenkaan suoranaisesti vaikuta omaan tutkimukseeni, sillä tavoitteena oli saada mahdollisimman paljon konkreettisia *manage*-verbin englannin-, suomen- ja saksankielisiä käyttöesimerkkejä, ja niitä löytyi aineiston rajoituksista huolimatta tarpeeksi. Täysin samanlaiset segmentit eivät myöskään toisi tutkimukseeni lisäarvoa. Aineiston tekstien epätäydellisyys vaikuttaa esimerkiksi siihen, ettei aineistosta voi laskea luotettavasti frekvenssejä. Tämä ei kuitenkaan ole olennaista omassa tutkimuksessani, sillä pystyn kuitenkin käsittelemään aineistosta saamiani *hakutuloksia* myös kvantitatiivisesti eli laskemaan todellisissa käännösyksiköissä esiintyneiden kielellisten piirteiden määriä sekä laskemaan niistä prosenttiosuuksia. Laskelmissa täytyy tietenkin ottaa huomioon se, että prosenttiosuuksien laskeminen kovin pienistä joukoista voi vääristää kuvaa todellisuudesta, mutta pyrin huomioimaan tämän analyysissa.

*TMXtract* muodostaa kaksikielisiä tmx-tiedostoja, ja vaikka TACT-käyttöliittymässä olisi periaatteessa ollut mahdollista rakentaa kolmikielisiä korpuksia, käännösmuistin valtava koko ja lähdetiedostojen kaksikielisyys tekivät sen käytännössä ongelmalliseksi. Tästä johtuen käytössäni on kaksi erillistä korpusta (suomi–englanti ja saksa–englanti), joiden avulla pyrin tutkimuskysymykseeni vastaamaan. Molemmissa korpuksissa on toisena kielenä englanti, jonka avulla pyrin löytämään yhteyden kahden erillisen korpuksen välille. Vertailemalla merkitysryhmien osuuksia eri aineistoissa pyrin myös varmistamaan, että molemmat aineistot koostuvat lopulta samoja aihepiirejä käsittelevistä teksteistä. Jos näin on, siitä voi varovaisesti päätellä myös sen, että aineistot sisältävät suurelta osin samoja englanninkielisiä tekstejä. Loppujen lopuksi käyttämäni aineisto antaa mielestäni riittävästi todellisia käyttöesimerkkejä nimenomaan siitä tekstiaineksesta, jota haluankin tutkia. Seuraavassa luvussa kerron tarkemmin analyysin eri vaiheista ja niiden tuloksista.

## 6. *Manage*-verbi suomeksi ja saksaksi

### 6.1 *Analyysin valmistelu*

TACT-käyttöliittymässä on erilaisia korpustyökaluja, joiden avulla laajoista tutkimuskorpuksista on mahdollista eritellä osia, jotka ovat tutkimuskysymyksen kannalta kiinnostavia. Korpustyökaluihin lukeutuvat frekvenssilistojen, yksi- ja kaksikielisen konkordanssin, kollokaatioiden ja sanatilastojen hakumahdollisuus. Lisäksi käyttöliittymä tuo kätevästi esiin kussakin korpuksessa olevan tekstien luettelon sekä muita tilastotietoja korpuksista, joita esittelin omien korpusteni osalta luvussa ”Tutkimuskorpusten kuvaus”.

Omaan tutkimuskysymykseeni vastaamisen kannalta oli mielekästä hyödyntää kaksikielistä konkordanssitoimintoa, koska sen avulla sain esille ne korpusten sisältämät käännösyksiköt, joissa käytettiin tarkastelemaani *manage*-verbiä sekä sen suomen- ja saksankielisiä vastineita. Käännösyksiköllä tarkoitan tässä sellaista segmenttiparia, joka sisältää sekä englannin- että saksan- tai suomenkielisen segmentin. Segmentillä puolestaan tarkoitan sellaista tekstikatkelmaa, joka on tallentunut käännösmuistiin omana erillisenä kokonaisuutenaan. Usein nämä segmentit ovat noin yhden lauseen mittaisia. Tekemäni haku oli samanlainen sekä englanti–suomi- että englanti–saksa-korpusten osalta. Hakusana oli *manage*, jonka määritin voivan olla ”mikä tahansa sanan osa”. Sanaluokaksi (*part of speech*) määritin verbin (V). Tämä haku tuotti englanti–suomi-korpuksesta 240 osumaa ja englanti–saksa-korpuksesta 198 osumaa. Englanninkielinen *manage* esiintyi hakutuloksissa pääasiassa finiittimuotoisena verbinä ja joitakin kertoja infiniittimuotoisena *ed*-partisiippina, mutta en erotellut näitä verbimuotoja toisistaan. Tarkistin vielä kyseisen haun ulkopuolelle jääneet *manage*-verbin *ing*-partisiipit (hakusana *managing*, sanaluokka V), mutta niitä ei aineistoista hakuohjelman avulla löytynyt. Osumat tallensin Excel-tiedostoihin, joista käytän jatkossa lyhenteitä en–fi ja en–de.

Ennen kuin aineistoa saattoi alkaa varsinaisesti analysoida, sitä oli käsiteltävä monin tavoin. Aivan aluksi etsin *manage*-verbin vastineet sekä suomen- että saksankielisistä segmenteistä. Vastineen tai käännösvastineen käsitettä ei ole tämän työn puitteissa tarkoituksenmukaista problematisoida, vaan tarkoitan sillä tässä yhteydessä sellaista lekseemiä tai sellaisia lekseemejä, joiden tulkitsin ilmaisevan englannin *manage*-verbiä vastaavan sisällön suomen- ja saksankielisissä segmenteissä. Suurimmaksi osaksi vastineet oli melko yksinkertaista paikantaa, mutta joissakin tapauksissa toinen tutkija olisi saattanut päätyä nimeämään

esimerkiksi eripituisen lekseemijonon *manage*-verbin vastineeksi. Kirjasin vastineet (verbien osalta a-infiniitivissä) omaan sarakkeeseensa aina kyseisen käännösyksikön kanssa samalle riville. Vastinesarakkeen nimeksi tuli en-fi-tiedostossa ”suomennos” ja en-de-tiedostossa vastaavasti ”saksannos”. Samalla poimin omaan sarakkeeseensa (”objekti (en)”) *manage*-verbin saamat englanninkieliset (suorat) objektit (esim. *to manage contracts*). Objekteihin rinnastan tässä tutkimuksessa myös adjektiivinomaisesti käytettyjen partisiippien pääsanat eli ne sanat, joihin partisiipit viittaavat (esim. *ports managed pursuant to article 6*, jossa *managed pursuant to article 6* on adjektiivina käytetty partisiippilauseke ja *ports* sen pääsana) (Online Writing Lab 2010). Näitä tapauksia ei ollut kovin monta. Lisäksi hyvin pienessä määrässä segmenttejä *manage*-verbiä seurasi perusmuotoinen verbi ja sen objekti, jolloin kirjasin molemmat objektisarakkeeseen (esim. *manage to find new markets*). Nämä valinnat perustelen sillä, että oikeastaan haluan tutkia paremminkin niitä *kohteita*, joihin *manage*-verbillä ilmaistu tekeminen kohdistuu, enkä niinkään välttämättä objekteja kieliopillisena luokkana. Kohteilla tarkoitan niitä asioita ja ilmiöitä, joita suomeksi esim. hallinnoidaan tai hoidetaan. Jatkossa puhun ainoastaan objekteista, mutta tarkoitan niillä myös edellä kuvattuja tapauksia. Objektit pyrin pitämään mahdollisimman lyhyinä. Päädyinkin kirjaamaan ainoastaan objektilausekkeen pääsanana (esim. *fishery resources*), jotta objektit voisi myöhemmin järjestää aakkosjärjestykseen pääsanansa mukaan (*resources*). Sulkuihin kirjasin sitten sekä pääsanana että siihen välittömästi liittyvän määritteen tai määritteet analyysin helpottamiseksi. Kirjasin objektit siten, että niistä näkyi, oliko objekti yksikössä vai monikossa, mikä ei kuitenkaan lopulta osoittautunut oleelliseksi aineistoa analysoitaessa. Kolmas asia, jonka kirjasin omaan sarakkeeseensa aineiston ensimmäisellä käsittelykierroksella, oli englanninkielistä objektia vastaava suomen- tai saksankielinen lekseemi- tai lekseemiryhmä. Nämä kirjasin sarakkeisiin ”en objektia vastaava fi” ja ”en objektia vastaava de”. Englanninkielisiä objekteja saattoi suomessa ja saksassa vastata esimerkiksi yhdyssanan alkuosa (*to manage a crisis* → *kriisinhallinta*), mistä johtuen sarakkeen otsikoksi ei soveltunut englannin sarakkeen otsikointia noudatteleva ”objekti de/fi”. Tämän sarakkeen sanat ovat kuitenkin niitä suomen- ja saksankielisten sanojen kollokaatteja, jotka tavallaan tulivat mutkan kautta luokitelleeksi merkitysryhmiin samalla, kun luokittelin englanninkieliset objektit.

Aineiston toisella käsittelykierroksella tarkistin vastineet ja poistin turhat kahdennukset. Kahdennuksilla tarkoitan sellaisia käännösyksiköitä, jotka olivat keskenään (lähes täysin) identtisiä ja peräisin samasta EurLex-asiakirjasta. Ainakin osa näistä oli syntynyt siten, että



sama segmentti oli käännetty jostain syystä kahteen kertaan, mutta sitä oli toisella kerralla alettu kääntää hieman eri kohdasta, jolloin alusta puuttui muutama sana. Tarkistin jokaisen kahdennuksen kohdalla, että kyseinen segmentti esiintyi varsinaisessa asiakirjassa vain kerran ja jos näin oli, poistin niistä toisen. Jos keskenään identtiset käännösyksiköt olivat kuitenkin peräisin eri EurLex-asiakirjoista, säilytin ne aineistossa. Turhien kahdennusten lisäksi poistin yhden en–fi-käännösyksikön ja kaksi en–de-käännösyksikköä, joissa molemmat segmentit olivat englanninkielisiä. Näissä tapauksissa se johtui siitä, että *manage* oli osa suomentamattoman ja saksantamattoman tietokoneohjelman komentoa. Havaintojeni mukaan kolmessa en–fi-tiedoston ja samoin kolmessa en–de-tiedoston käännösyksikössä *manage* esiintyi kahdesti samassa englanninkielisessä segmentissä, mutta käyttöliittymä oli poiminut käännösyksikön aineistoon ainoastaan yhtenä kappaleena. Tällöin etsin vastineet ja objektin sen sanan mukaisesti, jonka perusteella ohjelma oli poiminut käännösyksikön mukaan aineistoon (merkitty alaviivoin: *\_manage\_*). En katsonut aiheelliseksi kahdentaa kyseisiä yksiköitä niiden pienen määrän vuoksi ja siksi, etten lopulta voi olla täysin varma, että huomasin kaikki puuttuvat yksiköt. Tältä osin toimin siis käyttämäni korpustyökalun suomien mahdollisuuksien ja rajoitusten puitteissa. Turhien käännösyksiköiden poistamisen jälkeen en–fi-tiedostoon jäi 204 ja en–de-tiedostoon 170 käännösyksikköä.

Aineiston siistimisen ja vastineiden tarkistamisen jälkeen liitin vastineisiin tarkemman kieliopillisen analyysin: jos vastine ei ollut verbi, merkitsin sen perään kyseessä olevan sanaluokan. Lisäksi päädyin nimeämään erikseen verbinmuotoihin lukeutuvat partisiippimuodot, koska ne käyttäytyvät osin adjektiivin tavoin. Niitä oli sekä saksan- että suomenkielisissä vastineissa ja ne tuntuivat erottuvan ominaisuuksiltaan muista *manage*-verbin vastineina olleista verbinmuodoista, joiden muodostamisessa rakenteissa oli selvemmin samantyyppinen rakenne kuin englanninkielisissä segmenteissä (esimerkiksi en–fi-aineistossa: *organization which manages horses for competition or racing -> kilpailu- tai urheiluhevostia hallinnoiva yhdistys* (va-partisiippi) vs. *manage contracts -> hallinnoida sopimuksia* (muu verbinmuoto) ja en–de-aineistossa *the resources it manages -> von ihr verwalteten Mittel* (partisiippi II) vs. *manage a register -> verwalten ein Verzeichnis*) (muu verbinmuoto). Suomenkielisen aineiston kieliopillisen analyysin apuna käytin deskriptiivistä kielioppiteosta *Iso Suomen kielioppi* (Hakulinen ym. 2004a) sekä sen verkkoversiota (Hakulinen ym. 2004b). Englanninkielisen aineiston hahmottamisessa oli apuna teos *A Comprehensive Grammar of the English Language* (Quirk ym. 1985) ja saksankielisen aineiston analyysissa auttoi IDS Mannheimin verkkokielioppi *Grammis* (2010). En–fi-

tiedostossa vastineiden joukosta löytyi verbien lisäksi substantiiveja (”subst.”) ja adjektiiveja (”adj.”) sekä verbimuotoihin lukeutuvia partisiippeja (”va-partisiippi”, ”agenttipartisiippi”, ”tu-partisiippi”). En–de-tiedostossa oli verbin lisäksi vastineina substantiiveja (subst.) ja verbimuotoihin lukeutuvia partisiippeja (”partisiippi II”). En–fi-tiedostossa viidessä käännösyksikössä tilanne oli se, ettei *manage*-verbille voinut määrittää minkäänlaista vastinetta. Tämä johtui joko hyvin erilaisesta muotoilusta (kahdessa tapauksessa) tai siitä, että suomenkielisestä segmentistä puuttui osa englanninkielisen segmentin sisältämästä ajatussisällöstä (kolmessa tapauksessa). En–de-tiedostossa vain kolmen käännösyksikön osalta oli mahdoton määrittää vastinetta, sillä ne oli muotoiltu niin erilaista rakennetta käyttäen.

Toiseksi viimeinen vaihe aineiston käsittelyssä varsinaista analyysia varten oli se, että pyrin ”palauttamaan” kaikki suomen- ja saksankieliset vastineet verbeiksi. Palauttamisesta puhuessani tarkoitan muihin sanaluokkiin kuin verbeihin (yleensä substantiiveihin) kuuluvien vastineiden muuntamista niihin läheisesti liittyviksi verbeiksi, en niinkään halua vihjata, että ajattelisin kaikkien substantiivien saaneen alkunsa verbeistä. Tein tämän siksi, että suurin osa vastineista oli jo alun perinkin verbejä ja verbimuotoiseksi muunnettuna sain vertailuun mukaan myös loput vastineet. Lisäksi minua kiinnosti ryhmitellä vastineet niiden merkityksen mukaan ja ajattelin pintarakenteen alla piilevän merkityksen olevan samantyyppinen esimerkiksi suomen substantiivilla *hallinto* ja verbillä *hallinnoida* sekä saksan substantiivilla *Verwaltung* ja verbillä *verwalten*. Verbimuotoon palauttaminen helpotti vastineiden ryhmittelyä ja näin ollen myös niiden myöhempää vertailua. Verbimuotoon palauttaminen tuntui myös sikäli luontevalta, että myös tutkimani englanninkielinen *manage* on verbi. Viimeinen vaihe analyysin esivalmistelussa oli englanninkielisten objektien luokittelu merkitysryhmiin semanttisen preferenssin tutkimista varten. Loin omaan tutkimukseeni sopivat merkitysryhmät tutkimusaineistoni pohjalta eli korpusvetoisesti, ja näin hahmottui yhdeksän merkitysryhmää, joista neljään tuli lisäksi alaryhmiä. Näiden yhdeksän ryhmän lisäksi turvauduin vielä kaatoluokkaan ”muu” sellaisissa tapauksissa, jotka vaikuttivat yksittäisesiintymiltä ja joiden en katsonut sopivan yhteenkään laajemman joukon muodostamaan ryhmään. Yhteensä näin muodostui 16 merkitysryhmää. Koska objektit olivat lähes poikkeuksetta substantiiveja, tuntui luontevalta jaotella ne merkitysryhmiin jonkinlaisia yläkäsitteitä etsimällä. Luokittelussa otin huomioon koko objektin enkä vain sen pääsanaa, jonka nostin esiin lähinnä aakkostamistarkoituksessa. Monien objektien merkitys tuli esille vasta, kun objektia tarkasteli kokonaisuudessaan (esim. *fishery resources*). Joissakin

tapauksissa turvauduin vielä segmentin tai jopa koko asiakirjan antamaan laajempaan kontekstiin, mutta nämä tapaukset olivat vähäisiä ja mainitsen ne erikseen seuraavassa, kun esittelen käyttämäni merkitysryhmät ja kerron tarkemmin niiden sisällöistä. Laajempi kontekstin hyödyntäminen oli tarpeen esimerkiksi silloin, kun täytyi ratkaista, kuuluuko erilaisia varantoja kuvaava *resources* varallisuuden (tekstin aiheena pankkiala) vai konkreettisten luonnonvarojen (tekstin aiheena kalastus) merkitysryhmään. Seuraavassa esittelen merkitysryhmien sisällön ja luokitteluperusteet hieman tarkemmin sekä annan niihin luokittelemistani sanoista muutaman esimerkin.

### **järjestelmä**

Merkitysryhmään ”järjestelmä” kuuluu sanoja, jotka merkitsevät erilaisia poliittisia ohjelmia (*Community programme*), abstrakteja talouden rakennelmia (*regulated market*), toimenpide- ja tukijärjestelmiä (*aid scheme*) ja yleisemmin erilaisia yhteiskunnan toimintaan liittyviä järjestelmiä (*emergency communication and information system*).

### **kielteinen ilmiö**

Kielteiset ilmiöt nousivat selvästi esiin aineistosta. Tämän laajan ryhmän sisällä oli havaittavissa melko selvästi kolmentyyppisiä kielteisiä ilmiöitä: epävarmuuksia (*risk, exposure*), haittoja (*impacts, noise issues and effects*) ja ristiriitoja (*conflict, crisis*). Pääsanansa puolesta seuraukset (*consequences*) ei vielä ole aivan selvästi kielteinen, mutta viimeistään koko objektin (mahdolliset haittavaikutukset eli englanniksi *potential adverse consequences*) huomioiminen selventää tilannetta ja sijoittaa kyseisen objektin kielteisten ilmiöiden ryhmään. Pääsanansa puolesta melko neutraali *challenges* (suom. ’haasteet’) sai kontekstissa kuitenkin hieman kielteisen sävyn, minkä voi päätellä jo katsomalla objektia kokonaisuudessaan: *challenges of globalisation*. Nämä haasteet aiheuttavat vähintäänkin tiettyä tulevaisuuteen liittyviä epävarmuutta, minkä vuoksi katsoin perustelluksi luokitella objektin alaryhmään ”kielteinen ilmiö / epävarmuus”.

### **konkreettinen asia**

Erilaiset konkreettiset asiat olivat myös oma ryhmänsä, joka tuntui kuitenkin luonteelta erotella kahteen alaryhmään. Näitä olivat ihmisen toiminnan tulokset (*radioactive waste, ports*), jotka nimesin alaryhmäksi ”konkreettinen asia / ihmisen luomus” ja toisaalta luontoon liittyvät asiat (*forests, fish stocks*), joiden alaryhmän nimeksi tuli ”konkreettinen asia / luonnonvara”. Tiedetyt luontoon liittyvät objektit, kuten kilpahevokset, ilmatilan,

tietokoneohjelman käyttäjät ja oliivitarhat luokittelin ihmisen luomuksiksi, koska ne ovat saaneet varsinaisen asiakirjoissa tarkoitetun merkityksensä vasta, kun ihminen on alkanut hyödyntää niitä. Hevoset, ilma, ihmiset ja oliivipuut kuuluisivat luonnonvaroihin, mutta nämä tietyn roolin saaneet tai tiettyyn tarkoitukseen valjastetut luonnosta lähtöisin olevat tarkoitteet eivät. Luokittelin konkreettisiksi asioiksi myös nykyteknologiaan liittyvät lisäominaisuuksia ja aluetunnusta tarkoittavat objektit *special features, top level domain*. Näiden konkreettisuudesta voi olla monta mieltä, mutta tutkimuksen kannalta ne sopivat mielestäni tarpeeksi hyvin tähän merkitysryhmään.

### **oikeus**

Tämä merkitysryhmä voisi periaatteessa olla järjestelmien merkitysryhmän alaryhmä, sillä sen sisältämät sanat liittyvät usein johonkin järjestelmään, kuten talous- tai liikennejärjestelmiin. Katsoin kuitenkin parhaaksi erottaa ne omaksi ryhmäkseen, sillä ne muodostivat keskenään melko selkeän merkitysryhmän. Tämän ryhmän sanat tarkoittavat jonkinlaista lupaa (*permit*), oikeutta (*right*), ulkopuolisen toteamaa kelpoisuutta (*airworthiness*) tai kiintiötä (*quota*).

### **tiedot**

Merkitysryhmään ”tiedot” kuuluu erilaisia rekistereitä (*studbook*), luetteloita (*list*), koodeja (*waiting codes*) sekä muita tietoja (*fisheries data, information*). Nämä muodostivat luontevasti oman merkitysryhmänsä.

### **toiminta**

Toiminnaksi luokiteltavia objekteja oli runsaasti ja niiden luonne jakautui kolmeen alaryhmään: hallinnoivan (jne.) tahon näkökulmasta katsottuna ulkopuolisten eli muiden tekemään toimintaan (*fishing activities, fishing efforts*), hallinnoivan tahon omaan toimintaan (*project, operations relating to registration*) sekä yrityksen toimintaan (*undertaking's affairs*). Erotin yrityksen toiminnan kahdesta ensimmäisestä merkitysryhmästä, koska se liittyi niin selkeästi yritysmaailmaan ja muodosti laajan joukon. Oman toiminnan ryhmä voi luonnollisesti liittyä myös yrityksen toimintaan, mutta se on käsitettävä kuitenkin paljon yleisemmin toimintaa luonnehtivaksi. Esimerkiksi hankkeita ja kirjaamista koskevia tehtäviä kun voi olla myös poliittisilla elimillä, kansalaisjärjestöillä ynnä muilla sellaisilla.

### **varallisuus**

Tähän merkitysryhmään luin kuuluvaksi monenlaiset rahaan liittyvät asiat, kuten sijoitukset, osuudet, raha- ja muut varat sekä rahastot (*investments, holdings, monies and other assets, funds, trusts*). Tämän merkitysryhmän yhteydessä huomasin, että koko objektin huomioiminen on tärkeää. Jos objektin pääsana oli *holding*, luokittelin sen yleensä varallisuudeksi, mutta yhdistelmä *agricultural holding* viittaakin maatalousyrittäjyysryhmään, joten se kuului näin ollen yritysten ryhmään. Tähän merkitysryhmään laskin kuuluvaksi myös ”negatiivisen” varallisuuden kuten velat ja kulut (*public debt, expenditure*). Rahastojen hallinnoinnin olisi varmasti voinut rinnastaa myös yritysten johtamiseen, mutta toisaalta tuntui luontevammalta niputtaa kaikenlainen rahavaroihin liittyvä omaksi ryhmäkseen.

### **velvoite**

Erilaiset sopimukset, säännöt ja velvollisuudet (*contract, environmental responsibilities, acquis*) aiheuttavat velvoitteita, joten ne on luokiteltu omaksi merkitysryhmäkseen.

### **yritys**

Merkitysryhmällä ”yritykset” on kaksi alaryhmää: yksityisen sektorin yritykset ja julkishallinnon piiriin kuuluvat yksiköt ja vastaavat, jotka rinnastin yrityksiin, koska nekin muodostavat oman organisaationsa. Paremmin tätä alamerkitysryhmää kuvaa organisaation nimitys, mutta halusin säilyttää myös nimessä yhteyden kyseiseen päämerkitysryhmään. ”Yritys / julkinen” kuvaa siis julkishallinnon organisaatioita (*secretariats, centre*) ja ”yritys / yksityinen” kaupallisia yrityksiä (*investment companies, farm*). Luokittelin objektit *team* ja *personnel* myös yritysten merkitysryhmään, ja koska varsinaisesta EurLex-asiakirjasta tarkistamalla selvisi, että niillä viitattiin näissä yhteyksissä nimenomaan julkishallinnon palveluksessa oleviin ihmisryhmiin, ne päätyivät alaryhmään ”yritys / julkinen”.

### **muu**

Kaatoluokkaan ”muu” päätyivät esimerkiksi maantieteellistä käyttöaluetta tarkoittava *geographical distribution* ja tarjontaa merkitsevä *supply*. Tällaisia luokittelemattomia sanoja oli en-fi-aineistossa kahdeksan (8/204) ja en-de-aineistossa seitsemän (7/170).

## **6.2 Analyysi**

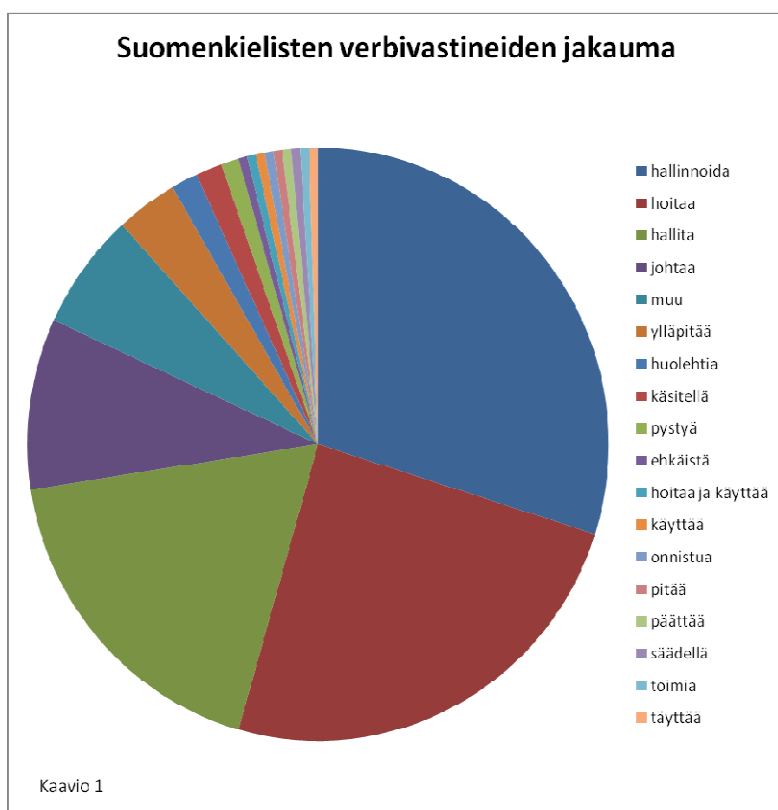
Tavoitteenani oli siis selvittää, mitä ja minkälaisia saksan- ja suomenkielisiä vastineita *manage*-verbi saa Euroopan komission teksteissä. Lisäksi halusin englanninkielisten

objektien merkityksen analysoimisen avulla selvittää, onko *manage*-verbin saaman objektin merkityksellä vaikutusta siihen, miten se suomennetaan tai saksannetaan. Yleensä semanttista preferenssiä tutkitaan yksikielisen tekstin avulla eli analysoidaan valitun hakusanan (esim. *manage*) kanssa esiintyvien sanojen (kollokaatit) merkityksiä ja luokitellaan niitä merkitysryhmiin. Tämän jälkeen katsotaan, onko hakusanalla taipumusta esiintyä yleisemmin jonkin tietyn tai joidenkin tiettyjen merkitysryhmien sanojen kanssa. Hakusanalla voi siis olla semanttista preferenssiä useampaankin merkitysryhmään (Jantunen 2004, 25). Omassa tutkimuksessani sovellan tätä ajatusta kollokaattien merkitysryhmistä ja niiden vaikutuksesta hakusanan merkitykseen kuitenkin niin, että pyrin selvittämään, vaikuttaako polyseemiseksi määrittelemäni *manage*-verbin objektin merkitys siihen, mitä sen alamerkityksistä kulloinkin tarkoitetaan. Yritän päästä tähän ongelmaan käsiksi tutkimalla ammattitaitoisten komission kääntäjien käännostratkaisuja, sillä Euroopan komission tekstien muodostama laaja käännosmuisti tarjoaa ainutlaatuisen mahdollisuuden saada runsaasti käyttöesimerkkejä samantyyppisistä teksteistä englanniksi, suomeksi ja saksaksi. Määrälliseen tutkimukseen kuuluisi tarkkaan ottaen lisäksi tilastollisen merkitsevyyden testaaminen. Pyrin tässä tutkimuksessa kuitenkin vasta alustavasti hahmottamaan mahdollista taipumuksena tai tendenssinä ilmenevää semanttista preferenssiä, joten en ryhdy testaamaan semanttisten preferenssitaipumusten tilastollista merkitsevyyttä. Tuloksiini on siis suhtauduttava suuntaa antavana yleiskatsauksena *manage*-verbin saksantamiseen ja suomentamiseen.

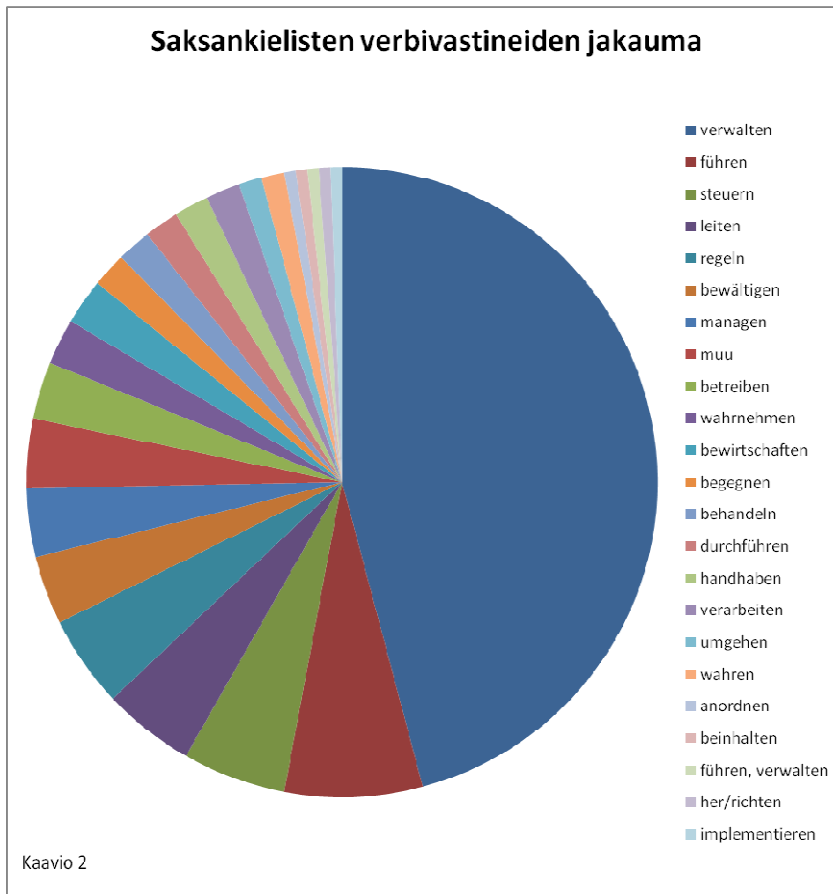
Tulosten jäsentämisessä olen alusta alkaen hyödyntänyt Excel-taulukkolaskentaohjelmaa. Taulukkolaskentaohjelman avulla huolellisesti kirjatusta sarakkeiden sisällöistä (pienikin kirjoitusvirhe tai ylimääräinen välilyönti vaikuttaisi tuloksiin) oli mahdollista erilaisten laskentakaavojen ja kommentojen avulla jäsentää ja järjestellä tutkimusaineistoa. Sekä en–fi-että en–de-tiedostojen ensimmäinen sivu on järjestetty ensin verbimuotoisen vastineen mukaan, sitten englanninkielisen objektin ja lopuksi tuon objektin merkitysryhmän mukaisesti aakkosjärjestykseen. Tämä helpotti kokonaistilanteen hahmottamista ja laskentakaavojen kehittämistä. Aivan ensimmäiseksi laskin Excelin avulla sen, miten verbimuotoon palautetut vastineet jakautuvat en–fi- ja en–de-aineistoissa.

En–fi-aineiston 204 käännoyksikköä sisälsivät yhteensä 17 erilaista suomenkielistä vastinetta ja lisäksi kaatoluokan ”muu”, johon sijoittui 13 vastinetta. En–de-aineiston 170 käännoyksiköstä löytyi 22 saksankielistä vastinetta kaatoluokan ”muu” lisäksi. Molempien aineistojen sisältämä kirjo vaikuttaa suurelta, mikä osaltaan tukee sitä ennako-odotusta, että

verbi *manage* sisältää monenlaisia merkityksiä ja voidaan näin ollen kääntää monella tavalla. Kaavioissa 1 ja 2 näkyvät näiden eri vastineiden jakaumat kielipareittain. Tarkat lukumäärät ja prosenttiosuudet ovat puolestaan liitteissä 1 ja 2. Prosenttiosuudet olen pyöristänyt kahden desimaalin tarkkuudelle tarkastelun helpottamiseksi. Suomenkielisten verbimuotoon palautettujen vastineiden jakaumaa esittävistä ympyrädiagrammista käy hyvin ilmi se, ettei yksikään vastine ole kovin selvästi muita yleisempi. Yleisin vastine *hallinnoida* kattaa 204:stä tapauksesta 61 eli 29,90 prosenttia. Toiseksi yleisin *hoitaa* esiintyy aineistossa 50 kertaa eli 24,51 prosentista tapauksia. Myös kolmanneksi yleisin vastine *hallita* muodostaa verrattain suuren joukon, 37 tapausta ja 18,14 prosenttia. Neljänneksi suurin ryhmä on *johtaa*-verbiin palautettavissa olleet vastineet, jotka nekin kattavat vielä 9,31 prosenttia kaikista suomenkielisistä vastineista 19 esiintymällänsä. Yhteensä nämä neljä suurinta vastineryhmää kattavat 81,86 prosenttia kaikista suomenkielisistä vastineista, mitä voidaan pitää merkittävänä osuutena, kun vertaa sitä jäljelle jäävien 13 verbin ja ”muu”-kategorian yhdessä muodostamaan hieman yli 18 prosenttiin. *Manage*-verbi vaikuttaisi siis saavan pääasiassa neljä eri suomennosta, joiden merkitykset ovat hyvinkin kaukana toisistaan. Suomentajan tekemällä leksikaalisella valinnalla on totisesti merkitystä, sillä onhan hyvin eri asia esimerkiksi johtaa jotakin kuin hoitaa sitä.



Saksankielisten vastineiden osalta tilanne on melko erilainen suomenkielisiin vastineisiin verrattuna. Selvästi suurimman vastineryhmän muodostavat *verwalten*-verbiin palautettavissa olleet vastineet, jotka kattavat 78 esiintymällään peräti 45,88 prosenttia kaikista 170 vastineesta. Loput vastineet vaikuttavat muodostavan hyvinkin pirstaloituneen joukon, sillä seuraavaksi yleisin *manage*-verbin saksankielinen vastine tässä tarkastelussa oli *führen*, joka löytyi vain 12 segmentistä ja muodosti näin ollen 7,06 prosenttia tapauksista. Tätä harvemmin esiintyvien 19 verbivastineen ja luokan ”muu” prosenttiosuudet vaihtelevat yksittäisten esiintymien 0,59 prosentista (5 kertaa) kolmanneksi yleisimmän *steuern*-verbin 9 esiintymän muodostamaan 5,29 prosenttiin. Näin pienten ryhmien prosenttiosuuden laskeminen voi kuitenkin olla hieman harhaanjohtavaakin, sillä kyse on usein yhden tai muutaman esiintymän muodostamasta joukosta. Oleellista tässä onkin se, että myös saksan kielessä englannin *manage*-verbi voidaan tämän tarkastelun perusteella kääntää hyvin monella eri tavalla. On muistettava, että 22 verbimuotoisen vastineen lisäksi aineistossa oli muihinkin sanaluokkiin kuuluvia käännöksiä, joten vastineiden kirjo on vielä monipuolisempi kuin tästä kaaviosta voisi päätellä.





Seuraavaksi tarkastelenkin suomen- ja saksankielisten *manage*-verbin vastineiden sanaluokkajakaumaa ja esitän siitä muutaman huomion. Suomenkieliset vastineet jakautuvat verbeihin, substantiiveihin ja luokkaan ”muut”, johon kuuluu 12 sellaista vastinetta (5,88 %), jotka eivät joko olleet lainkaan määriteltävissä (5 kpl), muodostuivat useammasta osasta (substantiivi ja partisiippi *hoidossa oleva*, verbi ja partisiippi *olla esitetty*, verbi ja substantiivi *varmistaa hallinto*, yhteensä 5 kpl) tai olivat adjektiiveja (*jnk alainen*, *vastuullinen*, 2 kpl). Verbimuotoisia vastineita oli 142, mikä muodostaa 69,61 prosenttia kaikista suomenkielisistä vastineista. Verbit olivat monissa eri taivutusmuodoissa ja esimerkiksi ma- ja a-infinitiiveinä, va-, nut-, tu- ja agenttipartisiippeina. Partisiippeja oli suomenkielisessä aineistossa 20 kappaletta eli 9,80 prosenttia kaikista vastineista. Substantiiveja vastineissa oli puolestaan 50 kappaletta eli 24,51 prosenttia. Kuudessa tapauksessa vastine oli osana yhdyssanaa (esim. *to manage a crisis* -> *kriisinhallinta*). Taulukko suomenkielisten vastineiden sanaluokkajakaumasta löytyy liitteestä 3.



Saksankielisten vastineiden kaaviosta 4 (ja liitteestä 3) ilmi käyvä sanaluokkajakauma poikkesi jonkin verran suomenkielisten vastineiden sanaluokkajakaumasta. Verbien osalta tilanne ei ollut dramaattisesti erilainen: saksankielisistä 170 vastineesta verbimuotoisia oli 103, mikä vastaa 60,59 prosenttia. Näistä partisiippimuotoisia (partisiippi II) oli 11 eli 6,47 prosenttia. Mitään suurta eroa suomen partisiippien osuuteen ei siis ollut tässä aineistossa havaittavissa, vaikka verbien osuus olikin suomenkielisessä aineistossa lähes kymmenen prosenttia saksankielisen aineiston verbien osuutta suurempi. Substantiivien

osuus oli puolestaan suurempi saksankielisessä aineistossa. Substantiivimuotoisia vastineita oli 60 kappaletta eli 35,29 prosenttia. Tämä on noin kymmenen prosenttia enemmän kuin suomenkielisessä aineistossa. Näistä yhdyssanan osia oli 9 kappaletta, joista yleisin oli *Management* yhdistelmässä *Risikomanagement*. Kaksi hyvää esimerkkiä yhdyssanarakenteista antaa käänössegmentti, jossa *to manage personnel and resources* on käännetty kahdella eri substantiiviyhdistelmällä *Personalführung und Ressourcenverwaltung*. Tämä kuvaa hyvin myös tutkimusongelmaa: *manage*-verbillä oli kaksi englanninkielistä objektia, jotka kuuluivat eri merkitysryhmiin ("yritys / julkinen" ja "varallisuus"), joten saksankielisiä vastineitakin oli yhden sijaa kaksi.

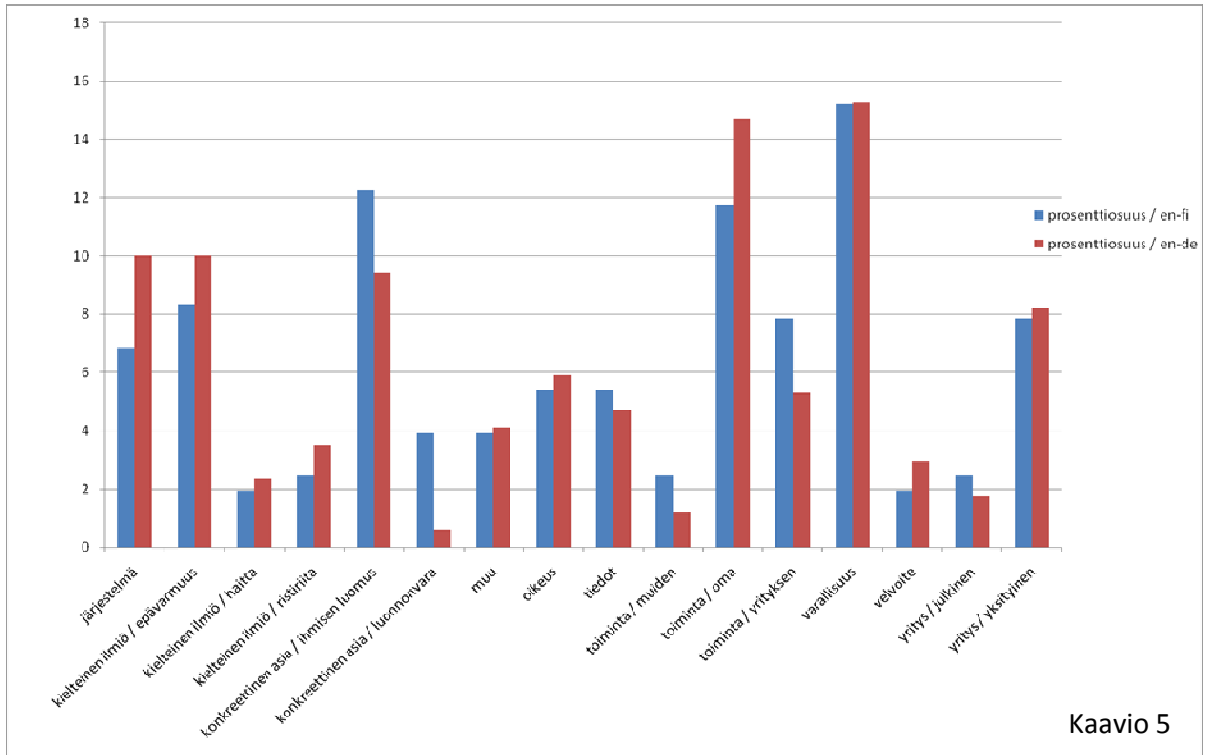


Tämän tutkimuksen tulosten valossa näyttäisi siltä, että saksaksi on hieman tyypillisempää käyttää substantiivirakennetta *manage*-verbin vastineena kuin suomeksi, joskaan ero suomeen ei ole kovinkaan merkittävä. Englanninkielisten verbien suomen- ja saksankielisten vastineiden sanaluokkajakoa laajemmin vertailemalla voisi saada mielenkiintoista tietoa kielten yleisluonteesta, mutta se ei kuitenkaan liity tutkimuskysymykseeni vastaamiseen. Luokkaan "muut" sisältyy yhteensä seitsemän vastinetta, joista kolmea ei voinut määrittellä (hyvin erilainen muotoilu kuin englanniksi, esim. *strategies and tools to manage the market transformation process* -> *Strategien und Instrumente für die Umstellung des Marktes*) ja neljän rakenne oli moniosainen (esim. *to manage [users]* -> *ausgelegt sein für [Benutzer]*). Kaiken kaikkiaan saksan- ja suomenkielisten vastineiden sanaluokkajakaumat vaikuttavat näin karkean jaottelun perusteella melko samankaltaisilta. Hienosyisemmässä analyysissä paljastuisi sitten enemmän kummallekin kielelle ominaisia tapoja niin verbi- kuin

substantiivi-ilmaisuunkin. Alustavana huomiona voin todeta, että esimerkiksi suomenkielisissä vastineissa oli runsaasti verbistä johdettuja teonnimiä (esim. *hallitseminen*, *hallinnoiminen*, *hoitaminen*), kun taas saksassa vastaavia (esim. *das Verwalten*) ei löytynyt yhtäkään. Saksankieliset substantiivit olivat tyypillisesti ung-loppuisia verbeistä johdettuja sanoja, kuten *Verwaltung*, *Bewältigung* ja *Leitung*. En kuitenkaan lähde tarkemmin analysoimaan vastineita, vaan siirryn seuraavaksi tarkastelemaan objektien merkitysryhmien jakautumista eri vastineiden välillä.

Aluksi esittelen kahden aineistoni merkitysryhmäjakaumat, jotta saadaan jonkinlainen yleiskatsaus toisaalta eri merkitysryhmien kokoon ylipäänsä ja toisaalta niiden aineistokohtaisiin jakaumiin. Näiden aineistokohtaisten jakaumien avulla on myös mahdollista arvioida, käsittelevätkö en-fi- ja en-de-aineistot ylipäänsä samanaiheisia asioita. Merkitysryhmien tarkat lukumäärät ja kahden desimaalin tarkkuudelle pyöristetyt prosenttiosuudet löytyvät taulukkomuotoisina liitteistä 4 ja 5. Seuraavan sivun kaaviossa 5 merkitysryhmäjakaumat on esitetty graafisesti.

Molemmissa aineistoissa englanninkielisten *manage*-verbin objektien yleisin merkitysryhmä oli ”varallisuus”, joka kattoi molemmista aineistoista noin 15 prosenttia. Tämä ei ole kovin yllättävää, sillä politiikanteko liittyy usein yhteisten rahavarojen hoitamiseen ja jakamiseen. Määrärahojen lisäksi lähdeasiakirjoissa runsaasti esillä ollut rahastojen säätely kasvatti tämän merkitysryhmän kokoa molemmissa aineistoissa. En-fi-aineiston toiseksi yleisin merkitysryhmä ”konkreettinen asia / ihmisen luomus” oli en-de-aineistossa vasta viidennellä sijalla, mutta prosentuaalisesti ero ei ollutkaan kovin suuri (12,25 vs. 9,41 %). Tätäkään ei voi pitää kovin yllättävänä yleisenä hallinnon (jne.) kohteena, sillä ihmisten luomuksiksi luokitellut asiat (esim. ilmatila) ovat tyypillisiä EU:n politiikan kohteita. Oma toiminta on myös suuri merkitysryhmä molemmissa aineistoissa sijoittuen en-fi-aineiston kolmanneksi yleisimmäksi ja en-de-aineiston toiseksi yleisimmäksi merkitysryhmäksi kattaen molemmissa aineistoissa vielä yli kymmenen prosenttia kaikista englanninkielisistä objekteista.



Kaavio 5

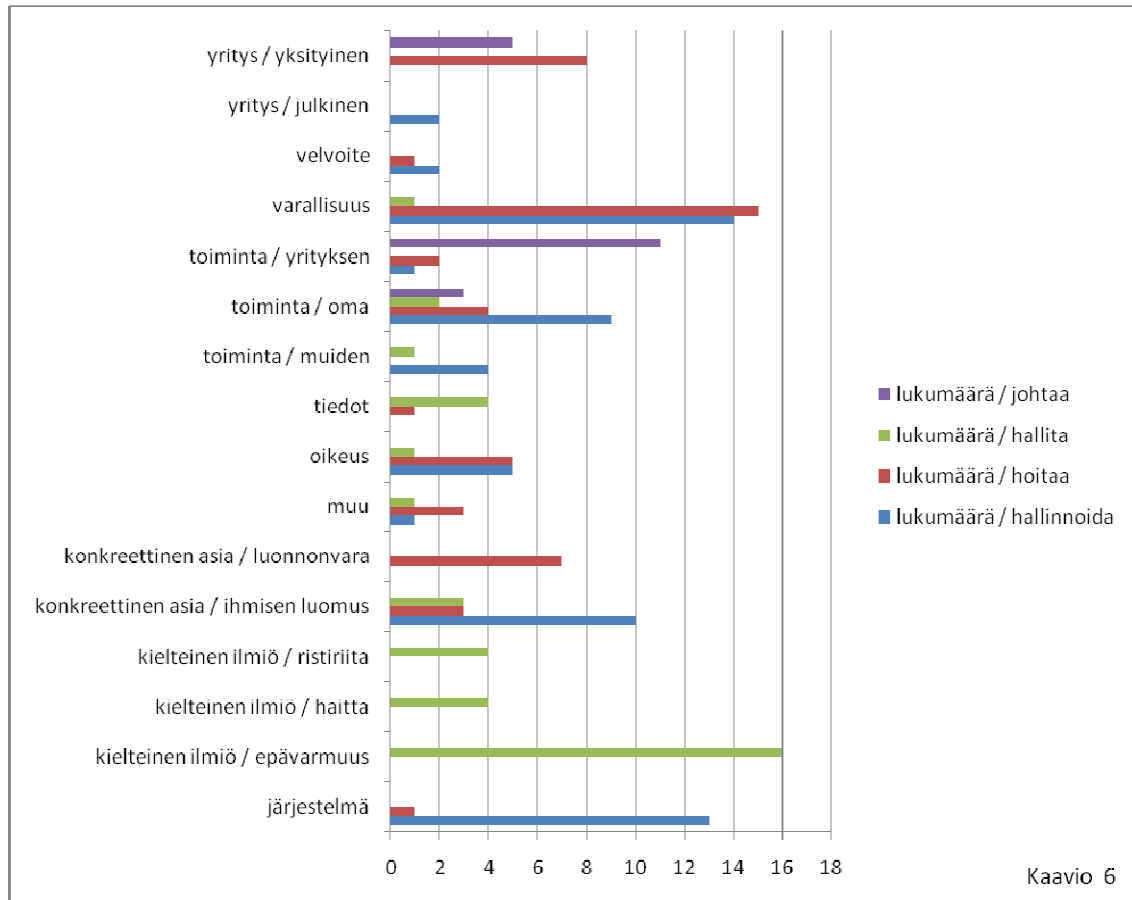
Merkitysryhmien jakaumat muuttuvat mielenkiintoisesti, jos laskee yhteen ne objektit, jotka on luokiteltu eri alamerkitysryhmiin, mutta samaan pääryhmään (esim. erilaiset kielteiset ilmiöt). Tällöin en–fi-aineiston suurin merkitysryhmä on ”toiminta” (45 esiintymää / 22,06 %), toiseksi suurin ”konkreettinen asia” (33/16,18 %), mutta kolmanneksi kirii silti edellisen laskutavan suurin merkitysryhmä ”varallisuus” (31/15,20 %). Tarkemmat tiedot liitteessä 4. En–de-aineistossa samanlainen samaan pääryhmään kuuluvien alamerkitysryhmien yhteennipputtaminen aiheuttaa sen, että toiminta on siinäkin suurin merkitysryhmä (36 esiintymää / 21,18 %). Toiseksi suurimmaksi ryhmäksi nousee merkitysryhmä ”kielteinen ilmiö” (27/15,88 %) ja edellisen laskutavan suurin ryhmä ”varallisuus” on tässäkin aineistossa kolmantena (26/15,29 %). Tarkemmat tiedot liitteessä 5. ”Kielteistä ilmiötä” voidaan pitää merkittävänä merkitysryhmänä molemmissa aineistoissa, sillä se on en–fi-aineistossakin heti neljäntenä (26/12,75 %). Vaikka merkitysryhmäjakaumia manipuloitiinkin näin, ei niiden jakaumassa silti ilmene kovin suuria eroja aineistojen välillä, mikä vahvistaa käsitystä aineistojen samankaltaisuudesta.

Kaiken kaikkiaan ainoa silmiinpistävä ero aineistojen merkitysryhmäjakaumissa on merkitysryhmän ”konkreettinen asia / luonnonvara” kohdalla. En–fi-aineiston kahdeksan

esiintymää muodostavat 3,92 prosenttia kaikista esiintymistä, kun taas en–de-aineistossa luonnonvaroihin luokiteltavia objekteja on ainoastaan yksi kappale (0,59 prosenttia kaikista). Tässä on kuitenkin otettava huomioon se seikka, että kyseessä on todella vain seitsemän esiintymän ero, joten se mikä näyttää dramaattiselta erolta kaaviossa voi todellisuudessa johtua yhden tai kahden lähdetekstin vaikutuksesta aineistossa. Merkitysryhmien jakaumat tukevat sitä vaikutelmaa, joka minulle syntyi jo vastineita etsiessä ja objekteja luokitellessa: tekstit tosiaan vaikuttaisivat pääosin käsittelevän hyvin samankaltaisia asioita ja suuri osa on luultavasti peräisin jopa samojen asiakirjojen eri kieliversioista. Viimeistä väittämää voisi arvioida vertailemalla kahden aineiston EurLex-asiakirjoihin viittaavia tunnisteita, mutta en katsonut tutkimuskysymykseeni vastaamisen kannalta tällaista vertailua oleelliseksi. On paljon kiinnostavampaa siirtyä tarkastelemaan eri suomennosten ja saksannosten merkitysryhmäprofiileja ja yrittää selvittää, onko tietty suomennos tai saksannos taipuvainen esiintymään tiettyjen merkitysryhmien yhteydessä.

Suomenkielisten verbeiksi palautettujen vastineiden joukosta erottuu siis neljä verbiä, jotka muodostavat yli 80 prosenttia kaikista vastineista. Tarkastelen seuraavassa juuri näiden verbien taipumusta esiintyä *manage*-verbin vastineena tiettyjen englanninkielisten objektien merkitysryhmien yhteydessä. Suurimpien verbiluokkien tarkastelu on perusteltua sikäläkin, ettei yksittäisten esiintymien semanttisen preferenssin tutkiminen olisi edes mahdollista: kaikilla vaikuttaisi olevan vahva preferenssi juuri siihen merkitysryhmään, jonka yhteydessä se sattuu aineistossa esiintymään. Neljää suurinta ryhmää seuraa luokka ”muu”, jonka jälkeen seuraavaksi yleisin verbi on *ylläpitää*. Tämä verbi esiintyy aineistossa kuitenkin vain seitsemän kertaa, joten merkitysryhmien prosenttiosuukien laskeminen antaisi tässäkin tapauksessa hyvin epäluotettavan kuvan tilanteesta, kun yksittäinen esiintymä muodostaisi yli kymmenen prosenttia kokonaisuudesta. Tarkastelemissani verbeissä esiintymien lukumäärän vaihteluväli on 19–61. Näistäkin luvuista laskettuna prosenttiosuuksiin on syytä suhtautua kriittisesti, mutta tarkoituksenani ei olekaan tilastollinen vakuuttavuus, vaan alustavien havaintojen esittäminen. Näitä havaintoja voisi sitten myöhemmin testata vieläkin laajemmalla aineistolla, jolloin analyysiin voisi ottaa mukaan myös tilastollisen merkitsevyyden testaamisen. Prosenttiosuukien laskeminen on ylipäätään tarpeen siksi, että kahden hieman erikokoisen aineiston tuloksista saataisiin keskenään vertailukelpoisia. Saksankielisessä aineistossa tilanne on sikäli harmillinen, että vastineet ovat voimakkaasti jakautuneet toisaalta *verwalten*-verbin enemmistöön ja toisaalta lukuisien muiden verbien pieniin esiintymämääriin. Onneksi *verwalten* muodostaa 78:lla esiintymällään hyvin

kiinnostavan tutkimuskohteen myös semanttisen preferenssin kannalta. Lisäksi päätin tarkastella toiseksi suurimman verbin *führen* (12 esiintymää) sekä sen kanssa suhteellisen synonyymiseltä vaikuttavan *leiten*-verbin (8 esiintymää) semanttisia preferenssejä, vaikka niihin on pienten lukumäärien vuoksi suhtauduttavakin hyvin varovaisesti.



Kaavioon 6 olen yhdistänyt neljä yleisintä suomenkielistä verbimuotoon palautettua vastinetta ja kaikki englanninkielisten objektien merkitysryhmät. Tässä kaaviossa käytän vertailuperusteena lukumäärää, sillä verbikohtaisten prosentiosuuksien vertailu ei kerro vielä siitä, missä suhteessa ne kussakin merkitysryhmässä ovat edustettuna. Kaavion tiedot ovat taulukkomuodossa liitteessä 6. Kaaviosta näkyy, että kolme verbeistä esiintyy melko monen merkitysryhmän yhteydessä: *hallita* (10 merkitysryhmää), *hoitaa* (11) ja *hallinnoida* (10). *Johtaa* puolestaan esiintyy ainoastaan kolmen merkitysryhmän kanssa, joista kaikki liittyvät yrityksiin tai toimintaan. Tätä ei liene voida pitää kovin yllättävänä tuloksena, jos ajatellaan *johtaa*-verbin merkitystä. Yhteiskaaviosta voi myös nähdä sen, että joidenkin merkitysryhmien yhteydessä esiintyy vain tietty tai tietyt verbit näiden neljän suurimman

osalta. Erityisen selvää tämä on kielteisten ilmiöiden kohdalla, joita ainoastaan *hallitaan*. Koko aineistossa kielteisten ilmiöiden yhteydessä esiintyy ainoastaan kaksi muuta verbiä, *ehkäistä* ja *säädellä*, kumpikin ainoastaan yhden kerran. Toinen melko selvä tapaus on ”järjestelmän” merkitysryhmä, jota lähes yksinomaan *hallinnoidaan*. Ainoa poikkeus koko aineistossa on yksittäinen *hoitaa*, joka esiintyy jakeluverkkoa tarkoittavan objektin (*distribution system*) yhteydessä. Konkreettisista asioista luonnonvaroja ainoastaan *hoidetaan*, koko aineistossakin ainoa poikkeus tähän on metsistä puhuttaessa *hoitaa ja käyttää*, jossa englannin *manage*-verbiä vastaa siis kaksi suomenkielistä verbiä. Yhteiskaaviosta erottuu vielä selvästi ainakin yrityksen toiminnan *johtaminen*, varallisuuden *hoitaminen* tai *hallinnoiminen* sekä yksityisen yrityksen *johtaminen* tai *hoitaminen*. Seuraavaksi analysoin neljän suurimman suomenkielisen vastineryhmän yksi kerrallaan. Näiden analyysien sisältämät tarkemmat tiedot löytyvät liitteestä 7.

**Hallinnoida**-verbin (61 esiintymää) osalta erottuu neljä englanninkielisten objektien merkitysryhmää, ”varallisuus” (22,95 %) ”järjestelmä” (21,31 %), ”konkreettinen asia / ihmisen luomus” (16,39 %) ja ”toiminta / oma” (14,75 %). Vaikuttaisi siltä, että hallinnoinnilla olisi jonkinlaista semanttista preferenssiä näihin merkitysryhmiin. *Hallinnoida* ei esiinny aineistossa yhtään kertaa minkään ”kielteiseen ilmiöön” kuuluvan merkitysryhmän eikä merkitysryhmien ”konkreettinen asia / luonnonvara”, ”tiedot” eikä ”yritys / yksityinen” yhteydessä. Yllättävimmältä näiden puuttuvien merkitysryhmien (6/16) suhteen vaikuttaa ehkä ensikatsomalta se, ettei luonnonvaroja tunnuta hallinnoitavan aineiston teksteissä. Esimerkiksi metsien tai merien hallinnoiminen vaikuttaisi luonnolliselta yhdistelmältä. Täytyy kuitenkin muistaa, että luokittelin muun muassa ilmatilan ihmisen luomukseksi juuri sillä perusteella, että se on hallinnollinen yksikkö.

**Hoitaa** verbin esiintymiä on kaikkiaan 50, joista 15 kappaletta (30 %) esiintyy merkitysryhmän ”varallisuus” yhteydessä. Muiden esiintymien lukumäärät ovat niin vähäisiä, ettei niistä kannata vetää kovinkaan suoraa johtopäätöksiä. Prosenttiosuuden puolesta yli kymmeneen prosenttiin yltävät kuitenkin merkitysryhmät ”yritys / yksityinen” (16 %) ja ”konkreettinen asia / luonnonvara” (14 %), joten näiden kanssa verbi *hoitaa* on ainakin tämän aineiston puitteissa taipuvainen esiintymään. Luonnonvarojen yhteydessä on lisäksi muistettava, että koko aineiston mittapuulla luonnonvarat esiintyivät yhtä poikkeusta lukuun ottamatta ainoastaan tämän verbin kanssa. Kokonaan puuttuvat (5/16) kielteisiin ilmiöihin liittyvät merkitysryhmät sekä merkitysryhmät ”yritys / julkinen” sekä ”toiminta / muiden”.

Näitä ei kuitenkaan voi pitää kovin voimakkaana negatiivisena preferenssinä, sillä monia muitakin merkitysryhmiä esiintyi *hoitaa*-verbin yhteydessä vain satunnaisia kertoja, mikä tarkoittaa, että nämä puuttuvatkin merkitysryhmät olisivat yhtä hyvin saattaneet sattuman kautta päätyä aineistoon.

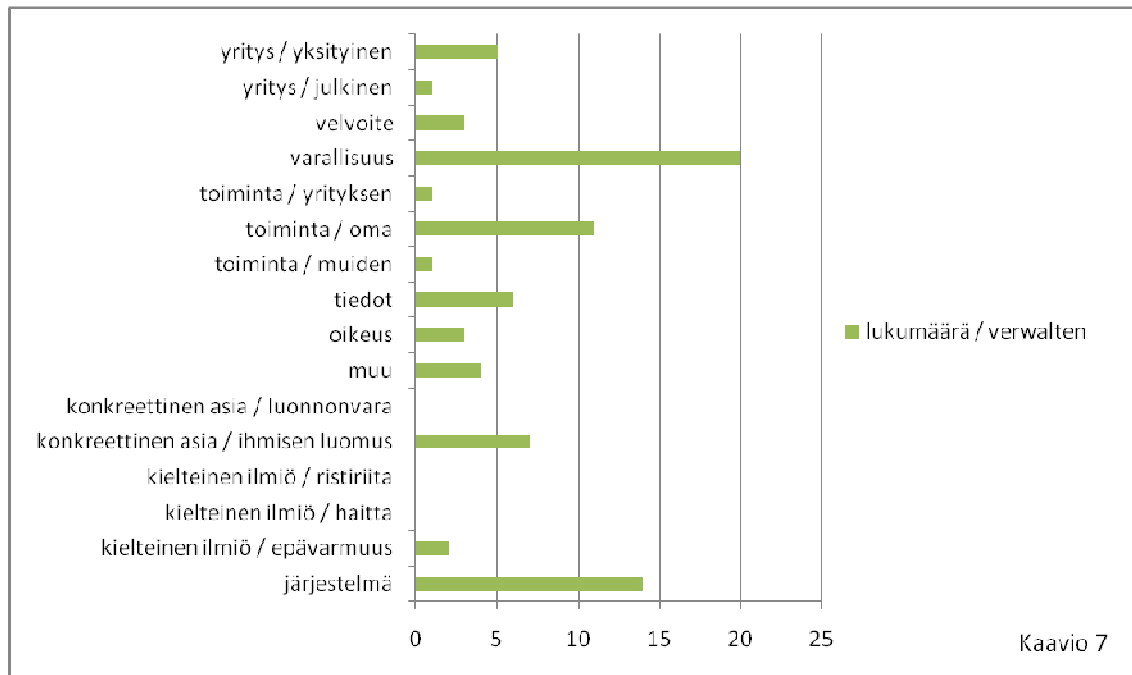
**Hallita**-verbillä voi tämän aineiston osalta sanoa olevan melko voimakas semanttinen preferenssi merkitysryhmään ”kielteinen ilmiö”, sillä sen alaryhmät muodostavat yhteensä 64,86 prosenttia kyseisen verbin merkitysryhmäjakaumasta. Epävarmuuteen liittyvä merkitysryhmä on näistä kolmesta alaryhmästä selvästi suurin 16 esiintymällään kaikista 37:stä *hallita*-verbin esiintymästä (43,24 %). Tämän merkitysryhmän objektit ovat suurimmaksi osaksi erilaisiin riskeihin liittyviä sanoja. Haittaan ja ristiriitaan liittyvät kielteiset ilmiöt ovat kaksi pienempää ryhmää (kummallakin 4 esiintymää / 10,81 % *hallita*-verbeistä). Kyseisiin merkitysryhmiin kuuluvia objekteja on kuitenkin koko aineistossa niin vähän (*haitta*: 4 esiintymää, *ristiriita*: 5 esiintymää), että nämä vaikuttavat ainakin tässä aineistossa tyypillisesti esiintyvän nimenomaan *hallita*-verbin kanssa. Lisäksi, kuten jo aiemmin tuli esiin, koko aineiston kaikista kielteisten ilmiöiden merkitysryhmien kanssa esiintyvistä suomennoksista ainoastaan kaksi on jokin muu kuin *hallita*. Muut merkitysryhmät jäävät yksittäisesiintymiksi (1–4 esiintymää) eikä kuutta merkitysryhmää esiinny lainkaan tämän verbin kanssa. Näistä ei kuitenkaan kannata tehdä kovin pitkälle meneviä johtopäätöksiä niiden vähäisyyden vuoksi.

**Johtaa**-verbin esiintymiä oli aineistossa 19 kappaletta, joista peräti 11 esiintymää (57,89 prosenttia) esiintyi merkitysryhmän ”toiminta / yrityksen” yhteydessä. Jos tähän lisää samaan päämerkitysryhmään kuuluvan ”toiminta / oma”, kattaa erilaisen toiminnan johtaminen yhteensä 73,68 prosenttia kaikista *johtaa*-verbin esiintymistä. Yksityisten yritysten merkitysryhmä muodostaakin sitten loput 26,32 prosenttia tämän verbin esiintymistä. Toisaalta voitaisiin niputtaa yhteen yrityksiin liittyvät merkitysryhmät eli yksityisten yritysten ja niiden toiminnan johtaminen, jolloin peräti 84,21 prosenttia *johtaa*-verbin kanssa esiintyvistä merkitysryhmistä liittyvät yrityksiin. Muut merkitysryhmät loistavatkin sitten poissaolollaan. Tämän aineiston perusteella johtaminen näyttäisi siis liittyvän vahvasti nimenomaan yritysmaailmaan.

**Verwalten**. Seuraavaksi siirrytään en–de-aineistoon ja katsotaan sen yleisimmän verbimuotoon palautetun *manage*-verbin vastineen, *verwalten*-verbin, esiintymistä eri

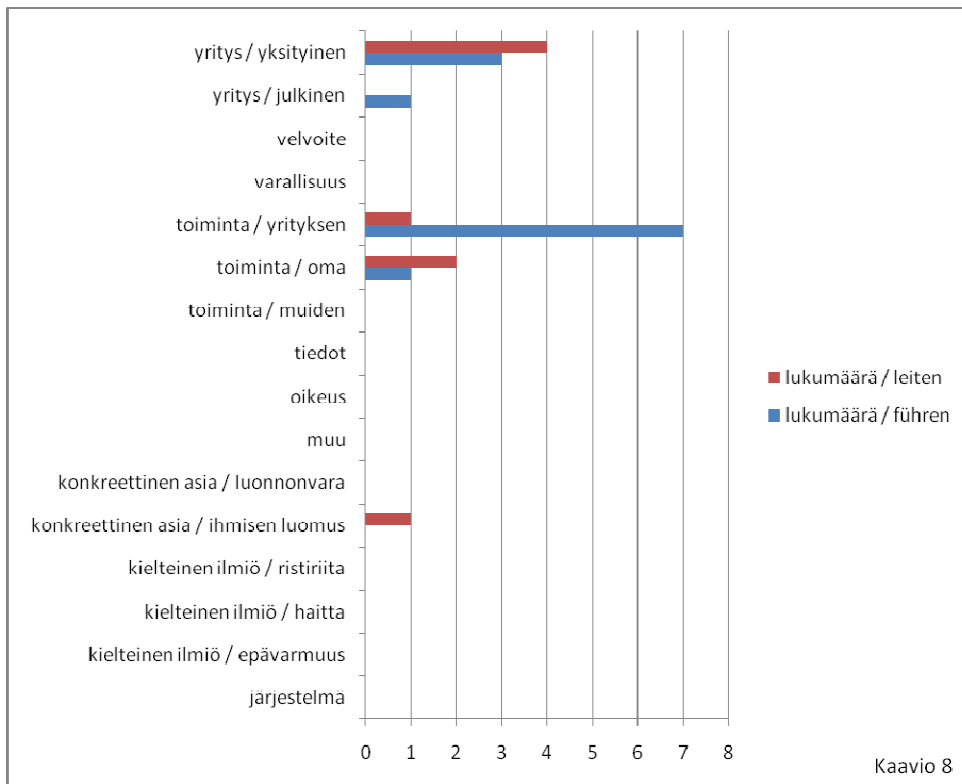


merkitysryhmien yhteydessä. *Verwalten* esiintyi aineistossa yhteensä 78 kertaa, mikä muodosti 45,88 prosenttia kaikista saksankielisistä vastineista. Muut vastineet muodostivat kovin sirpaleisen ryhmän, joten nyt onkin kiinnostavaa selvittää, minkälaisen merkitysryhmien kanssa tämä ehdottomasti yleisin saksankielinen *manage*-verbin vastine esiintyy. Seuraavassa kaaviossa on esitetty *verwalten*-verbin esiintymien jakautuminen eri merkitysryhmien välillä. Tarkat lukumäärät ja prosentiosuudet on esitetty taulukon muodossa liitteessä 8.



Huomionarvoista tässä jakaumassa on se, että *verwalten*-verbi esiintyy vähintään kerran kaikkien muiden merkitysryhmien paitsi merkitysryhmien ”kielteinen ilmiö / haitta”, ”kielteinen ilmiö / ristiriita” ja ”konkreettinen asia / luonnonvara” yhteydessä. Tämän perusteella voidaan sanoa, että sillä tuntuu olevan melko laaja erilaisten merkitysryhmien joukko, jonka yhteydessä se voi esiintyä. Tässä yhteydessä näyttää myös perustellulta, että merkitysryhmien luokittelussa eroteltiin kielteisiin ilmiöihin kolme eri alaryhmää, sillä vain yksi näistä esiintyy tämän verbin kanssa, nimittäin epävarmuuksiin liittyvien kielteisten ilmiöiden alaryhmä. Se esiintyy kuitenkin vain kaksi kertaa, molemmilla kerroilla riskeihin liittyen (*credit exposures* ja *risks*). Riskejä (*risk, exposure*) voidaan kaiken kaikkiaan tämän aineiston perusteella yhdistää lisäksi ainakin verbien *steuern* (4 esiintymää), *begegnen* (3 esiintymää), *bewältigen* (1 esiintymä) sekä *managen* (*Risikomanagement*, 5 esiintymää) kanssa, joten mistään erityisen tyypillisestä yhdistelmästä ei *verwalten* + ”kielteinen ilmiö /

epävarmuus” -yhdistelmän osalta tämän tarkastelun perusteella vaikuttaisi olevan kyse. Olennaista tässä yhteydessä on kuitenkin huomata, että *verwalten* on ammattikäytäntäjien mielestä ainakin mahdollista yhdistää myös tämän monien muiden verbien yhteydessä esiintyvän merkitysryhmän kanssa. Voimakkaimmin *verwalten* vaikuttaa tässä aineistossa suosivan merkitysryhmiä ”varallisuus” (20 esiintymää / 25,64 %), ”järjestelmä” (14 esiintymää / 17,95 %) sekä ”toiminta / oma” (11 esiintymää / 14,10 %). Varallisuudeksi luokiteltuja objekteja oli kaikkiaan aineistossa 26 kappaletta, joten niiden yhteydessä *manage* tunnutaan käännettävän huomattavan usein verbiin *verwalten* palautettavissa olevalla vastineella. Järjestelmäksi luokiteltavien objektien osalta tilanne on samansuuntainen: kaikista 17:stä kyseiseen merkitysryhmään luokitellusta objektista 14 esiintyi käännösyksikössä, jossa *manage*-verbin saksankielinen vastine oli *verwalten*. Merkitysryhmään ”toiminta / oma” tämän verbin yhteys ei ole yhtä selvä, sillä tähän merkitysryhmään kuuluvia objekteja on koko aineistossa 25 kappaletta.



**Kokeilu: *führen* vs. *leiten*.** Jarmo Harri Jantunen (2004) tutki omassa väitöskirjassaan nimenomaan synonyymisten sanojen (suomen määritteet *hyvin*, *kovin* ja *oikein*) semanttista preferenssiä muiden näiden sanojen kontekstiin liittyvien ominaisuuksien ohella. Tästä

innostuneena minäkin tein pienen kokeellisen vertailun melko synonyymisiltä vaikuttavien *manage*-verbin saksankielisten vastineiden *führen* (12 esiintymää) ja *leiten* (8 esiintymää) välillä. Aineisto on näiden sanojen osalta äärimmäisen pieni, joten tämä toimikoon todella lähinnä kokeiluluonteisena ajatusten herättäjänä. Verbit esiintyvät tämän perusteella ainakin merkitysryhmien ”yritys / yksityinen”, ”toiminta / yrityksen” ja ”toiminta / oma” yhteydessä sekä lisäksi *führen* merkitysryhmän ”yritys / julkinen” ja *leiten* merkitysryhmän ”konkreettinen asia / ihmisen luomus” yhteydessä. Kahdesta viimeisestä huomiosta ei vielä kannata päätellä, ettei myös toinen näistä synonyymeistä voisi esiintyä kyseisen merkitysryhmän yhteydessä, sillä erossa on kyse yhdestä(!) esiintymästä. Sen sijaan luetellut merkitysryhmät kuvaavat maalaisjärjellä ajateltuna melko hyvin kyseisten verbien merkitysympäristöjä eikä merkitysryhmien luettelosta tulisi oikopäätä mieleen lisätä näihin muita kuin ”toiminta / muiden”. Mielenkiintoisen paljon aineiston kokoon nähden eroavat merkitysryhmän ”toiminta / yrityksen” yhteydessä esiintyvät vastineet: *führen* esiintyy seitsemän kertaa ja *leiten* vain kerran. Tämä saattaa kuitenkin selittyä vakiintuneella ilmaisulla *Geschäfte führen*, joka esiintyy aineistossa kaikkiaan 7 kertaa, kun merkitysryhmällä ”toiminta / yrityksen” on yhteensä 9 esiintymää koko aineistossa. Kyse voi siis pikemminkin olla idiomaattisesta ilmaisusta kuin varsinaisesta merkityserosta. Tällaisia synonyymeja voisi tutkia joko käyttämästäni aineistosta tai jostain suoraan saksaksi kirjoitetusta aineistosta analysoimalla niiden saksankielisten kollokaattien merkitysympäristöä ja pyrkimällä etsimään semanttista preferenssiä. Se on kuitenkin jo toinen tutkimus, joten jätän tämän kokeilun tältä erää tähän.

### **6.3 Suomen- ja saksankielisten vastineiden vertailua**

Seuraavaksi päästään käsiksi analyysin siihen vaiheeseen, joka selittää koko tutkimusasetelman eli englannin-, saksan- ja suomenkielisen aineiston tarpeen. Pyrin vertailemaan suomen- ja saksankielisten vastineiden merkitysympäristöjen eroja verbien englanninkielisen vastineen *manage* saamien objektien merkitysryhmien avulla. Tavallaan tämän voi katsoa olevan suomen- ja saksankielisten verbien semanttisen preferenssin vertailua, sillä ne käsitteet, jotka englanninkielisissä rakenteissa olivat verbin objekteja, esiintyvät luonnollisesti näidenkin verbien kollokaatteina, vaikka kieliopillinen rakenne voikin poiketa englannin lauserakenteesta (esim. yhdyssanan alkuosa *kriisinhallinta* tai *Risikomanagement*). Molempien aineistojen yleisimpien vastineiden vertailu vaikuttaa kiinnostavalta lähtökohdalta etenkin kun neljästä *hallinnoida*-verbin yleisimmästä merkitysryhmästä kolme muodostaa myös *verwalten*-verbin yleisimmät merkitysryhmät.

”Varallisuus” on molemmissa tärkein (fi: 22,95 %, de: 25,64 %), toisena tulee kummassakin ”järjestelmä” (fi: 21,31 %, de: 17,95 %) ja suomenkielisen verbin neljänneksi yleisin merkitysryhmä ”toiminta / oma” on saksankielisen verbin kolmanneksi yleisin merkitysryhmä (fi: 14,74 %, de: 14,19 %). Suomenkielisen verbin merkitysryhmistä kolmanneksi yleisin eli ”konkreettinen asia / ihmisen luomus” on itse asiassa saksankielisen verbin yhteydessä esiintyvistä merkitysryhmistä neljänneksi yleisin, mutta se jätettiin aiemmasta tarkastelusta pois, koska sen osuus ei yltänyt yli kymmeneen prosenttiin (seitsemän esiintymää / 8,97 %). Yleisimmät merkitysryhmät vaikuttavat siis samoilta kummassakin verbissä. On kuitenkin mielenkiintoista havaita, että siinä missä varallisuudesta puhutaan saksaksi pääasiassa nimenomaan verbin *verwalten* yhteydessä (20/26 esiintymää), suomeksi varallisuutta myös *hoidetaan*, *hallitaan* ja *käytetään*. Merkitysryhmän ”varallisuus” yhteydessä esiintyvät suomenkieliset verbit jakautuvat aineistossa kutakuinkin kahtia hallinnon (14/31 esiintymää) ja hoitamisen (15/31 esiintymää) välillä. Saksan *verwalten*-verbi kattaa siis suurelta osin tutkimusaineiston merkitysryhmän ”varallisuus”, kun taas suomenkielisessä aineistossa kaksi verbiä kattaa enemmistön tämän merkitysryhmän yhteydessä esiintyneistä vastineista.

Verrataan seuraavaksi kokonaan puuttuvia merkitysryhmiä. *Hallinnoida* ei esiinny aineistossa kertaakaan kielteisiin ilmiöihin, luonnonvaroihin, tietoihin eikä yksityisiin yrityksiin liittyvien merkitysryhmien yhteydessä. *Verwalten* ei puolestaan esiinny haittoihin ja ristiriitoihin liittyvien kielteisten ilmiöiden eikä luonnonvaroihin liittyvän merkitysryhmän yhteydessä. On kuitenkin huomattava, että luonnonvaroja ei ylipäänsä käsitellä saksankielisessä aineistossa kuin yhdessä segmentissä (*to manage land* -> *Flächen her/richten*), jolloin tästä ei vielä voi päätellä kovin paljon. Ristiriitoihin ja haittoihin kuuluvia objekteja on kummassakin aineistossa suunnilleen yhtä monta, joten tästä voi ehkä varovaisesti päätellä, että ainakaan näiden aineiston puitteissa verbejä *hallinnoida* eikä *verwalten* käytetään niiden yhteydessä. Suomenkielisen verbin muut puuttuvat merkitysryhmät sen sijaan esiintyvät saksankielisen verbin kanssa. Tietojen merkitysryhmä esiintyy *verwalten*-verbin kanssa kuusi kertaa ja yksityisiin yrityksiinkin liittyvä merkitysryhmäkin viisi kertaa, mikä tässä aineistossa riittää sijoihin viidenneksi ja kuudenneksi yleisin merkitysryhmä. Suomeksi tietoja ei *hallinnoida* vaan *hallitaan*, *käsitellään*, *ylläpidetään*, *hoidetaan* tai niistä *huolehditaan*. Saksaksi tietojen merkitysryhmään liittyviä verbejä oli tässä aineistossa *verwalten*-verbin lisäksi ainoastaan yksi: kahdesti esiintynyt *verarbeiten*, mikä voisi vastata suunnilleen suomen käsittelyä. Jälleen esiintymien määrä on sen verran

pieni, ettei mitään ehdottomia päätelmiä voi tehdä, mutta suuntaa antavana tuloksena tämä on mielenkiintoinen. Merkitysryhmän ”yritys / yksityinen” yhteydessä suomeksi esiintyviä vastineita ovat *hoitaa*, *johtaa* ja *ylläpitää*. Saksaksi tämän merkitysryhmän yhteydessä esiintyvät *verwalten*-verbin lisäksi verbit *betreiben*, *bewirtschaften*, *führen* ja *leiten*. Näissä suomen- ja saksankielisissä verbeissä on joitakin samankaltaisia sanoja, etenkin suomen johtamiseen viittaavat saksaksi *führen* ja *leiten*, mikä onkin odotettavissa yrityksistä puhuttaessa. *Betreiben*-verbin voisi ajatella karkeasti ottaen vastaavan suomen ylläpitämistä.

Eroavatko nämä verbit sitten tutkimusaineistoissa semanttisen preferenssin puolesta toisistaan vai voisiko verbin *verwalten* kääntää suomeksi aina verbillä *hallinnoida*? Vastaus näin kategoriseen kysymykseen on tietysti kääntämisen yhteydessä melkein aina kieltävä, mutta tämän semanttisen preferenssin analyysin avulla kieltävän vastauksen voi myös mielekkäästi perustella. Molemmilla verbeillä esiintyy kollokaatteina sanoja monipuolisesti eri merkitysryhmistä, mutta selviä erojakin löytyy. Erityisen merkittävältä erolta tuntuu se, että varallisuuteen liittyen saksaksi käytetään tutkimusaineistossa *manage*-verbin vastineena pääasiassa verbiä *verwalten*, kun taas suomeksi puhutaan lähes yhtä usein *hallinnoimisesta* kuin *hoitamisestakin*.

Suomenkielisen aineiston *hallita*-verbiä tarkasteltaessa voidaan huomata, että sillä oli melko voimakas semanttinen preferenssi kielteisten ilmiöiden merkitysryhmään. Kaikkiaan 26:sta kielteisiksi ilmiöiksi luokitelluista objekteista 24 esiintyi nimenomaan tämän verbin yhteydessä, lisäksi suomeksi käytetään verbejä *säädellä* ja *ehkäistä*. Saksan 27:stä vastaavasta objektista kuusi esiintyi verbin *bewältigen*, viisi verbin *managen*, viisi verbin *regeln*, neljä verbin *steuern*, kolme verbin *begegnen* ja yksi verbin *umgehen* yhteydessä. Tässä on huomattava ero suomenkieliseen aineistoon verrattuna. Jälleen tulee hyvin näkyviin se, miten eri tavoin eri kielet rakentuvat – jos yksittäisesiintymät jätetään pois laskuista, voidaan todeta, että siinä missä suomeksi riittää tällä kertaa yksi verbi, käytetään saksassa ainakin viittä. Olisi mielenkiintoista pureutua tarkemmin tähänkin havaintoon tutkimalla saksankielisten vastineiden semanttista preferenssiä, mutta se ei tämän tutkimuksen puitteissa ole tarkoituksenmukaista. *Johtaa*-verbin (19 esiintymää) semanttinen preferenssi vaikuttaisi käyvän *verwalten*-verbiä paremmin yhteen saksankielisten verbien *führen* ja *leiten* yhdistetyn semanttisen preferenssin kanssa. Jos saksankieliset verbit yhdistetään (12+8=20), niiden yhteydessä esiintyvistä merkitysryhmistä yleisimmäksi nousee ”toiminta / yrityksen”, toiseksi

yleisemmäksi ”yritys / yksityinen” ja kolmanneksi ”toiminta / oma” aivan kuten suomen *johtaa*-verbin kohdallakin.

## 7. Loppupäätelmät

Tutkielmastani saattaa virheellisesti huokua se ajattelutapa, että käsittäisin englannin kielen olevan jotenkin suomea ja saksaa köyhempi, kun tässä tutkimusasetelmassa yhtä englanninkielistä sanaa kohti löytyy aineistoista runsaasti suomen- ja saksankielisiä vastineita. Haluan nyt esimerkin avulla korjata tämän mahdollisen väärinkäsityksen. Valtioneuvoston kanslian käännös- ja kielipalveluiden (2010a) sanastoja ja ohjeita sisältävällä internet-sivulla on useampikin termiselvitys, joka käsittelee suomalaisia verbejä ja niiden mahdollisia englanninnoksia. Yksi näistä verbeistä on suomenkielinen *tehdä* (Valtioneuvoston kanslian käännös- ja kielipalvelut 2010b), jonka mahdollisia suomennoksia löytyy tiedostosta runsain mitoin. Näiden käyttöä ohjeistetaan käyttöyhteyksien eli kontekstin avulla, joka hyvin usein perustuu juuri objektin merkitykseen (esim. *tehdä väärin* -> *to do wrong* vs. *tehdä vaikutus* -> *to make an impression*). Omassa työssäni asetelma on periaatteessa käänteinen – lähtökielenä on englanti, ja tutkin yhtä englanninkielistä sanaa sekä sen moninaisia suomennoksia ja saksannoksia sekä pyrin hahmottamaan niiden käyttöyhteyksiä lähikontekstin avulla. Kohdekieli vaikuttaa näissä asetelmissa aina jotenkin hienosyisemmin rakentuneelta ja ehkä vivahteikkaammaltakin. Kukin kieli pelaa kuitenkin tavallaan omilla säännöillään ja nimeää sekä käsitteellistä asioita omalla tavallaan. Tästä aiheutuu se, ettei ole olemassa mitään perusrunkoa, joka luettelisi kaikkien kielten käsitteet ja josta voisi sitten rakentaa monikielisen sanaston lisäämällä kaikkien luonnollisten kielten vastineet niille kuuluville paikoille. Juuri tämä tekee kääntämisestä kahden kielen välillä niin haastavaa, mutta samalla myös niin mielenkiintoista. Kun samasta tekstistä tehdään käännöksiä yli 20 kielelle, kuten Euroopan komissiossa, asetelma on tietysti vieläkin monimutkaisempi ja jännittävämpi.

Oman tutkimukseni tuloksia tulkitessa on syytä pitää mielessä aineiston syntykonteksti. Englanninkieliset tekstit ovat suureksi osaksi syntyneet monikielisen ja -kulttuurisen virkamiesyhteisön neuvotteluprosessien tuloksena eikä niiden näin ollen voi ajatella edustavan tyypillisiä alun perin englanniksi kirjoitettuja säädöstekstejä. Lisäksi alun perin englanniksi kirjoitettuja säädöstekstejäkin tehdään ympäri maapalloa: jo Yhdistyneessä kuningaskunnassa laaditut säädöstekstit poikennevat aika lailla yhdysvaltalaisen poliittisen

järjestelmän piirissä syntyneistä teksteistä. Tässä tutkimuksessa aineisto koostuu siis yhdestä englannin kielen varieteetista, jota voisi luonnehtia euroenglanniksi. Euroopan unionin käännösvolyymit huomioon ottaen tämä on kuitenkin arvokas tutkimuskohde. Lisäksi olisi tietenkin hyödyllistä ulottaa *manage*-verbin semanttisen preferenssin tutkiminen myös muihin englannin varieteetteihin, minkä avulla voitaisiin puolestaan saada tietoa Euroopan unionissa käytettävän *lingua francan* ominaispiirteistä. Englanninkielisen aineiston lisäksi toimintakontekstilla on tietenkin vaikutusta myös suomen- ja saksankielisiin teksteihin. Ne ovat suurimmaksi osaksi käännösprosessin tuloksena syntyneitä kieltä, jota voisi myöskin olla mielenkiintoista vertailla alun perin suomeksi ja saksaksi kirjoitettuihin säädösteksteihin vaikkapa käännösuniversaalien tutkimisen muodossa. Lisäksi on muistettava, että komission käännöstoimen pääosastossa käytetään monipuolisia kääntämisen apuvälineitä, joiden avulla pyritään tekstien yhtenäistämiseen. Lisäksi yksittäisillä kääntäjillä saattaa olla henkilökohtaisia kääntämisstrategioita hankalimpien usein toistuvien sanojen kääntämiseksi. Sekä käännösmuistin käyttäminen ja muut apuvälineet että strategiat saattavat vaikuttaa siten, että kerran tietyssä yhteydessä käytettyä vastinetta käytetään samantyyppisissä yhteyksissä yhä uudestaan, vaikka ilman tietoa aiemmista ratkaisuista saatettaisiinkin valita jokin muu vastine. Tutkimuksen tulokset kertovat siis maailman suurimman käännösorganisaation toiminnan tuloksena syntyneiden tekstien piirteistä. Rajoituksistaan huolimatta tämäntyyppinen aineisto antoi ainutlaatuisen mahdollisuuden vertailla samojen tai vähintään hyvin samankaltaisten tekstien englannin-, suomen- ja saksankielisiä versioita keskenään *manage*-verbin ja sen vastineiden osalta.

Juuri kolmen eri kielen välillä tapahtuvan vertailun mahdollisuus oli oleellista tutkimuskysymykseeni vastaamisen kannalta. Tässä aineistossa *manage*-verbin suomen- ja saksankieliset vastineet vaikuttaisivat eroavan toisistaan monella tapaa. Vastineiden kieliopillisessa jakaumassa on huomattavissa saksan hieman suomea suurempi taipumus ilmaista *manage*-verbin sisältö substantiivilla. Kieliopillisia jakaumia olisi ollut erittäin mielenkiintoista tutkia pidemmällekin, mutta se ei olisi auttanut vastaamaan käsillä olevan tutkimuksen tavoitteisiin, vaan paremminkin luonnehtimaan toisenlaisia suomen ja saksan eroja. Tutkimuskysymyksen kannalta mielenkiintoisempaa oli ilmaista kaikki vastineet verbimuodossa ja tarkastella näin saatujen suomen- ja saksankielisten vastineiden jakaumia aineistoissa. Kuten jo johdannossa mainitsin, saksankielinen harjoittelijakollegani ei pitänyt *manage*-verbin kääntämistä kovinkaan ongelmallisena. Tämä on tulostenkin valossa tavallaan ymmärrettävää, sillä saksankielisten vastineiden joukossa yksi on selvästi ylitse muiden:

*verwalten*. Se kattaa yli 45 prosenttia kaikista vastineista ja sitä esiintyy kaikkien paitsi kolmen merkitysryhmän yhteydessä. Voisikin siis ajatella, että ainakin kovan aikapaineen alla saksankielinen kääntäjä voisi melkein aina saksantaa *manage*-verbin joko verbillä *verwalten* tai sen substantiivijohdoksella. Toki laaja kirjo muita saksankielisiä vastineita (peräti 21 muuta verbimuotoon palautettavissa ollutta vastinetta) antaa myös ymmärtää, että saksastakin löytyy monia tapoja ilmaista *manage*-verbin sisältö. Jos pyritään vivahteikkaaseen ja idiomaattiseen kieleen, saksantajankin täytyy varmasti tarkoin harkita, minkä vastineen valitsee.

Suomenkielisten vastineiden osalta tilanne on siinä suhteessa sama, että erilaisia vastineita löytyy runsaasti (tässä aineistossa yhteensä 17 verbimuotoista vastinetta ryhmän ”muu” lisäksi). Muilta osin tilanne onkin sitten huomattavan erilainen. Verbimuotoon palautettujen vastineiden joukosta nousee esiin neljä selvästi muita suurempaa ryhmää: *hallinnoida*, *hoitaa*, *hallita* ja *johtaa*. Yhdelläkään näistä vastineista ei ole semanttisen preferenssin suhteen yhtä monipuolista merkitysryhmien kirjoa kuin saksankielisellä verbillä *verwalten*. Kullakin vastineella on lisäksi havaittavissa omanlaisensa semanttisen preferenssin muodostama profiili. Yleisimmällä *hallinnoida* verbillä on tässä aineistossa taipumusta esiintyä peräti neljän eri merkitysryhmän yhteydessä ja toiseksi yleisimmällä *hoitaa*-verbilläkin ainakin kolmen eri merkitysryhmän kanssa. Nämä semanttiset preferenssit eivät kuitenkaan ole kovin vahvoja muiden merkitysryhmien määriin verrattuna, joten näiden verbien osalta merkitysryhmien jakauma on suomenkielisistä vastineista parhaiten verrattavissa *verwalten*-verbin monipuoliseen semanttiseen preferenssiin. Mielenkiintoista näiden kolmen verbin (*verwalten*, *hallinnoida* ja *hoitaa*) suhteessa on se, että kaikkien verbien yhteydessä yleisimmin esiintyvä merkitysryhmä on ”varallisuus”. Saksankielisessä aineistossa varallisuuden yhteydessä käytetään lähes yksinomaan verbiä *verwalten*, kun taas suomenkielisessä aineistossa tämä merkitysryhmä jakautuu näiden kahden verbin kesken lähes tasan. Olisikin mielenkiintoista tutkia jatkossa tarkemmin en–fi-aineiston englanninkielisiä objekteja, pyrkiä etsimään hienojakoisempaa luokittelua varallisuudelle ja katsoa, pystyttäisiinkö sen avulla saamaan tarkempi semanttinen preferenssi verbeille *hoitaa* ja *hallinnoida*. Toinen vaihtoehto olisi tarkastella näiden verbien kollokationaalista valintaa eli siirtyä merkitysryhmistä tarkastelemaan lekseemitasolla tapahtuvaa myötäesiintymistä.

Kaksi muuta yleisesti esiintyvää suomenkielistä vastinetta ovat *hallita* ja *johtaa*. Näiden semanttinen preferenssi poikkeaa kahden yleisimmän verbin preferenssistä siinä mielessä,



että niiden osalta on havaittavissa voimakkaampaa taipumusta esiintyä tiettyjen merkitysryhmien kanssa. Verbi *hallita* esiintyy aineistossa yli 60 prosentissa tapauksista jonkin kielteiseen ilmiöön liittyvän seikan yhteydessä. Suurin näistä kielteisen ilmiön ryhmistä ovat epävarmuuteen liittyvät kielteiset ilmiöt, jotka ylipäänsä muodostavat aineistossa suurimman kielteisten ilmiöiden merkitysryhmän. Kyseisen verbin ja kielteisten ilmiöiden välillä tuntuu muutenkin vallitsevan voimakas suhde, sillä ainoastaan yksittäistapauksissa kielteisten ilmiöiden yhteydessä esiintyy jokin muu vastine kuin *hallita*. *Johtaa* verbin semanttinen preferenssi näyttää siltä, että se on taipuvainen esiintymään yksityisiin yrityksiin tai niiden toimintaan liittyvien merkitysryhmien tai omaan toimintaan liittyvän merkitysryhmän kanssa. Tässä aineistossa se ei esiinny minkään muun merkitysryhmän yhteydessä. Voimakkain preferenssi sillä vaikuttaisi olevan merkitysryhmään ”yritys / yksityinen”. *Manage*-verbin käsittäminen polyseemiseksi sai tuloksista vahvistusta vähintään polysemian laskennallisen määritelmän mukaan ja erityisesti suomenkielisestä aineistosta. Ainakin suomennettaessa kääntäjän on todella täytynyt tehdä leksikaalinen valinta melko erityyppisten verbien välillä – onhan aika eri asia *johtaa* tai *hoitaa*, *hallinnoida* tai *hallita*. Saksan osalta tulokset eivät olleet yhtä selviä, sillä *verwalten* vaikutti monissa tapauksissa olevan käypä käännös *manage*-verbille.

Tutkimuksen menetelmistä ja tutkimusotteesta voi todeta sen verran, että metodina korpustutkimus oli hyödyllinen, sillä sen avulla oli mahdollista valjastaa valtava tekstimassa yhden tietyn verbin tarkasteluun. Kaksikielinen rinnakkaiskorpus oli luonteva korpusmuoto, kun tutkimuskorpuksia muodostettiin käännösmuistin pohjalta. Kontekstuaalisen semantiikan ajatukset siitä, että sanat saavat merkityksensä vasta kontekstissaan, sopivat mielestäni hyvin käännöstieteelliseen ajattelutapaan. Heti käännöstieteellisen koulutuksen alkutaipaleella opiskelijoihin iskostetaan tietoisuus siitä, ettei kaikki selviä sanakirjasta katsomalla. Ajatus sanojen semanttisesta preferenssistä vaikuttaa tästä tutkimuksesta saamani kokemusten perusteella hyödylliseltä tavalta jäsentää yksittäisten sanojen ja niitä ympäröivän kontekstin suhdetta. Sanakirjoissakin annetaan usein verbien käyttöesimerkkejä nimenomaan antamalla sanakirjan käyttäjälle tiedoksi verbille tyypillinen objekti. Verbien saamiin objektien merkitysryhmien tutkiminen sopivista korpuksista saattaisi antaa systemaattisesti hyödynnettynä arvokasta tietoa myös leksikografeille. Kääntäjät puolestaan voisivat hyötyä alun perin kohdekielellä kirjoitettujen tekstien tutkimuksesta semanttisen preferenssin näkökulmasta, erityisesti synonyymisiltä vaikuttavien sanojen ja *manage*-verbin tavoin polyseemisen sanan mahdollisten vastineiden osalta. Tämä voisi auttaa tekemään

käännetyistä teksteistä entistäkin luontevampia ja kohdekielen ehdoilla toimivia. Semanttisen preferenssin tunteminen voisi nimittäin auttaa kääntäjää tekemään perusteltuja päätöksiä valitessaan sopivaa vastinetta käännökseensä. Asioiden käsitteellistäminen voi olla hyödyllistä, vaikka samalla on tärkeää tunnustaa äidinkielen ammattilaisen kieltä koskevan intuition ja huolellisen selvitystyön merkitys. Toisaalta jo semanttisen preferenssin tutkimisessa hyödynnetään kielitaitoa ja kielellistä intuitiota, mikä on tunnustettava tämänkin tutkimuksen tuloksia arvioitaessa. Olisikin mielenkiintoista pyytää lisää henkilöitä luokittelemaan englanninkieliset objektit samojen kriteereiden mukaisesti merkitysryhmiin ja verrata näin saatuja merkitysryhmäjakoja keskenään. Tämän tutkimuksen puitteissa semanttisen preferenssin tutkiminen tapahtui kiertotietä ryhmittelemällä englanninkieliset *manage*-verbin objektit merkitysryhmiin ja tarkastelemalla sitten kyseisten verbin saamien suomennosten ja saksannosten tendenssejä esiintyä tiettyjen merkitysryhmien yhteydessä. Tämäkin tapa hyödyntää semanttisen preferenssin ajatusta nosti aineistoistani esiin mielenkiintoisia piirteitä. Eri suomennosten ja saksannosten semanttisen preferenssin vertailemisen avulla saattoi havaita paitsi eroja, myös kahden kielen sanoja yhdistäviä piirteitä.

## Lähteet

- Arppe, Antti 2002. Lingvististen korpusten esittelyä. [www-dokumentti].  
<http://www.ling.helsinki.fi/kit/kurssit/verkkokurssit/korpuskurssi/corpus-list.html>  
Päivitetty 25.11.2002.
- Asetus N:o 1 Euroopan talousyhteisössä käytettäviä kieliä koskevista järjestelyistä [www-dokumentti]. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1958R0001:20070101:FI:PDF> Päivitetty 1.1.2007.
- Baker, Mona 1995. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target* 7:2. 223–243.
- Birmingham City University 2010. RDUES - fundamental and applied research in corpus linguistics. [www-dokumentti]. <http://rdues.bcu.ac.uk/>. Luettu 2.5.2010.
- BNC 2009a. Creating the BNC. [www-dokumentti].  
<http://www.natcorp.ox.ac.uk/corpus/creating.xml.ID=design> Päivitetty 26.1.2009.
- BNC 2009b. Creation stage. [www-dokumentti].  
<http://www.natcorp.ox.ac.uk/corpus/creating.xml.ID=creation> Päivitetty 26.1.2009.
- BNC 2009c. Collection of texts. [www-dokumentti].  
<http://www.natcorp.ox.ac.uk/corpus/creating.xml.ID=collection> Päivitetty 26.1.2009.
- Brown Corpus Manual 1997. [www-dokumentti].  
<http://icame.uib.no/brown/bcm.html> Päivitetty 11.9.1997.
- Chesterman, Andrew 2004. Beyond the particular. *Translation Universals: Do They Exist?* Mauranen, Anna & Kujamäki, Pekka (toim.). John Benjamins Publishing Company. 39–49.
- The Concise Oxford German Dictionary 2004. Toim. Clark, Michael & Thyen, Olaf. [sähköinen sanakirja].  
<http://www.oxfordreference.com/views/ENTRY.html?entry=t65b.e15080&srn=1&ssid=342571496#FIRSTHIT> Oxford University Press.
- DGT & JRC 2008. The DGT Multilingual Translation Memory of the Acquis Communautaire: DGT-TM. [www-dokumentti]. <http://langtech.jrc.it/DGT-TM.html#DGT-TM> Päivitetty 12.8.2008.
- Eskola, Sari 2002. Syntetisoivat rakenteet käännessuomessa – Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä. Joensuun yliopisto.

- Euroopan komissio 2009a, EU:n kielipolitiikka. [www-dokumentti].  
[http://ec.europa.eu/education/languages/eu-language-policy/doc99\\_fi.htm](http://ec.europa.eu/education/languages/eu-language-policy/doc99_fi.htm) Päivitetty 19.8.2009.
- Euroopan komissio 2009b, EU-Rechtsvorschriften. [www-dokumentti].  
[http://ec.europa.eu/dgs/translation/translating/legislation/index\\_de.htm](http://ec.europa.eu/dgs/translation/translating/legislation/index_de.htm) Päivitetty 11.8.2009.
- Euroopan komissio 2009c, Official EU languages. [www-dokumentti].  
[http://ec.europa.eu/dgs/translation/translating/officiallanguages/index\\_en.htm](http://ec.europa.eu/dgs/translation/translating/officiallanguages/index_en.htm) Päivitetty 12.8.2009.
- Euroopan komissio 2009d, Translation and the European Union. [www-dokumentti].  
[http://ec.europa.eu/dgs/translation/translating/index\\_en.htm](http://ec.europa.eu/dgs/translation/translating/index_en.htm) Päivitetty 11.8.2009.
- Euroopan komissio 2009e. Kääntäminen monikielisessä yhteisössä. [esitekirjanen].  
 Euroopan yhteisöjen virallisten julkaisujen toimisto, Luxemburg.
- Euroopan komissio 2010. Commission directory. Search by organization chart. DGT: Directorate-General for Translation. [www-dokumentti].  
[http://ec.europa.eu/staffdir/plsql/gsys\\_page.display\\_index?pLang=EN](http://ec.europa.eu/staffdir/plsql/gsys_page.display_index?pLang=EN) Päivitetty 4.5.2010.
- Euroopan unioni 2009. Euroopan unionin elinten käännöskeskus. [www-dokumentti].  
[http://europa.eu/agencies/community\\_agencies/cdt/index\\_fi.htm](http://europa.eu/agencies/community_agencies/cdt/index_fi.htm) Päivitetty 7.1.2009.
- Europa-sanasto 2010. Acquis communautaire. [www-dokumentti]  
[http://europa.eu/scadplus/glossary/community\\_acquis\\_fi.htm](http://europa.eu/scadplus/glossary/community_acquis_fi.htm) Luettu 10.2.2010.
- Granqvist, Tytti 2010. [Henkilökohtainen sähköpostiviesti]. Vastaanottaja: Katariina Franssila. Lähetetty ja luettu 20.5.2010.
- A. Hakulinen, M. Vilkuna, R. Korhonen, V. Koivisto, T.R. Heinonen ja I. Alho (toim.) 2004a.  
*Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seura, Helsinki.
- A. Hakulinen, M. Vilkuna, R. Korhonen, V. Koivisto, T.R. Heinonen ja I. Alho (toim.) 2004b.  
*Iso suomen kielioppi*. [www-dokumentti]. <http://scripta.kotus.fi/visk> Suomalaisen Kirjallisuuden Seura, Helsinki.
- IDS = Institut für Deutsche Sprache, Mannheim. *Grammis – das grammatische Informationssystem des IDS*. [www-dokumentti] <http://hypermedia.ids-mannheim.de/grammis/> Päivitetty 7.5.2010.
- Jantunen, Jarmo Harri 2004. Synonymia ja käännössuomi – korpusnäkökulma samanmerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin. Joensuun yliopisto.
- Kennedy, Graeme 1998. *An Introduction to Corpus Linguistics*. Studies in Language and Linguistics. Longman, London.

- Kujamäki, Pekka & Jääskeläinen, Riitta 2003. Korpukset kääntäjän apuna. Oittinen, Riitta & Mäkinen, Pirjo (toim.). *Alussa oli käänös*. Tampere University Press. 239–252.
- Langenscheidt Taschenwörterbuch Englisch-Deutsch 2003. Langenscheidt, Berlin.
- Laviosa, Sara 2002. *Corpus-based Translation Studies – Theory, Findings, Applications*. Approaches to Translation Studies, Volume 17. Rodopi B.V. Amsterdam.
- Longman dictionary of contemporary English 2007. Pearson Education Limited, England.
- McEnery, Tony & Wilson, Andrew 2001, *Corpus Linguistics*, toinen painos. Edinburgh Textbooks in empirical Linguistics. Edinburgh University Press, Edinburgh.
- McEnery, Tom & Xiao, Richard 2008. Parallel and Comparable Corpora: What is Happening. Anderman, Gunilla & Rogers, Margaret (toim.). *Incorporating corpora*. Translating Europe. Multilingual Matters, Clevedon. 18–31.
- Miller, George A. & Leacock, Claudia 2000. Lexical Representations for Sentence Processing. *Polysemy – Theoretical and Computational Approaches*. Ravin, Yael & Leacock, Claudia (toim.). Oxford University Press, New York. 152–160.
- MOT Englanti 4.8 englanti-suomi. Kielikone Oy & Gummerus Kustannus Oy.
- MOT Kielitoimiston sanakirja 2.0. Kotimaisten kielten tutkimuskeskus ja Kielikone Oy.
- Neves, Anabela 2010. [Henkilökohtainen sähköpostiviesti]. Vastaanottaja: Katariina Franssila. Lähetetty ja luettu 3.5.2010.
- Online Writing Lab 2010 at Purdue University. Participles. <http://owl.english.purdue.edu/owl/resource/627/02/>. Luettu 30.5.2010.
- The Penguin English Dictionary 2007. Penguin Group, London.
- R. Quirk, S. Greenbaum, G. Leech & J. Svartvik 1985. *A Comprehensive Grammar of the English Language*. Longman Group Limited, New York.
- Ravin, Yael & Leacock, Claudia 2000. Polysemy: An Overview. *Polysemy – Theoretical and Computational Approaches*. Ravin, Yael & Leacock, Claudia (toim.). Oxford University Press, New York. 1–29.
- Suursanakirja englanti-suomi 2003. Werner Söderström Osakeyhtiö, Helsinki.
- Tiedeman, Jörg 2010. OPUS – an open source parallel corpus. <http://urd.let.rug.nl/tiedeman/OPUS/>. Päivitetty 28.5.2010.
- Ulkoasiainministeriö 2010. Juhani Lönnroth: ”Monikielisyys on osa Euroopan unionin

aateperustaa.[www-dokumentti]. <http://www.finlande.be/public/default.aspx?contentid=188884&nodeid=37417&contentlan=1&culture=fi-FI> Päivitetty 26.3.2010.

Valtioneuvoston kanslian käännös- ja kielipalvelut 2010a. Sanastot ja ohjeet. [www-dokumentti]. <http://www.vnk.fi/toiminta/kielipalvelut/sanastot/fi.jsp> Luettu 31.5.2010.

Valtioneuvoston kanslian käännös- ja kielipalvelut 2010b. Tehdä. [www-dokumentti]. <http://www.vnk.fi/toiminta/kielipalvelut/sanastot/pdf/tehdä.pdf> Luettu 31.5.2010.

Vuorinen, Erkkä 2010. [Henkilökohtainen sähköpostiviesti]. Vastaanottaja: Katariina Franssila. Lähetetty ja luettu 7.5.2010.

WebCorp Live 2010. Background. [www-dokumentti] <http://www.webcorp.org.uk/guide/> Luettu 2.5.2010.

Yhteinen tutkimuskeskus 2008. The DGT Multilingual Translation Memory of the Acquis Communautaire: DGT-TM. <http://langtech.jrc.it/DGT-TM.html> Päivitetty 12.8.2008.

## Liitteet

Liite 1.

Suomenkielisten verbivastineiden lukumäärät ja prosenttiosuudet

<b>verbi</b>	<b>lukumäärä</b>	<b>prosenttiosuus</b>
hallinnoida	61	29,90
hoitaa	50	24,51
hallita	37	18,14
johtaa	19	9,31
muu	13	6,37
ylläpitää	7	3,43
huolehtia	3	1,47
käsitellä	3	1,47
pystyä	2	0,98
ehkäistä	1	0,49
hoitaa ja käyttää	1	0,49
käyttää	1	0,49
onnistua	1	0,49
pitää	1	0,49
päättää	1	0,49
säädellä	1	0,49
toimia	1	0,49
täyttää	1	0,49
<b>yhteensä</b>	<b>204</b>	

Liite 2.

Saksankielisten verbivastineiden lukumäärät ja prosenttiosuudet

<b>verbi</b>	<b>lukumäärä</b>	<b>prosenttiosuus</b>
verwalten	78	45,88
führen	12	7,06
steuern	9	5,29
leiten	8	4,71
regeln	8	4,71
bewältigen	6	3,53
managen	6	3,53
muu	6	3,53
betreiben	5	2,94
wahrnehmen	4	2,35
bewirtschaften	4	2,35
begegnen	3	1,76
behandeln	3	1,76
durchführen	3	1,76
handhaben	3	1,76
verarbeiten	3	1,76
umgehen	2	1,18
wahren	2	1,18
anordnen	1	0,59
beinhalten	1	0,59
führen, verwalten	1	0,59
her/richten	1	0,59
implementieren	1	0,59
<b>yhteensä</b>	<b>170</b>	



Liite 3.

Suomenkielisten vastineiden sanaluokkajakauma

<b>vastineiden sanaluokka</b>	<b>lukumäärä</b>	<b>prosenttiosuus</b>
verbi	122	59,80
(subst.)	50	24,51
partisiippi	20	9,80
ei määriteltävissä	5	2,45
(subst.) + (va-partisiippi)	2	0,98
(V) + (tu- tai va-partisiippi)	2	0,98
(adj.)	2	0,98
(V) + (subst.)	1	0,49
<b>yhteensä</b>	<b>204</b>	

Saksankielisten vastineiden sanaluokkajakauma

<b>sanaluokka</b>	<b>lukumäärä</b>	<b>prosenttiosuus</b>
verbi	92	54,12
(subst.)	60	35,29
(partisiippi II)	11	6,47
(subst.) + (V)	3	1,76
ei määriteltävissä	3	1,76
(partisiippi II) + (V)	1	0,59
<b>yhteensä</b>	<b>170</b>	

## Liite 4.

## Merkitysryhmäjakauma en-fi-aineistossa

merkitysryhmä / en-fi-aineisto	lukumäärä	prosenttiosuus
1. varallisuus	31	15,20
2. konkreettinen asia / ihmisen luomus	25	12,25
3. toiminta / oma	24	11,76
4. kielteinen ilmiö / epävarmuus	17	8,33
5. toiminta / yrityksen	16	7,84
6. yritys / yksityinen	16	7,84
7. järjestelmä	14	6,86
8. oikeus	11	5,39
9. tiedot	11	5,39
10. konkreettinen asia / luonnonvara	8	3,92
11. muu	8	3,92
12. kielteinen ilmiö / ristiriita	5	2,45
13. toiminta / muiden	5	2,45
14. yritys / julkinen	5	2,45
15. kielteinen ilmiö / haitta	4	1,96
16. velvoite	4	1,96
<b>yhteensä</b>	<b>204</b>	

merkitysryhmä / en-fi-aineisto	lukumäärä	prosenttiosuus
varallisuus	31	15,20
konkreettinen asia	33	16,18
toiminta	45	22,06
kielteinen ilmiö	26	12,75
yritys	21	10,29
järjestelmä	14	6,86
oikeus	11	5,39
tiedot	11	5,39
muu	8	3,92
velvoite	4	1,96
<b>yhteensä</b>	<b>204</b>	

## Liite 5.

## Merkitysryhmäjakauma en–de-aineistossa

Merkitysryhmä / en–de-aineisto	lukumäärä	prosenttiosuus
1. varallisuus	26	15,29
2. toiminta / oma	25	14,71
3. järjestelmä	17	10,00
4. kielteinen ilmiö / epävarmuus	17	10,00
5. konkreettinen asia / ihmisen luomus	16	9,41
6. yritys / yksityinen	14	8,24
7. oikeus	10	5,88
8. toiminta / yrityksen	9	5,29
9. tiedot	8	4,71
10. muu	7	4,12
11. kielteinen ilmiö / ristiriita	6	3,53
12. velvoite	5	2,94
13. kielteinen ilmiö / haitta	4	2,35
14. yritys / julkinen	3	1,76
15. toiminta / muiden	2	1,18
16. konkreettinen asia / luonnonvara	1	0,59
<b>yhteensä</b>	<b>170</b>	

merkitysryhmä / en–de-aineisto	lukumäärä	prosenttiosuus / en–de
varallisuus	26	15,29
toiminta	36	21,18
järjestelmä	17	10,00
kielteinen ilmiö	27	15,88
konkreettinen asia	17	10,00
yritys	17	10,00
oikeus	10	5,88
tiedot	8	4,71
muu	7	4,12
velvoite	5	2,94
<b>yhteensä</b>	<b>170</b>	

Liite 6.

Neljä yleisintä suomenkielistä verbimuotoon palautettua vastinetta ja kaikki englanninkielisten objektien merkitysryhmät

merkitysryhmä	lukumäärä / hallinnoida	lukumäärä / hoitaa	lukumäärä / hallita	lukumäärä / johtaa
järjestelmä	13	1	0	0
kielteinen ilmiö / epävarmuus	0	0	16	0
kielteinen ilmiö / haitta	0	0	4	0
kielteinen ilmiö / ristiriita	0	0	4	0
konkreettinen asia / ihmisen luomus	10	3	3	0
konkreettinen asia / luonnonvara	0	7	0	0
muu	1	3	1	0
oikeus	5	5	1	0
tiedot	0	1	4	0
toiminta / muiden	4	0	1	0
toiminta / oma	9	4	2	3
toiminta / yrityksen	1	2	0	11
varallisuus	14	15	1	0
velvoite	2	1	0	0
yritys / julkinen	2	0	0	0
yritys / yksityinen	0	8	0	5
<b>yhteensä</b>	<b>61</b>	<b>50</b>	<b>37</b>	<b>19</b>

Liite 7. Sivun 1/2.

Neljän yleisimmän suomenkielisen verbimuotoon palautetun vastineen merkitysryhmäjakaumat

<i>hallinnoida</i>		
<b>merkitysryhmä</b>	<b>lukumäärä</b>	<b>prosenttiosuus</b>
varallisuus	14	22,95
järjestelmä	13	21,31
konkreettinen asia / ihmisen luomus	10	16,39
toiminta / oma	9	14,75
oikeus	5	8,20
toiminta / muiden	4	6,56
velvoite	2	3,28
yritys / julkinen	2	3,28
muu	1	1,64
toiminta / yrityksen	1	1,64
kielteinen ilmiö / epävarmuus	0	0
kielteinen ilmiö / haitta	0	0
kielteinen ilmiö / ristiriita	0	0
konkreettinen asia / luonnonvara	0	0
tiedot	0	0
yritys / yksityinen	0	0
<b>yhteensä</b>	<b>61</b>	

<i>hoitaa</i>		
<b>merkitysryhmä</b>	<b>lukumäärä</b>	<b>prosenttiosuus</b>
varallisuus	15	30
yritys / yksityinen	8	16
konkreettinen asia / luonnonvara	7	14
oikeus	5	10
toiminta / oma	4	8
konkreettinen asia / ihmisen luomus	3	6
muu	3	6
toiminta / yrityksen	2	4
järjestelmä	1	2
velvoite	1	2
tiedot	1	2
toiminta / muiden	0	0
yritys / julkinen	0	0
kielteinen ilmiö / epävarmuus	0	0
kielteinen ilmiö / haitta	0	0
kielteinen ilmiö / ristiriita	0	0
<b>yhteensä</b>	<b>50</b>	

Liite 7. Sivun 2/2.

<i>hallita</i>		
<b>merkitysryhmä</b>	<b>lukumäärä</b>	<b>prosenttiosuus</b>
kielteinen ilmiö / epävarmuus	16	43,24
kielteinen ilmiö / haitta	4	10,81
kielteinen ilmiö / ristiriita	4	10,81
tiedot	4	10,81
konkreettinen asia / ihmisen luomus	3	8,11
toiminta / oma	2	5,41
muu	1	2,70
oikeus	1	2,70
toiminta / muiden	1	2,70
varallisuus	1	2,70
järjestelmä	0	0
konkreettinen asia / luonnonvara	0	0
toiminta / yrityksen	0	0
velvoite	0	0
yritys / julkinen	0	0
yritys / yksityinen	0	0
<b>yhteensä</b>	<b>37</b>	

<i>johtaa</i>		
<b>merkitysryhmä</b>	<b>lukumäärä</b>	<b>prosenttiosuus</b>
toiminta / yrityksen	11	57,89
yritys / yksityinen	5	26,32
toiminta / oma	3	15,79
yritys / julkinen	0	0
kielteinen ilmiö / epävarmuus	0	0
kielteinen ilmiö / haitta	0	0
kielteinen ilmiö / ristiriita	0	0
tiedot	0	0
konkreettinen asia / ihmisen luomus	0	0
muu	0	0
oikeus	0	0
toiminta / muiden	0	0
varallisuus	0	0
järjestelmä	0	0
konkreettinen asia / luonnonvara	0	0
velvoite	0	0
<b>yhteensä</b>	<b>19</b>	

Liite 8.

Yleisimmän saksankielisen verbimuotoon palautetun vastineen  
merkitysryhmäjakauma

<i>verwalten</i>		
<b>merkitysryhmä</b>	<b>lukumäärä</b>	<b>prosenttiosuus</b>
varallisuus	20	25,64
järjestelmä	14	17,95
toiminta / oma	11	14,10
konkreettinen asia / ihmisen luomus	7	8,97
tiedot	6	7,69
yritys / yksityinen	5	6,41
muu	4	5,13
oikeus	3	3,85
velvoite	3	3,85
kielteinen ilmiö / epävarmuus	2	2,56
toiminta / muiden	1	1,28
toiminta / yrityksen	1	1,28
yritys / julkinen	1	1,28
kielteinen ilmiö / haitta	0	0
kielteinen ilmiö / ristiriita	0	0
konkreettinen asia / luonnonvara	0	0
<b>yhteensä</b>	<b>78</b>	

# DEUTSCHE KURZFASSUNG

Universität Tampere  
Institut für Sprach- und Translationswissenschaften  
Translationswissenschaft (Finnisch-Deutsch)

FRANSSILA, KATARIINA: Das Kind beim rechten Namen nennen – finnisch- und deutschsprachige Äquivalente des englischen Verbs *to manage* in den Texten der Europäischen Kommission

Magisterarbeit: 63 Seiten  
Anhang: 9 Seiten  
Deutsche Kurzfassung: 13 Seiten  
Juni 2010

---

## 1. Einleitung

Im Frühjahr 2009 arbeitete ich fünf Monate als Praktikantin bei der Europäischen Kommission in Brüssel, in der Generaldirektion Übersetzung. Mein Übersetzungsreferat war das Finnische FI 02, das auf Forschung, Wirtschaft, Technik und Naturressourcen spezialisiert ist. Die Arbeit dort hat mein Forschungsinteresse geweckt, denn täglich stolperte ich über das englische Verb *to manage* oder genauer gesagt darüber, wie man es ins Finnische übersetzen sollte. Es schien, als ob es mehrere Möglichkeiten gäbe, und ich als Übersetzerin eine lexikalische Wahl zwischen sehr unterschiedlichen Wörtern treffen müsste. Als ich mich in der Cafeteria darüber beschwerte, sagte eine österreichische Praktikantin, sie hätte kaum das Problem, da sie fast immer eine und dieselbe Übersetzung dafür (*verwalten*) verwende. Ihre Äußerung regte mich zum Nachdenken an. Besteht dieser Unterschied einfach darin, dass verschiedene Sprachen ihre Begriffe unterschiedlich bilden und ihre eigene Logik haben? Sollte ich das Phänomen einfach akzeptieren? Eines Tages meinte aber eine finnische Kollegin, die eine meiner Übersetzungen prüfte, dass ein anderes finnisches Verb wegen des Objekts von *manage* besser zum Kontext passen würde als das von mir verwendete Verb. Mein Interesse war nun endgültig geweckt. Diese aus der Praxis stammenden Fäden habe ich später im Seminar zusammengeknüpft und daraus meine Forschungsfragen formuliert. Meine erste Forschungsfrage ist, welche Äquivalente man für das englische Verb in den finnisch- und deutschsprachigen Texten der Europäischen Kommission findet. Zweitens versuche ich herauszufinden, ob der Sinngehalt der Kollokate des englischen Verbs einen Einfluss darauf ausübt, welches Äquivalent im finnisch- oder deutschsprachigen Text auftritt. Die Tatsache, dass ich existierende Texte untersuchen wollte, führte dazu, dass ich ein breites empirisches



Untersuchungsmaterial brauchte. Zwei maschinenlesbare Korpora wurden für diesen Zweck gestaltet, aber dieses Thema gehe ich später näher ein.

## **2. Polysemie und semantische Präferenz**

Zuerst versuchte ich das Phänomen, dass das Verb *to manage* in Übersetzungssituationen mehrere Bedeutungen zu haben scheint, begrifflich zu fassen. Das Konzept der Polysemie wurde kurz dargestellt, wobei ich überlegte, ob die Polysemie des Verbs *to manage* zu den vielen möglichen Übersetzungen beiträgt. Polysemie ist ein Begriff der Semantik. Polysemie bedeutet, dass ein Wort mehrere Bedeutungen hat, d.h. dass mit ihm auf verschiedene Begriffe verwiesen wird. Polysemie unterscheidet man von Homonymie unter etymologischen Kriterien: Polyseme sind etymologisch und deshalb auch semantisch verwandt, wohingegen Homonyme nur zufällig gleich geschrieben werden (Homografie) (Ravin & Leacock 2000, 2). Ein anderer Begriff, der einen engen Bezug zur Polysemie aufweist, ist Mehrdeutigkeit. Diese verweist jedoch auf die Tatsache, dass die Sprache von ihrer Natur her flexibel ist und dass Wörter ihre jeweilige Bedeutung erst im Kontext bekommen. Zum Beispiel kann das Wort *Kind* sowohl auf weibliche als auch auf männliche junge Menschen verweisen, was nicht unbedingt bedeutet, dass das Wort *Kind* ein Polysem ist, sondern dass eine semantische Eigenschaft des Wortes unbestimmt ist. Die Unterscheidung zwischen Polysemie und Mehrdeutigkeit ist eine der Kernfragen der semantischen Theorie. (Ravin & Leacock 2000, 3.) Viele Tests sind von Theoretikern für diesen Zweck entwickelt worden, aber die Resultate sind unentschieden, was dazu führt, dass die Unterscheidung unklar bleibt (Ravin & Leacock 2000, 4).

Ravin und Leacock betrachten Polysemie aus der Sicht der Computerlinguistik und konstatieren, dass auch wenn Polysemie kaum Probleme für die zwischenmenschliche Kommunikation bereitet, sie dennoch eine Herausforderung für die semantische Theorie samt ihrer semantischen Anwendungen wie Übersetzung und Lexikografie darstellt (Ravin & Leacock 2000, 1). Auch wenn sich meine Arbeit nicht mit dem Maschinenübersetzen beschäftigt, ist die rechnerische Definition der Polysemie relevant für sie. Semantiker und Lexikografen mögen unterschiedliche Einstellungen zu Polysemie haben, aber für eine Übersetzerin oder einen Übersetzer handelt es sich jedes Mal um ein polysemisches Wort, wenn man im Übersetzungsprozess eine lexikalische Wahl zwischen Wörtern mit unterschiedlichen Bedeutungen treffen muss (vgl. *to manage* → *verwalten*, *führen*,

*bewältigen*). In dieser Hinsicht kann man durchaus auch das Verb *to manage*, als Polysem betrachten.

Als Nächstes musste ich bestimmen, wie ich die Bedeutung der Kollokaten des Verbs *to manage* ermitteln kann. Der finnische Forscher Jarmo Harri Jantunen (2004) hat in seiner Dissertation die Kontextualität der Synonymie und lexikalische Besonderheiten der übersetzten Sprache untersucht und verschiedene Korpora als Forschungsmaterial eingesetzt. Er hat das Bedeutungsumfeld einiger finnischer Synonyme mithilfe von Konzepten der Kontext-Semantik (wie semantische Prosodie und semantische Präferenz), betrachtet. Wenn man die semantische Prosodie eines Wortes untersucht, berücksichtigt man, ob die Kollokaten z. B. eine positive oder negative „Bedeutungsaura“ haben. (Jantunen 24–26.) Oft wird gar nicht zwischen semantischer Prosodie und semantischer Präferenz unterschieden. Beim Untersuchen semantischer Präferenz entwickelt man jedoch genauere distinktive Klassen oder Bedeutungsgruppen als beim Untersuchen semantischer Prosodie. In diese Bedeutungsgruppen werden die Kollokate eines Wortes eingeteilt (Jantunen 2004, 133). Es geht also darum, was der Sinngehalt der Kollokaten eines bestimmten Wortes ist. Aus den Kollokaten des Verbs *to manage* betrachtete ich die Wörter, die man „managt“, welche in den meisten Fällen die Objekte des Verbs waren (*to manage resources*). Viele Forscher haben unterschiedliche Klassifikationen entwickelt (vgl. eine Zusammenfassung auf Finnisch, Jantunen 2004, 136), aber meiner Meinung nach eignete sich keine von diesen für mein Forschungsmaterial. Aus diesem Grunde bildete ich die Bedeutungsgruppen auf korpusbezogene Vorgehensweise (*corpus driven*), d.h. sie wurden durch das Betrachten des konkreten Materials geschaffen. Im Kapitel, das sich mit der Analyse befasst, gehe ich näher auf die verschiedenen Bedeutungsgruppen ein. Als Nächstes beschreibe ich aber die Bedingungen, unter denen das Material meiner Arbeit entstanden ist.

### **3. Entstehungskontext des Forschungsmaterials**

Die Generaldirektion Übersetzung (Directorate-General for Translation, später im Text „die DGT“) ist der interne Übersetzungsdienst der Europäischen Kommission, der alle Amtssprachen der Union umfasst. Sie ist eine der größten Übersetzungsorganisationen weltweit (Euroopan komissio 2009e, 0). Die Hauptgeschäftstellen liegen in Brüssel und in Luxemburg, und die DGT beschäftigt insgesamt etwa 1 750 fest angestellte Übersetzerinnen und Übersetzer (Euroopan komissio 2009e, 5). Organisatorisch ist die DGT sprachbedingt in Abteilungen unterteilt und diese wiederum gliedern sich in verschiedene Themenbereiche

spezialisierende Referate. Eine Ausnahme ist die so genannte Web Unit, in der kleinere Teams mit unterschiedlichen Arbeitssprachen kooperieren (Euroopan komissio 2009e, 11). Neben *in house* -Übersetzerinnen und -Übersetzer verwendet die DGT Freiberufler, die aber im Prinzip keine Rechtsvorschriften übersetzen (Euroopan komissio 2009e, 9–10; Neves 2010). Die DGT übersetzt jährlich eine enorme Textmenge, die aus einer breiten Palette an Textsorten besteht. Im Jahre 2008 wurden insgesamt 1 805 689 Seiten übersetzt, wovon 72,5 Prozent ursprünglich auf (Euro)englisch geschrieben waren. Von den Ausgangstexten waren im Jahr 2008 nur 11,8 Prozent auf Französisch und 2,7 Prozent auf Deutsch geschrieben worden, was einen bedeutenden Unterschied im Vergleich zum Jahr 1997 darstellt (1997: 45,4 % auf Englisch, 40,4 % auf Französisch und 5,4 % auf Deutsch). (Euroopan komissio 2009e, 6.) Bemerkenswert ist die Tatsache, dass die meisten übersetzten Texte zwar auf Englisch geschrieben sind, aber nur ein Bruchteil dieser Texte von Muttersprachlern geschaffen oder editiert wird (Euroopan komissio 2009e, 5; Granqvist 2010).

Bei der DGT stehen den Übersetzenden viele Hilfsmittel zu Verfügung, die zur Vereinheitlichung der Texte dienen und die Arbeit erleichtern. Eines der wichtigsten Werkzeuge ist das Translation Memory (TM) *Translator's Workbench* von SDL Trados, das an die Anforderungen der DGT angepasst worden ist und seit 1997 verwendet wird (Euroopan komissio 2009e, 13). Auf das lokale TM werden relevante, bereits übersetzte Textsegmente aus dem zentralen TM geladen. Das zentrale TM ist ein wichtiger Teil des *European advanced multilingual information system*, kurz *Euramis*. Hierbei handelt es sich um ein bei der Kommission entwickeltes System. Wenn die Übersetzung fertig ist, werden die in dem lokalen TM befindlichen Übersetzungssegmente in das *Euramis* geschickt. (Euroopan komissio 2009e, 13.) Ein Teil dieses zentralen TM wurde von der DGT im November 2007 im Internet veröffentlicht (DGT & JRC 2008). Das Forschungsmaterial dieser Arbeit stammt aus diesem veröffentlichten TM, aber dieser Thematik widme ich mich nach der Vorstellung der Korpusanalyse als Methode.

#### **4. Korpuslinguistik als Methode**

Mein Forschungsmaterial bestand aus zwei elektronischen Korpora und in meiner Arbeit benutzte ich korpuslinguistische Werkzeuge, wie z. B. zweisprachige Konkordanz. Um mein Material und meine Vorgehensweise besser zu verstehen, ist es nützlich, kurz die Geschichte der Korpusstudien zu erläutern und danach den Versuch zu unternehmen, Korpus nach

heutigem Verständnis zu definieren. Erstens ist es wichtig zu verstehen, dass Korpuslinguistik *eine Methode* ist, mithilfe derer man viele verschiedene Aspekte der Sprache in vielen Bereichen der Linguistik untersuchen kann (McEnery & Wilson 2001, 2). Heutzutage wird sie außerdem gern und oft in weiteren Disziplinen wie z. B. in der Translationswissenschaft genutzt. Die Geschichte der Korpora beginnt aber schon bevor man die Translationswissenschaft als eigenständige Disziplin anerkannt hat. Am Anfang des 20. Jahrhunderts gehörte die Korpusanalyse zur Basismethodik der Linguisten. Damals waren Korpora große empirische Textsammlungen und z. B. die Linguisten, die Feldstudien unternahmen, übten sehr wohl Korpusanalyse (McEnery & Wilson 2001, 2–3). Die frühe Korpuslinguistik begegnete aber sehr praktischen Herausforderungen: umfangreiche Korpora mussten hauptsächlich ohne Hilfe von Maschinen verarbeitet werden (McEnery & Wilson 2001, 12). Ende der 1950er Jahre wurde die frühe Korpuslinguistik scharf vom berühmten Linguist Noam Chomsky kritisiert. Er war der Ansicht, dass ein Linguist das Wissen über die Sprache (*competence*) beschreiben sollte anstatt den Sprachgebrauch (*performance*) zu erforschen, weil externe Faktoren einen störenden Einfluss auf den Sprachgebrauch haben und das Material damit die Sprache in täuschendem Licht darstellt (McEnery & Wilson 2001, 6). Ein zweiter Kritikpunkt waren die möglichen Unstimmigkeiten des Materials. Sie beziehen sich auf Konstruktionen, die generell nur selten in der untersuchten Sprachvarietät vorkommen, aber zufällig im Korpus überrepräsentiert sind und im Gegenteil. Diesen Kritikpunkt kann man jedoch durch sorgfältiges Sammeln der Korpora und das Konzentrieren auf Repräsentativität zumindest teilweise annullieren (McEnery & Wilson 2001, 75).

Die ersten elektronischen Korpora sammelte man schon in den 1960ern, aber besonders schnell begann sich die Korpuslinguistik in den 1980ern zu entwickeln (McEnery & Xiao 2008, 18), als die Computer üblicher wurden. In der modernen Korpuslinguistik wird die Sprache also empirisch und systematisch mithilfe von umfangreichen maschinenlesbaren Textmassen untersucht (Kujamäki & Jääskeläinen 2003, 243). Die Sammlungskriterien (*corpus design*) und Computerverfahren zur Korpusanalyse bilden die Methodologie der Korpuslinguistik (Laviosa 2002, 6). In der quantitativen Korpusanalyse werden sprachliche Eigenschaften klassifiziert, berechnet und die Resultate der Analyse möglicherweise statistisch getestet, aber man kann Korpora auch qualitativ als Quelle realer Beispiele behandeln. (McEnery & Wilson 2001, 76.)

Heutzutage bezieht sich der Begriff Korpus demnach hauptsächlich auf elektronisches Forschungsmaterial. Eine der zentralen Figuren in der korpuslinguistischen Translationswissenschaft, Mona Baker, betont, dass ein wichtiges Merkmal eines Korpus ist, dass es unter klaren Kriterien für einen bestimmten Forschungszweck gesammelt worden ist (1995, 225). Die finnische Translationswissenschaftlerin Sari Eskola hält die Repräsentativität für die allerwichtigste Voraussetzung für ein wohlgebildetes Korpus (2002, 25). Es gibt aber keine einheitliche Meinung darüber, unter welchen Kriterien Korpora gesammelt werden sollten. Diese hängen im Prinzip von den Zielen der Untersuchung ab. (Eskola 2002, 26.) Seit Anfang der 1990er Jahre sind Korpusstudien ein fester Bestandteil der Translationswissenschaft (Laviosa 2002, 1). Sara Laviosa ist neben Mona Baker einer der Pioniere auf diesem Feld. Mona Baker hat als Erste vorgeschlagen, dass man von korpuslinguistischen Methoden auch in der Translationswissenschaft profitieren könne. Sie hat auch den Begriff Übersetzungsuniversalien geprägt (Eskola 2002, 28–29). Was für Korpusstypen werden denn in korpuslinguistischen Übersetzungsstudien eingesetzt? Zum einen sind Korpora, die aus Ausgangstexten und ihren Übersetzungen bestehen (parallele Korpora), von Nutzen. Zum anderen werden so genannte vergleichbare Korpora genutzt. Sie beinhalten Texte in einer Sprache, aber die Texte in einem Teilkorpus sind ursprünglich in der betreffenden Sprache geschrieben und die Texte in einem anderen/einigen anderen Teilkorpus/Teilkorpora sind übersetzt worden. (Eskola 2002, 32.)

## **5. Forschungsmaterial**

Mein eigenes Forschungsmaterial bestand wie gesagt aus zwei Korpora, die ich im Folgenden beschreibe. Sie sind aus dem von der DGT veröffentlichten TM gebildet worden, die man im Internet finden kann<sup>10</sup>. Dieses TM beinhaltet den größten Teil der so genannten *acquis communautaire*, womit man den gemeinschaftlichen Besitzstand bezeichnet. Genauer gesagt beinhaltet es den Gesamtbestand an Rechten und Pflichten, der für die Mitgliedstaaten der EU verbindlich ist. Wegen dieser Verbindlichkeit ist dieser Besitzstand in alle Amtssprachen der EU mit Ausnahme des Irischen übersetzt worden. (Yhteinen tutkimuskeskus 2008.) Aus diesem Grunde gibt es in diesem TM reichlich Texte in den Sprachen, die mich wegen meiner Forschungsfragen interessierten. Ein Universitätslektor der Universität Tampere, Herr Doktor Mikhail Mikhailov, hat die in meiner Arbeit benutzten Korpora gesammelt. Die Texte sind in dem TM in Übersetzungssegmenten gespeichert und die Segmente bestehen oft aus

---

<sup>10</sup> <http://langtech.jrc.it/DGT-TM.html>

einem Satz. Die DGT kann nicht garantieren, dass alle Texte komplett in dem TM gespeichert sind oder dass die Reihenfolge der Segmente der Reihenfolge des Originals entspricht (Neves 2010). Die Segmente beinhalten aber die EurLex-Nummer des Dokumentes, mithilfe derer man das Original in der EurLex-Datenbank finden kann. Diese Eigenschaft war bei der Analyse nützlich. Aus dem TM kann man zweisprachige Korpora zusammenstellen. Für meine Arbeit wurde ein Korpus mit den Sprachen Englisch und Deutsch hergestellt und eines mit den Sprachen Englisch und Finnisch. Diese werde ich im Folgenden als Korpus en–de und Korpus en–fi bezeichnen. Konkret liegen sie im Server „mustikka“ vom Institut für Sprach- und Translationswissenschaften der Universität Tampere und für ihre Bearbeitung wurde die Benutzeroberfläche TACT (Tampere University Corpus Tools) beansprucht.

Im Korpus en–de gab es insgesamt 19 976 277 Wörter, von denen 10 620 764 auf Englisch waren und 9 355 513 auf Deutsch. Im Korpus en–fi gab es 16 782 105 Wörter, von denen der englische Teil 9 981 904 Wörter ausmachte und der finnische 6 800 201 Wörter. Wie man sieht, sind die beiden Korpora ziemlich umfassend und verfügen im Prinzip über eine ähnliche Größe. Die Größe war auch ein wichtiger Grund dafür, dass ich dieses Material als meinen Forschungsgegenstand wählte. Auch wenn es einige Nachteile in diesem Material gibt, hätte ich nie im Rahmen einer Pro Gradu -Arbeit eine so umfangreiche Textmenge manuell sammeln können. Für die Beantwortung meiner Forschungsfragen ist es wichtig, viele authentische Verwendungsbeispiele zu haben, und für diesen Zweck war das Material ideal. Genau genommen sind die Korpora eigentlich TM-Sammlungen und entsprechen somit nicht der striktesten Definition von Korpus. Sehr unterschiedliche Korpora werden jedoch immer noch als Korpus bezeichnet. Am allerwichtigsten ist es zu fragen, wie gut ich auf meine Forschungsfragen mit diesem Material antworten kann. Es ist wahr, dass die Ausgangssprache der Texte aus dem Material nicht hervorgeht. Die Texte sollten aber einander ganz genau entsprechen, woraus folgt, dass ich Äquivalente des englischen Verbs *to manage* in den finnischen und deutschen Texten suchen kann. Die Tatsache, dass die Texte nicht unbedingt komplett sind führt dazu, dass es sich nicht lohnt, Frequenzen aus dem gesamten Material zu zählen. Ich machte aber eine Suche im Material und bekam eine Menge an Beispielen für den Gebrauch des Verbs *to manage*. Diese Resultate konnte ich wiederum auch quantitativ behandeln. Das Wichtigste war, dass ich genügend Beispiele bekam. Als Nächstes stelle ich dar, wie ich das Material vor der Analyse behandelte.

## 6. Analyse

Die Vorbereitung der Analyse erfolgte nach der Zusammenstellung der Korpora in folgender Weise. Mithilfe der Bedieneroberfläche TACT führte ich in beiden Korpora dieselbe zweisprachige Konkordanz-Suche durch. Das Stichwort war *manage*, das einen beliebigen Teil des Wortes bilden durfte (damit bekam ich auch die Imperfektformen und einige ed-Partizipien des Verbs als Resultat) und ein Verb (V) sein sollte. Aufgrund dieser Suche bekam ich 240 Treffer im Korpus en–fi und 198 im Korpus en–de. Danach prüfte ich noch, ob das Konkordanz-Programm Partizipien (auch V) in der Form *managing* findet, aber bekam keine Treffer. Die Resultate speicherte ich in zwei Excel-Dateien, die ich weiter behandelte. Vor der tatsächlichen Analyse suchte ich die finnisch- und deutschsprachigen Äquivalente des Verbs *to manage* in den finnisch- und deutschsprachigen Segmenten und speicherte sie in einer eigenen Spalte. Gleichzeitig suchte ich auch die Objekte des englischen Verbs (oder bei Partizipien das Hauptwort, z. B. *the funds managed*; Partizipien gab es nicht sehr viele, und von hier an spreche ich nur von Objekten auch wenn ich Objekte samt dieser Hauptwörter meine). Zuletzt suchte ich noch die deutsch- und finnischsprachigen Äquivalente der englischen Objekte.

In der zweiten Runde prüfte ich die Äquivalente und löschte doppelte Segmente. Mit doppelten Segmenten meine ich solche Segmente, die zweimal im Suchergebnis aber nur einmal in der tatsächlichen EurLex-Urkunde vorkommen. Zusätzlich löschte ich ein Übersetzungssegment im Korpus en–fi und zwei im Korpus en–de, die auch in den deutsch- und finnischsprachigen Segmenten das Verb *to manage* auf Englisch enthielten (Computerbefehle). Nach diesen Maßnahmen hatte ich 204 Übersetzungssegmente im Korpus en–fi und 170 Übersetzungssegmente im Korpus en–de. Danach markierte ich die Wortart aller finnisch- und deutschsprachigen Äquivalente, die keine Verben waren. Zusätzlich markierte ich auch Partizipien, weil ihre Eigenschaften teilweise den Eigenschaften von Adjektiven ähneln. Als Hilfsmittel bei dieser Sortierung benutzte ich eine deskriptive Grammatik des Finnischen, *Iso Suomen kielioppi* (Hakulinen ym. 2004a), und ihre Online-Version (Hakulinen ym. 2004b) sowie die Online-Grammatik vom IDS Mannheim *Grammis* (2010). Im Korpus en–fi gab es fünf Fälle, in denen ich wegen einer sehr unterschiedlichen Formulierung oder einem Mangel an Elementen, die im englischsprachigen Segment vorkamen, kein Äquivalent für das Verb *to manage* fand. Im Korpus en–de war dies nur in drei Segmenten der Fall. Die vorletzte Phase vor der eigentlichen Analyse war, dass ich alle Äquivalente, auch z. B. die Substantive, als Verben

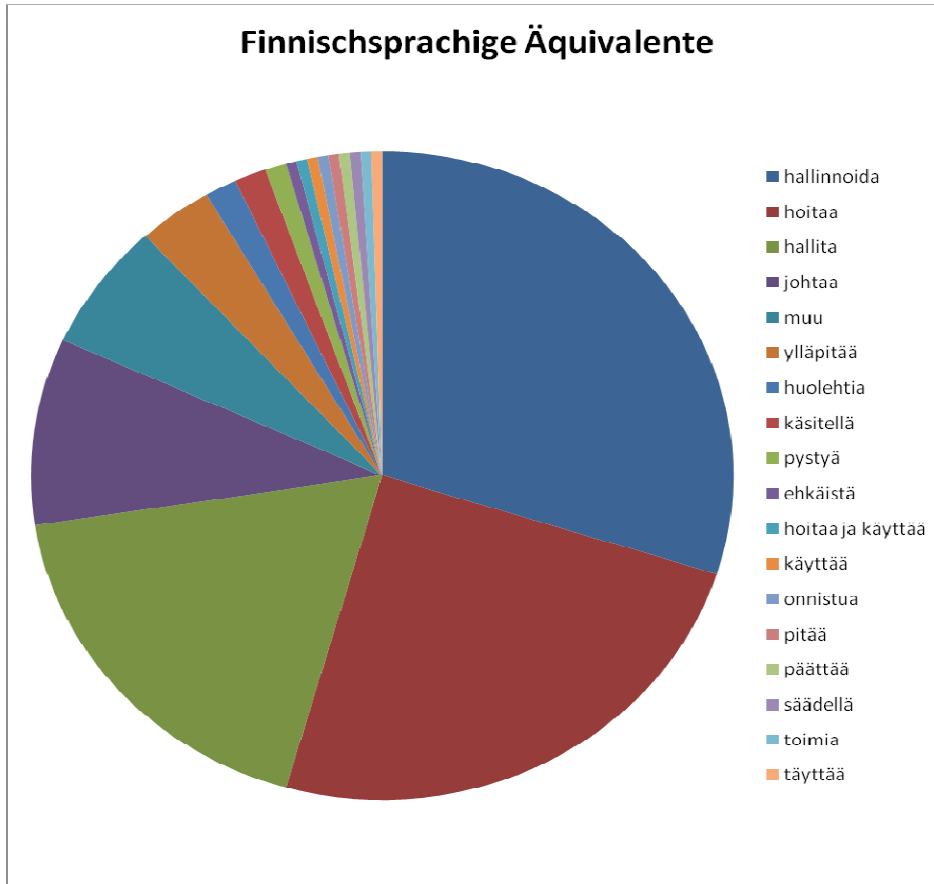
anführte (z. B. *die Verwaltung* -> *verwalten*). Dies machte ich, damit der Vergleich erleichtert wurde und weil mein Gedanke war, dass die Bedeutung des Substantivs ungefähr der Bedeutung des Verbs entspricht.

Die allerletzte Vorbereitungsphase war die Klassifizierung der englischen Objekte in Bedeutungsgruppen um die Analyse der semantischen Präferenz zu ermöglichen. Ich schuf aufgrund meines Forschungsmaterials neun Bedeutungsgruppen, von denen einige noch Untergruppen bekamen. Neben diesen Gruppen gab es noch eine gemischte Gruppe „andere“, weil eine lückenlose Klassifizierung nicht sachgemäß schien (insgesamt 16 Bedeutungsgruppen). Beim Klassifizieren berücksichtigte ich das vollständige Objekt und nicht nur das Hauptwort. Die Bedeutungsgruppen waren „System“ (Objekte wie *Community programme, regulated market, aid scheme, emergency communication and information system*); „negatives Phänomen“ mit drei Untergruppen „negatives Phänomen / Unsicherheit“ (*risk, exposure*), „negatives Phänomen / Beschwerden“ (*impacts, noise issues and effects, potential adverse consequences*) und „negatives Phänomen / Konflikte“ (*conflict, crisis*); „konkrete Sache“, die zwei Untergruppen, „konkrete Sache / menschliche Kreation“ (*radioactive waste, ports*) und „konkrete Sache / Naturressourcen“ (*forests, fish stocks*) bekam; „Rechte“ (*permit, right, airworthiness, quota*); „Information“ (*studbook, list, waiting codes, fisheries data, information*); „Handlung“, die drei Untergruppen, „Handlung / von anderen“ (*fishing activities, fishing efforts*), „Handlung / eigene“ (*project, operations relating to registration*) und „Handlung / eines Unternehmens“ (*undertaking's affairs*), bekam; „Eigentum“ (*investments, holdings, monies and other assets, funds, trusts* aber auch negatives Eigentum wie *public debt, expenditure*); „Verpflichtung“ (*contract, environmental responsibilities, acquis*); „Unternehmen“ mit zwei Untergruppen „Unternehmen / privates“ (*investment companies, farm*) und „Unternehmen / öffentliches“, die staatliche Organisationseinheiten wie *secretariats, centre* sind; und die Mischgruppe „andere“.

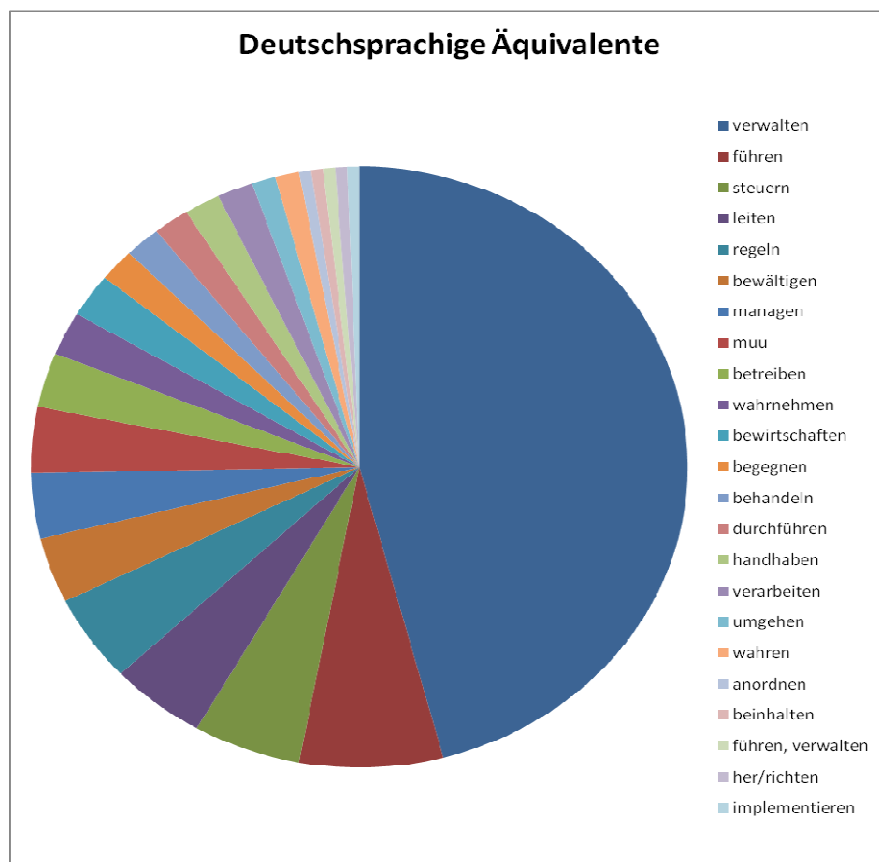
Normalerweise untersucht man semantische Präferenz durch Klassifizierung der Kollokate eines Wortes in Bedeutungsgruppen. Ich konzentrierte mich auf die Klassifizierung der englischsprachigen Objekte, wodurch ich auch über einen Umweg an die Bedeutung der Kollokate der finnisch- und deutschsprachigen Äquivalente gelangte (z. B. *to manage a register* -> ein **Verzeichnis** zu *verwalten*, englischsprachiges Objekt als „Information“ klassifiziert, was auch für das deutschsprachige Objekt gilt). Mein Ziel war, herauszufinden, ob ich bei den deutsch- und finnischsprachigen Äquivalenten irgendeine Präferenz im



Hinblick auf bestimmte Bedeutungsgruppen beobachten kann. Zu einer quantitativen Untersuchung gehörten strikt genommen statistische Tests, aber mein Ziel ist nur, präliminare Resultate zu erzielen, die dann später mit anderen Methoden weitergetestet werden könnten.



Es gab insgesamt 17 verschiedene finnischsprachige Äquivalente samt Mischgruppe „andere“ (auf Finnisch *muu*). Die genauen Prozentzahlen (fi *prosenttiosuus*) samt Anteilen (fi *lukumäärä*) findet man in der Beilage (fi *Liite*) 1. Hier kann man deutlich sehen, dass es vier Äquivalente gab, die den überwiegenden Anteil ausmachten: *hallinnoida*, *hoitaa*, *hallita* und *johtaa* besitzen voneinander abweichende Bedeutungen (lexikalische Wahl muss getroffen werden) und machen zusammen 81,86 Prozent von allen finnischsprachigen Äquivalenten aus.



Wenn man das Diagramm über die deutschsprachigen Äquivalente betrachtet, bemerkt man einen deutlichen Unterschied zu den finnischsprachigen Äquivalenten: ein Verb erfreute sich großer Beliebtheit, nämlich das Verb *verwalten*, das allein 45,88 Prozent aller deutschsprachiger Äquivalente ausmachte. An zweiter Stelle folgte *führen*, das aber nur noch 7,06 Prozent der Äquivalente bildete. Von *verwalten* abgesehen war das Feld sehr zersplittert. Die genauen Angaben zum Diagramm sind in der Beilage 2 aufgeführt. Was die Wortartaufteilung angeht, waren 35,29 Prozent der deutschsprachigen Äquivalente Substantive, wohingegen von den finnischsprachigen Äquivalenten 24,51 Prozent Substantive waren.

Als ich die Bedeutungsgruppeneinteilung in den zwei Korpora verglich, schien sie in den beiden Korpora ziemlich gleich zu sein, was als relative Gleichheit der Themenbereiche der Texte in beiden Korpora interpretiert werden könnte. Ich betrachtete die semantische Präferenz der vier üblichsten finnischen Verben näher und machte folgende Beobachtungen:

Das Verb *hallinnoida* trat am üblichsten in Verbindung mit den Bedeutungsgruppen „Eigentum“ (22,95 %), „System“ (21,31 %), „konkrete Sache / menschliche Kreation“

(16,39 %) und „Handlung / eigene“ (14,75 %) auf. Es gab sechs Bedeutungsgruppen von den insgesamt 16 Bedeutungsgruppen, in deren Zusammenhang dieses Verb gar nicht vorkam. Das Verb *hoitaa* trat recht oft in Verbindung mit der Bedeutungsgruppe „Eigentum“ (30 %) auf. Bemerkenswert ist auch die Tatsache, dass die Bedeutungsgruppe „konkrete Sache / Naturressource“ fast ausschließlich in Verbindung mit diesem Verb auftrat. Das Verb *hallita* schien eine klare Präferenz für negative Phänomene zu haben, die in 64,86 Prozent der Fälle mit dem Verb auftreten. Die semantische Präferenz des Verbs *johtaa* (~‘führen‘ oder ‘leiten‘) war nicht sehr überraschend: es wurde in Verbindung mit „Handlung / eines Unternehmens“ (57,89 %), „Unternehmen / privates“ (26,32 %) und „Handlung / eigene“ (15,79 %) benutzt. Die genauen Anteile und die konkrete Anzahl sind in der Beilage 7 aufgeführt.

Als Nächstes betrachtete ich das dominante deutschsprachige Verb, *verwalten*. Bei diesem Verb ist es erwähnenswert, dass es in Verbindung mit allen Bedeutungsgruppen außer „negatives Phänomen / Beschwerden“, „negatives Phänomen / Konflikt“ und „konkrete Sache / Naturressource“ auftrat. Es schien also, als ob das Verb *verwalten* in Verbindung mit vielen unterschiedlichen Bedeutungsgruppen vorkommen könnte, auch wenn es im vorliegenden Material am stärksten die Bedeutungsgruppen „Eigentum“ (25,64 %), „System“ (17,95 %) und „Handlung / eigene“ präferierte.

Wenn ich die deutsch- und finnischsprachigen Äquivalente mithilfe ihrer semantischen Umgebung verglich, bemerkte ich interessante Unterschiede. Das finnische Verb *hallinnoida* und das deutsche Verb *verwalten* wiesen aber viele Gemeinsamkeiten auf. Sie wurden beide in meinem Material am üblichsten in Zusammenhang mit der Bedeutungsgruppe „Eigentum“ benutzt und überhaupt waren die vier üblichsten Bedeutungsgruppen gleich, auch wenn die Reihenfolge leicht variierte. Bemerkenswert bei der Bedeutungsgruppe „Eigentum“ war aber, dass auf Finnisch gleich oft ein anderes Verb, *hoitaa*, genauso oft benutzt wurde, wobei auf Deutsch in Verbindung mit Eigentum fast ausschließlich das Verb *verwalten* vorkam.

Zusammenfassend kann man sagen, dass diese Untersuchung den Grundgedanken hinter der Forschungsfrage bestätigte, dass es für das englische Verb *to manage* mehrere gängige Äquivalente in den deutsch- und finnischsprachigen Texten der Europäischen Kommission gibt. Vor allem scheint es aber, dass die ins Finnische Übersetzenden eine bedeutende lexikalische Wahl treffen müssen, wobei die ins Deutsche Übersetzenden zumindest im Notfall auf das Verb *verwalten* zurückgreifen können. Eine idiomatische Übersetzung

verlangt auch in der Deutschen Sprache ein intensiveres Nachdenken, wofür die vielen möglichen Äquivalente ein Zeichen sind. Die Kenntnis der semantischen Präferenz der Äquivalente könnte der Übersetzerin oder dem Übersetzer eventuell bei der lexikalischen Wahl helfen. Es wäre vielleicht nützlich, die semantische Präferenz der möglichen deutsch- und finnischsprachigen Äquivalente noch mit ursprünglich auf Finnisch und Deutsch geschriebenen Forschungsmaterialien zu vergleichen.