

## **Mistä kertoo valtiosopimusten kieli? Tapaustutkimus interferenssistä Suomen ja Venäjän välisissä valtiosopimuksissa**

Mikhail Mikhailov & Miia Santalahti  
Tampereen yliopisto

### **Abstract**

This article deals with corpus-based research methods applied to the study of translation and introduces the Parallel Electronic corpus of State Treaties (PEST), which is currently being compiled as part of the authors' research project. In this article, we analyse certain features discovered in state treaties between Finland and Russia. These analyses were conducted in order to study whether evidence can be found of one party's dominant role in the treaty preparation process – in other words, whether the text of the treaties reflects the power relations between the parties. We approach the topic by analysing word frequency lists and keyword statistics and comparing the language of treaties in various periods of time. This tentative analysis suggests that a lot can, indeed, be interpreted from the language of state treaties. Information describing the typical discourse of treaties in different languages can be obtained by observing the vocabulary used in them. Moreover, based on deviations from the typical language of the genre, some conclusions can be made regarding segments that have most likely been translated in a certain direction, which would indicate domination by the corresponding party.

**Keywords:** corpus-based research, parallel corpora, quantitative research, frequency lists, language and power

**Asiasanat:** korpuspohjainen tutkimus, rinnakkaistekstikorpuksat, määrällinen tutkimus, frekvenssilistat, kieli ja valta

### **1 Taustaa**

Kahdenväliset valtiosopimukset ovat usein pitkien neuvottelujen ja monien kompromissien tulosta. Valmisteluvaiheessa sopimustekstiä korjataan, täydennetään ja täsmennetään jatkuvasti. Tekstiä myös käännetään useita kertoja molempiin suuntiin: kielellä tuotetaan sekä alkuperäistekstiä että käännöstä niin, että lähde- ja kohdetekstiä ei enää pystytä identifioimaan (Probirskaja 2009: 47). Myöskään oikeustieteellisestä näkökulmasta mitään valtiosopimusten tai kansainvälisten sopimusten kieliversioista ei pidetä käännökseenä, vaan kaikkien katsotaan olevan alkuperäistekstejä. Valtiosopimukset ovat tyypillisesti yhtä todistusvoimaisia kaikkien osapuolten kielillä, ellei sopimustekstiin erikseen lisätä merkintää, jonka mukaan ristiriitatapauksissa noudatetaan tiettyä tekstiä (Valtiosopimusoikeutta koskeva Wienin Yleissopimus 33:1, ks. myös Schäffner 1997).

On kuitenkin syytä uskoa, että on olemassa melko paljon valtiosopimuksia, joista yksi kieliversio on alkuperäisteksti ja muut alkuperäisteksteiksi nimitettyjä käännöksiä. Tällaisia ovat esimerkiksi sopimukset, joiden luonnokset on laadittu molemmille osapuolille vieraalla kielellä eli ns. välittäjäkielellä ja vasta lopullinen versio on käännetty osapuolten kielille. Ennen toista maailmansotaa välittäjäkielenä käytettiin ranskaa, mutta nykyään tämä rooli on siirtynyt englannin kielelle. Välittäjäkielen käyttö on yleistynyt Suomessa tehtäessä sopimuksia sellaisten kaukaisten maiden kanssa, joiden kielille on vaikeaa järjestää käännöspalveluita. Vuonna 2016 englannin kielen välityksellä laadittiin esimerkiksi valtiosopimus Suomen ja Vietnamin välillä (35/2016) sekä Suomen ja Uzbekistanin välillä (43/2016). Jopa Ruotsin ja Suomen välisiä valtiosopimuksia laaditaan englanniksi (esimerkiksi 82/2012). Myös Venäjän kanssa laadittiin 1990-luvulla joitakin sopimuksia englanniksi (esim. 110/2002), mutta 2000-luvulla on jälleen luovuttu välittäjäkielen käytöstä.

Käytettiinpä välittäjäkieltä tai ei, valtiosopimusten teksti on aina monikielisen viestinnän tulos. Tästä huolimatta Fernando Prieto Ramos (2015) on todennut, että sopimusten kieliversioiden eroja tutkitaan yllättävän vähän, koska uskotaan, että niitä ei ole tai ne ovat pieniä, vaikka tämä ei välttämättä pidä paikkaansa.

Suoraan osapuolten kielillä laaditut sopimustekstit ovat hybriditekstejä, joissa on sekä käännettyä että suoraan kyseisellä kielellä tuotettua ainesta. Tämä ei kuitenkaan tarkoita sitä, että molempien osapuolten panos olisi aina täysin sama. Voidaan ajatella, että jos toinen kieliversio on käännos, siinä on toisen version vaikutusta. Kysymys kuuluukin: Kuinka paljon eri sopimusteksteissä on havaittavissa toisen kielen vaikutusta eli interferenssiä? Ja ehkä vielä mielenkiintoisempi kysymys: Mistä mahdollinen interferenssi kertoo?

Voidaan olettaa, että jossain tapauksissa sopimusasiakirjoja laaditaan suurimmalta osin vain toisen osapuolen kielellä, eli toinen osapuoli dominoi sopimusprosessia ja esittää toiselle valmiita tekstejä käännettäviksi. Tällainen on esimerkiksi vuonna 1944 solmittu rauhansopimus, jossa toisena osapuolena olivat Neuvostoliitto ja Iso-Britannia ja toisena Suomi: sopimustekstin mukaan suomenkielinen versio ei ole edes todistusvoimainen. Edelleen voidaan olettaa, että tällaisissa toisen osapuolen dominoivassa ohjauksessa laadituissa sopimuksissa olisi myös selvästi havaittavissa käännostekstien piirteitä, kuten yksinkertaistumista, eksplisiittistämistä, konservatiivista ilmaisua, tasoittamista sekä suoranaista lähdekielen vaikutusta eli interferenssiä (ks. esim. Baker 1996 ja Toury 2012).

Gideon Toury (2012: 310–311) esittää, että *interferenssiä* ilmenee käytännössä väistämättä kaikessa käännöstoiminnassa. Lähdekielen vaikutus voi ilmetä joko negatiivisina heijasteina (kohdejärjestelmälle epätavallisten elementtien käyttönä) tai positiivisina heijasteina (kohdejärjestelmälle ominaisten elementtien lisääntyneenä käyttönä), mutta täysin vaikutteista vapaan käännoksen tekeminen edellyttäisi erityisolosuhteita ja/tai erityistä panostusta kääntäjältä.

Valtiosopimus kontekstissa vastapuolen kielelle tyyppisten elementtien käyttö voi myös kertoa kielen tasolla tehdystä kompromissiratkaisusta: toinen osapuoli on välttämättä halunnut sopimustekstiin tietyn tekstuaalisen muotoilun ja toinen on suostunut siihen saadakseen vastaavasti oman tahtonsa läpi jossain toisessa kohdassa

(ks. esim. Probirskaja 2009: 74). Molemmissa tapauksissa sopimuskielessä ilmenevän interferenssin voidaan katsoa kertovan jotain sopijaosapuolten välisistä suhteista.

Tämän tutkimuksen tavoitteena on selvittää, miten Suomen ja Neuvostoliiton/Venäjän väliset suhteet ovat vaikuttaneet valtiosopimusten laadintaan. Hypoteesina on, että varsinkin YYA-sopimusten aikakaudella (jota myös suomettumisen ajaksi kutsutaan) Neuvostoliitto olisi usein ollut dominoivana osapuolena Suomen kanssa solmituissa valtiosopimuksissa. Odotuksena on, että tuota aikakautta edustavissa suomenkielisissä sopimusteksteissä ilmenisi eniten venäjämäistä suomeen päin tapahtuneen käännöstoiminnan vaikutusta. Aineistona käytämme kokoamaamme valtiosopimusten rinnakkaiskorpusta, jonka esittelemme luvussa 2. Ensin tarkastelemme koko aineistoa frekvenssilistojen ja niistä nousevien epätavallisten ilmaistusten kautta (luku 3). Tämän jälkeen vertailemme eri aikakausia avainsana-analyysin avulla (luku 4).

## 2 Valtiosopimusten rinnakkaiskorpus

Osapuolten rooleja valtiosopimuksen laadinnassa voidaan tarkastella tekstin kvantitatiivisen analyysin keinoin. Tutkimuksen objektiivisuus ja validiteetti vahvistuvat, kun käytetään monipuolista aineistoa: eri kielipareja, eri maita, eri aikakausia. Myös tekstien kohdistaminen antaa uusia mahdollisuuksia tutkimukseen. Tällaista tutkimusta varten olemme aloittaneet laajan kahdenvälisen valtiosopimusten rinnakkaiskorpuksen PEST:n kokoamisen (PEST = Parallel Electronic corpus of State Treaties). Tarkoituksena on koota siihen Suomen ja Venäjän sekä Suomen ja Ruotsin välillä tehtyjä valtiosopimuksia eri aiheista ja eri aikakausilta alkaen vuodesta 1918. Myöhemmin korpukseen lisätään myös Venäjän ja Ruotsin välillä solmittuja sopimuksia sekä kansainvälisiä sopimuksia.

Venäjän ja Suomen väliset suhteet ovat olleet hyvinkin dramaattisia ja vaihtelevia – niihin mahtuu sekä vihollisuutta että ystävyyttä. Ruotsin ja Suomen väliset suhteet taas ovat olleet aivan erilaiset, ja ruotsi on ollut Suomen virka- ja hallintokielenä vuosisatojen ajan, aina vuoteen 1919 asti. Tuleekin olemaan kiinnostavaa selvittää, näkyykö tämä seikka myös sopimusten kielessä: onko Suomen ja Venäjän välisten sopimusten suomen kieli erilaista kuin Suomen ja Ruotsin välisten sopimusten kieli?

Työ on aloitettu suomi–venäjä-korpuksesta. Tätä artikkelia viimeisteltäessä korpuksessa on 221 tekstiparia ja runsaat 230 000 sanaa eli sanaesiintymää suomeksi ja lähes 300 000 sanaa venäjäksi (sanamääräerot selittyvät pääasiallisesti kielten rakenne-eroilla). Tekstit ovat kohdistettuja (toisiaan vastaavat segmentit erikielisissä versioissa on merkitty), lemmatisoituja (jokaisen sanan perusmuoto eli lemma on määritetty) sekä annotoituja morfologisella ja syntaktisella tasolla (jokaisen sanan sijamuoto ja syntaktinen funktio on ilmaistu). Lisäksi teksteistä on tallennettu metadataa, kuten otsikko, päivämäärä, aihealue yms. Korpusta kasvatetaan ja uusia tekstejä lisätään jatkuvasti, mutta teksteistä voi jo tehdä hakuja ja alustavia analyysejä. Korpusta tutkitaan erityisesti sitä varten kehitetyllä TextHammer-verkkosovelluksella, jonka avulla saadaan monenlaisia kvantitatiivisia ja kvalitatiivisia tietoja. TextHammer-

työkaluilla voi hakea esimerkiksi saneiden, lemموjen, sanaluokkien tai sijamuotojen frekvenssilistoja, N-grammeja eli 2–5 sanan kokonaisuuksia, sanojen frekvenssejä eri teksteissä tai eri osakorpuksissa sekä kollokaatteja eli haetun sanan lähiympäristössä esiintyviä sanoja.

Korpustutkimuksessa voidaan tutkia koko korpuksen tekstimassaa tai erilaisia tekstien ryhmiä eli osakorpuksia. Korpus voidaan jakaa osakorpuksiin eri perusteilla: aihe, aikakausi, jne. Tässä artikkelissa keskitytään eri aikakausina laadittujen sopimusten välisiin eroihin, joten Suomen ja Neuvostoliiton/Venäjän väliset sopimukset on jaettu aikakauden mukaan kolmeen osakorpukseen (ks. Taulukko 1). Osakorpus A käsittää ajanjakson Suomen itsenäistymisestä toisen maailmansodan päättymiseen, osakorpus B puolestaan ns. YYA-sopimusten aikakauden ja osakorpus C Neuvostoliiton hajoamisen jälkeisen ajan. Kuten Taulukosta 1 voidaan havaita, osakorpukset ovat hieman epätasapainossa: osakorpus B on paljon suurempi kuin muut osakorpukset, mikä selittyy osin sillä, että se käsittää myös muita pidemmän ajanjakson.

**Taulukko 1:** PEST-korpuksen kokoonpano: aikakaudet

Osakorpus	Kausi	Tekstiparien määrä	Sanemäärä: suomi	Sanemäärä: venäjä
A	1918–1944	49	69 330	81 059
B	1945–1991	128	115 076	141 683
C	1992–2016	44	52 546	68 797
Yhteensä		221	236 952	209 928

Tekstejä tarkasteltaessa on luonnollisesti huomioitava myös muita taustamuuttujia, kuten yleinen kehittyminen sekä käänösstrategioiden muuttuminen ajan mittaan. Jäljempänä tässä artikkelissa käsiteltävät sanat *yhteistoiminta* ja *yhteistyö* ovat esimerkki tällaisesta kehityksestä. Aika-aspektin huomioiminen myös osaltaan lisää tutkimuksen objektiivisuutta.

### 3 Frekvenssilistoihin perustuva analyysi

Yksi tapa analysoida tekstejä on sanojen frekvenssi- eli taajuuslistojen tarkastelu. Frekvenssilistat antavat monenlaista tietoa tekstin erityispiirteistä: mitkä sanat ovat teksteissä yleisimpiä, kuinka yleisiä jotkin tietyt sanat ovat jne. Luvut ja tilastot ovat korpustutkimuksessa kuitenkin vain lähtökohta, jonka perusteella huomio kiinnittyy joihinkin tiettyihin seikkoihin, ja vasta tarkemman analyysin avulla voidaan edetä aineiston kuvauksesta ilmiöiden selittämisen tasolle (Baker 2004: 17). Niinpä tässäkin analyysissä on yhdistetty kvantitatiivista ja kvalitatiivista tutkimusta. Aluksi koottiin korpuksen hakutoiminnon avulla sanojen lemmatisoidut frekvenssilistat<sup>1</sup> sekä venäjän-

<sup>1</sup> Listoissa sanan frekvenssi tarkoittaa kaikkien sen sijamuotojen yhteenlaskettuja esiintymiä, esim. 'tuoli' 20, 'tuolia' 15, 'tuolilla' 5, 'tuolissa' 3 → TUOLI 43.

että suomenkielisistä teksteistä. Tarkasteluun otettiin vain vähintään 10 kertaa esiintyneet sanat. Venäjänkielisessä listassa on noin 1700 sanaa ja suomenkielisessä noin 1600 sanaa. Tämän jälkeen frekvenssilistat käytiin läpi ja tiettyjä, huomionarvoisilta vaikuttavia esiintymiä otettiin lähempään tarkasteluun.

Diskurssille ominaisten sanojen tarkastelu auttaa muodostamaan käsityksen sopimuskielen tyypillisistä piirteistä ja niiden kehityksestä. Tällainen tieto on hyödyllistä esimerkiksi diskurssin tutkijoille, sopimustekstien kirjoittajille sekä monikielisen viestinnän opiskelijoille. Tässä artikkelissa keskitytään kuitenkin käännöstoiminnan vaikutusten ilmenemiseen eli nimenomaan sopimuskielen genrelle epätyypillisten sanojen tarkasteluun.

Epätavallisuutta lähestyttiin sen ennako-odotuksen kautta, että valtiosopimusten sanasto koostuisi neutraaleista sanoista (*talo, paljon, tehdä, pitkä*), viralliseen rekisteriin ja lakikielen diskurssiin kuuluvista sanoista (*täten, tahi, taho*), sopimustermeistä (*sopimuposapuoli, voimassaolo*) ja sopimusten aiheina olevien erikoisalojen termeistä (*rajanylityspaikka, liikenteenharjoittaja*). Frekvenssilistoja tarkasteltaessa kiinnitettiin huomiota sellaisiin sanoihin, jotka eivät kuulu mihinkään edellä mainituista ryhmistä. Lähempään tarkasteluun valittiin erityisesti sanoja, joita voidaan pitää sopimuskielen kontekstissa epätavallisina. Sanojen luokittelu tyyliä varten ei ymmärrettävästikään ole yksiselitteinen toimenpide, johon olisi olemassa absoluuttisia sääntöjä ja rajanvetoja. Tässä analyysissä sanat on luokiteltu ”manuaalisesti” käyttäen referenssinä esimerkiksi sanakirjojen rekisterimerkintöjä sekä hakuja muista suomen ja venäjän kielen korpuksista (kuten Araneum-kokoelman korpuksista, <http://ella.juls.savba.sk> ja Kielipankista, [corp.csc.fi](http://corp.csc.fi)).

Valittuja elementtejä tarkasteltiin monisanaisten ilmaisujen (N-grammien), kollokaattien (hakusanan lähikontekstissa usein esiintyvien sanojen) ja konkordanssien haun avulla. Luetteloissahan sanat esiintyvät yksittäisinä ja irrallisina, ja monissa tapauksissa on vaikeaa tehdä päätelmiä niiden merkityksistä ja sävyistä ilman kontekstia. Esimerkiksi venäjän sana *мирный* tarkoittaa yksin esiintyessään rauhanomaista ja on selkeästi emotionaalisesti väritynyt ilmaisu, mutta sanaliitto *мирный договор* 'rauhansopimus' on täysin neutraali sopimustermi. N-grammihaku osoitti, että sana *мирный* 'rauhanomainen' esiintyy korpuksessa useimmiten (138 kertaa) nimenomaan neutraalissa merkityksessä eli yhdessä sanan *договор* 'sopimus' kanssa, mutta lisäksi löytyi 30 esiintymää sanaparilla *мирный + цель* 'rauhanomainen + tarkoitus'. Konkordanssihaku kyseisellä sanaparilla puolestaan osoitti, että lähes kaikki sen esiintymät liittyivät ydinenergian käyttöön (esimerkiksi *Sopimus yhteistyöstä atomienenergian rauhanomaisen käytön alalla / Соглашение [---] о сотрудничестве в области использования атомной энергии в мирных целях*). Ainoa poikkeus on sopimus yhteistyöstä tieteen ja teknologian aloilla, jossa on maininta *rauhanomaisiin tarkoituksiin liittyvää yhteistyötä tieteen ja teknologian aloilla / сотрудничество в мирных целях в области науки и техники*. Tämän analyysin periaatteiden mukaan sana *мирный* luokitellaan emotionaalisesti latautuneeksi kahdessa viimeksi mainitussa kontekstissaan eli silloin, kun se on merkityksessä *rauhanomainen*, mutta ei *rauhansopimus*-yhdyssanan osana.

Aineistossa esiintyy jossain määrin myös ekspressiivisesti latautuneita sanoja. Venäjänkielisestä luettelosta löytyvät esimerkiksi sanat *держава* ‘suurvalta’, *дружба* ‘ystävyyks’, *опасный* ‘vaarallinen’, *защита* ‘suoja’, *бедствие* ‘häätä’, *угроза* ‘uhka’ ja *добрососедский* ‘naapurisovun merkeissä’. Kaiken kaikkiaan ekspressiivisiksi tai tyyliilajille muuten epätavallisiksi määriteltävissä olevia sanoja oli 63. Vastaavasti suomenkielisessä luettelossa oli sellaisia sanoja kuin *ystävyyks*, *kunnioittavasti*, *lujittaa*, *historiallinen* yms., yhteensä 25 ekspressiivistä sanaa. Tämän perusteella voidaan siis alustavasti päätellä, että suomenkielisissä valtiosopimuksissa pitäydyttäisiin venäjänkielisiä tiukemmin tekstilajille ominaisessa tyylirekisterissä.

Tekstilajille epätyypilliset sanavalinnat antavat aiheutta selvittää, voisiko niiden syynä olla käännostoiminnasta johtuva interferenssi ja toisen kielen vaikutus. Seuraavassa on esimerkkejä tapauksista, joissa on käytetty ekspressiivisiä ilmauksia, vaikka neutraalimpikin ilmaisu olisi ollut käytettävissä:

- (1) Sopimuspuolet pyrkivät luomaan mahdollisimman **suotuisat edellytykset** yhteisyritysten perustamiselle ja toiminnalle. / Стороны будут стремиться создавать наиболее **благоприятные условия** для учреждения и деятельности совместных учреждений. (Sopimus tavaravaihdosta ja maksuista vuosina 1991-1995, 1989)

Venäjänkielinen sanaliitto *благоприятные условия* kuuluu lain ja liike-elämän viralliseen rekisteriin, ja sitä käytetään esimerkiksi Venäjän federaation laissa 52-FZ. Sen sijaan suomenkielinen vastine *suotuisat edellytykset* on huomattavasti harvinaisempi sanaliitto kuin *hyvät edellytykset*.

- (2) ... haluten vahvistaa ja edelleen kehittää maiden välisiä **ystävällisiä suhteita** ja vuorovaikutusta kulttuurin, opetuksen ja tieteellisen tutkimuksen alalla ... /... желая укреплять и далее развивать **дружественные отношения** и взаимодействие между своими странами в области культуры, образования в научно-исследовательской деятельности. (Sopimus yhteistyöstä kulttuurin, opetuksen ja tutkimuksen alalla, 1992)

Myös esimerkissä 2 esiintyvä venäjänkielinen ilmaisu *дружественные отношения* on siinä määrin tavanomainen, että siihen viitataan jopa sanakirjoissa ja kielioppaissa, kun taas suomen kielessä *ystävälliset suhteet* on melko epätavallinen sanaliitto, jota käytetään lähinnä epävirallisessa viestinnässä. Näiden esimerkkien voidaan katsoa viittaavan siihen, että kyseiset kohdat sopimusteksteistä on todennäköisesti käännetty venäjältä suomeksi ja kääntäjä on syystä tai toisesta mukailnut sanavalinnallaan venäjänkielistä versiota.

#### 4 Avainsana-analyysi

Edellisessä luvussa käsiteltiin yleisiä tendenssejä, jotka ovat havaittavissa koko korpuksen frekvenssilistoista. Kuten edellä on todettu, tämän tutkimuksen hypoteesina on, että varsinkin YYA-sopimusten aikakaudella Neuvostoliitto olisi usein ollut dominoivana osapuolena Suomen kanssa solmituissa valtiosopimuksissa. Toisin sanoen odotuksena on, että B-osakorpuksen teksteissä ilmenisi eniten merkkejä venäjältä

suomeen päin tapahtuneesta käännostoiminnasta. Tämän hypoteesin testaamiseksi olemme vertailleet PEST-korpuksen aikakausittain jaoteltuja osakorpuksia avainsana-analyysin avulla. Avainsanahaku on toiminto, joka vertaa kahden valitun tekstin tai tekstijoukon frekvenssilistoja toisiinsa ja antaa tulokseksi avainsanalistan eli listan sanoista, joiden taajuudet näissä tekstijoukoissa poikkeavat toisistaan merkittävästi. Hakutuloksista saa ensisijaisesti tietoja tekstin aiheesta sekä teksteissä mainituista paikoista ja henkilöistä.

Avainsanahaku otettiin käyttöön ensi kerran WordSmith Tools -sovelluksessa (<http://www.lexically.net/wordsmith/>). Tätä työkalua käytetään sanastotyössä ja opetuksessa (Berber Sardinha 2000), mutta sen avulla tutkitaan myös esimerkiksi käännoskielen erikoispiirteitä (ks. esim. Kemppanen 2008, Probirskaja 2009), poliittista diskurssia (Fidler & Cvrček 2015) sekä tyyli- ja rakenne-eroja eri tekstityyppien välillä (Biel 2014).

Avainsanojen haku tapahtuu pääpiirteittäin näin:

1. Yhdistetään kaksi frekvenssilistaa, joita halutaan vertailla keskenään.
2. Lasketaan frekvenssien odotetut arvot.
3. Jokaiselle sanalle lasketaan *log likelihood* -indeksi, jolla voi tarkastella erojen merkittävyyttä. Indeksillä on positiivinen, jos sana esiintyy useammin ensisijaiseksi koeaineistoksi valitussa korpuksessa, ja negatiivinen, jos sanan frekvenssi on korkeampi toisessa osakorpuksessa.<sup>2</sup>
4. Sanat, joilla on korkeimmat LL-indeksin arvot, ovat ns. avainsanoja.
5. Sanat, joilla on matalimmat LL-indeksin negatiiviset arvot, ovat negatiivisia avainsanoja.

Menetelmä ja kaavat on kuvattu yksityiskohtaisesti esim. Lancasterin yliopiston tutkijan Paul Raysonin sivuilla (<http://ucrel.lancs.ac.uk/llwizard.html>) sekä artikkelissa Rayson ym. (2004).

WordSmith Tools -sovelluksen heikkoutena on, että sillä pystyy tutkimaan vain annotoimattomia raakatekstejä. Tässä tutkimuksessa käytettävillä TextHammer-työkaluilla voidaan vertailla myös lemmatisoituja listoja. Vertailtavina ovat siis lekseemien eivätkä erillisten saneiden taajuudet, mikä parantaa tulosten luotettavuutta.

Avainsanojen lista heijastaa tutkittavien tekstien sanaston erityispiirteitä. Tulokset ovat kuitenkin aina sidoksissa vertailuaineistoon, eli ottamalla erilaisia tekstejä vertailuun voidaan saada varsin erilaisia tuloksia. Vaikka tutkijoiden tavoitteena onkin yleensä mahdollisimman neutraalin kontrolliaineiston käyttäminen, valinta on kuitenkin aina subjektiivinen. Isompi, tasapainotettu otanta voi kuitenkin parantaa tulosten ennustettavuutta. Esimerkiksi tšekkiläiset tutkijat Fidler ja Cvrček käyttävät kontrollikorpuksina sekä erikoiskielen että yleiskielen korpusta tutkiessaan Tšekkoslovakian presidentin uudenvuodentervehdysten kielen kliseisyyttä, ja näiden

---

<sup>2</sup> Indeksillä on aina positiivinen, merkki "+" tai "-" lisätään osoittamaan, missä osakorpuksessa sana esiintyy useammin.

kahden kontrollikorpuksen perusteella saadut tulokset erosivat toisistaan merkittävästi (Fidler & Cvrček 2015).

Svetlana Probirskaja (2009) vertaili väitöskirjassaan Suomi–Venäjä-sopimuksia muiden maiden kanssa tehtyihin vastaavia aiheita käsitteleviin sopimuksiin, ja hänen lähestymistapansa oli synkroninen. Tässä artikkelissa aineistoa käsitellään puolestaan diakronisesta näkökulmasta, eli vertailun kohteena ovat eri aikakausina tehdyt sopimukset. Lemmatsoituja listoja vertaillaan periaatteella osakorpus B versus osakorpus A ja osakorpus B versus osakorpus C (osakorpusjako on esitetty Taulukossa 1), ja tuloksena on kaksi avainsanalistaa.

Alla on pieni otos hakutuloksista puhdistetussa muodossa.<sup>3</sup> Sarakkeessa B on B-osakorpuksen frekvenssejä, sarakkeessa C on C-osakorpuksen frekvenssejä ja sarakkeessa LL on log likelihood -indeksin arvoja.

**Taulukko 2:** Osakorpus B vs. osakorpus C, listan kärki.

Sana	B	C	LL
raja#merkki <sup>4</sup>	1146	10	772,16
maa	1136	111	364,00
liitto	683	36	313,35
sosialistinen	669	36	305,22
neuvosto#tasa#valta	668	36	302,90
Neuvostoliitto	562	35	239,89
pituus	335	5	212,73
käyttää	945	160	169,94
kulkea	396	26	164,99
maan#tie	285	10	151,82
asettaa	403	35	142,37
rakenne	279	12	139,93
Nuija#maa	299	18	130,32
korkeintaan	171	1	118,93

<sup>3</sup> Puhdistaminen tarkoittaa, että listoista on manuaalisesti korjattu parserin tekemiä virheitä. Sanojen sijamuotojen tunnistus ja perusmuotojen (esim. substantiivien nominatiivit ja verbien infinitiivit) muodostus on vaikeaa jopa ihmisille, ja parserit tekevät melko paljon virheitä: laatu jää noin 90 %:n tasolle. Parserit eivät usein osaa valita luonnollisinta vaihtoehtoa monitulkintalisissa lauseissa, kuten *Kuusi tuli* (numeraali *kuusi* näyttää paremmalta vaihtoehdolta kuin substantiivi *kuu* possessiivisuffiksilla). Lisäksi monet sanat puuttuvat ohjelman sanastosta, joten ohjelma määrittelee lemmoja heuristisesti korkeahkolla virhemahdollisuudella. Parseri esimerkiksi rakentaa lemman *\*sopimuspuola* monille sanan *sopimuspuoli* muodoille (...ja muilla molempia *sopimuspuolia* kiinnostavilla aloilla...).

<sup>4</sup> #-merkillä jaetaan yhdyssanojen vartalog.



Sana	B	C	LL
pitkin	170	5	94,63
SNTL	153	3	92,79
raja	567	111	81,25
metri	140	5	75,05
suomalainen	190	15	72,52
Suomi	1880	592	67,67

Odotetusti lista kertoo paljon B-korpuksen aihepiireistä:

- ”Neuvosto-aika”: *Neuvostoliitto, neuvostotasavalta, sosialistinen, SNTL*
- ”Valtioiden rajat”: *rajamerkki, Nuijamaa (rajanylityspaikka), raja, linja*; paljon mittayksiköitä ja maamerkkejä, kuten *metri, maantie* yms.

Listassa on kuitenkin myös yllättäviä tuloksia. Miksi sanojen *maa, kulkea* ja *asettaa* frekvenssit ovat näin korkeita B-korpuksessa? Tämä voi mahdollisesti johtua joistakin toistuvista ilmaisuista. Tällaisiin tapauksiin kannattaa paneutua lähemmin muiden työkalujen avulla.

Tutkitaan esimerkkitapauksena sanaa *maa*. Ensinnä haetaan sanan kollokaatteja eli sanoja, jotka esiintyvät usein sen lähikontekstissa. Taulukossa 3 on lueteltu *maa*-sanalle lemmakollokaatit B-osakorpuksessa.

**Taulukko 3:** Sanan *maa* kollokaatit B-osakorpuksessa ( $F \geq 60$ ).

Sana	L3 <sup>5</sup>	L2	L1	R1	R2	R3	Summa	LL
käyttää	1	3	0	661	5	4	674	757,13
raja#merkki	1	0	0	0	659	3	663	498,25
N:o	5	0	0	0	0	635	640	396,37
pituus	0	0	209	0	0	0	209	213,05
raja#viiva	0	206	0	0	0	0	206	119,13
kivikkoinen	0	0	62	0	0	0	62	74,39
kolmas	2	0	70	0	1	1	74	66,93
keski#viiva	0	81	5	0	0	0	86	63,61
molemmat	0	0	113	2	0	0	115	59,49
joka	5	6	5	0	69	3	88	34,1

<sup>5</sup> L<sub>x</sub> = frekvenssi x:n etäisyydellä vasemmalle, R<sub>x</sub> = frekvenssi x:n etäisyydellä oikealle.

Taulukosta näkyy, että valtaosa konteksteista on muodoissa ”maa + käyttää + rajamerkki + N:o” (yli 640) ja ”rajaviiva + pituus + maa” (yli 200). Seuraavaksi katsotaan käyttöesimerkkejä eli tehdään konkordanssihaku hakusanoilla ”maa” ja ”käyttää”. Alla muutamia tyypillisiä konteksteja:

- (5) Vuokrattujen alueiden rajojen merkitsemistä varten on asetettu 229 rajamerkkiä, joista **maalla käytettäviä rajamerkkejä** 213 kappaletta ja vesillä käytettäviä rajamerkkejä 16 kappaletta. (Saimaan kanavan ja Malyj Vysotskij-saaren<sup>6</sup> rajojen rajankuvauspöytäkirja, 1964)
- (6) **Rajaviivan pituus maalla käytettävien rajamerkkien N:o 11 ja N:o 12 välillä** on 158.6 m. (Saimaan kanavan ja Malyj Vysotskij-saaren rajojen rajankuvauspöytäkirja, 1964)
- (7) Suuntakulma **maalla käytettävästä rajamerkistä N:o 4** rajaviivan taitepisteeseen on 117°34'. (Saimaan kanavan ja Malyj Vysotskij-saaren rajojen rajankuvauspöytäkirja, 1964)

Konkordanssihaulla selviää myös se, että tässä tapauksessa kaikki hakutulokset ovat peräisin samasta sopimuksesta, rajankuvauspöytäkirjasta, joka on melko pitkä (17 239 sanetta) ja sisältää tarkkoja rajamerkkien kuvauksia. Näin ollen myös sanat *maa*, *kulkea* ja *käyttää* kuuluvat edellä mainittuun valtioiden rajat -aihepiiriin. Tämä tapaus on myös hyvä esimerkki siitä, että yksikin teksti voi vaikuttaa tilastoihin, jos korpus ei ole iso. Siksi kannattaa aina tarkastella myös sanojen esiintymiä teksteissä.

Analysoiduista avainsanalistoista löytyi monta sanaa, jotka olivat jo esillä tämän artikkelin luvussa 3: *ystävyyys*, *yhteistoiminta*, *lujittaa* jne. Havainnot tyylilajille epätavallisen sanaston käytöstä tuntuvat siis tosiaan painottuvan B-osakorpuksen eli YYA-aikakauteen. Seuraavassa tutkitaan joitain näistä sanoista lähemmin.

### **Ystävyys**

Sana *ystävyyys* on todella tyypillinen B-osakorpuksessa: sen frekvenssi on 75 eli 0,65 esiintymää 1000 sanetta kohti. Vastaavasti sillä on A-korpuksessa vain neljä esiintymää ja C-osakorpuksessa yksi. Venäjänkielisissä teksteissä sana *дружба* ‘ystävyyys’ on todella tavallinen, ja se esiintyy yleisesti sopimusten otsikoissa ja johdannoissa. Sen yleisin R2-kollokaatti on *сотрудничество* ‘yhteistoiminta’, eli tyypillinen konteksti on *дружба и сотрудничество* ‘ystävyyys ja yhteistoiminta’. Vastaavasti suomalaisen *ystävyyys*-sanan kollokaatteja ovat odotetusti *yhteistoiminta* ja *keskinäinen*.

- (8) Korkeat Sopimuspuolet vakuuttavat päättäneensä toimia yhteistyön ja **ystävyyden** hengessä Suomen ja Neuvostoliiton taloudellisten ja kulttuurisuhteiden jatkuvaksi kehittämiseksi ja lujittamiseksi. / Высокие договаривающиеся стороны подтверждают свою решимость действовать в духе сотрудничества и **дружбы** в целях дальнейшего развития и укрепления экономических и культурных связей между Советским Союзом и Финляндией. (YYA-sopimus, 1948)

---

<sup>6</sup> Kielenhuollon vastainen kirjoitusasu on peräisin alkuperäisestä asiakirjasta.

PEST-korpuksen teksteissä on myös paljon esiintymiä adjektiivilla *ystävällinen* (32 ja 0,28) ja yhdyssanoilla, joissa on *ystävyy*-alkuinen vartalo (esim. *ystävyyssuhde*, *ystävyyssseura* yms. ja jopa *ystävyysskuukausi*).

- (9) lujittaakseen edelleen molempien maiden keskisiä **ystävällisiä** suhteita / В целях дальнейшего укрепления **дружественных** отношений между обоими государствами (Konsulisopimus, 1966).

### ***Yhteistoiminta***

Myös sana *yhteistoiminta* on tavallinen B-osakorpuksessa (103/0,90) ja harvinainen muissa osakorpuksissa. Venäjänkielisissä versioissa vastaavissa konteksteissa esiintyy sana *сотрудничество*. Tämä sana oli todella tavallinen neuvostoaikana, mutta nykyään sen frekvenssi on laskussa, erityisesti ilmaisuihin kuten edellä mainittu *дружба и сотрудничество* ‘ystävyy ja yhteistoiminta’.

- (10) Suomen Tasavallan hallituksen toivomukset vartenottaen, ilmaissut olevansa valmis ryhtymään **yhteistoimintaan** poronhoitoa raja-alueilla koskevissa kysymyksissä... / ...идя навстречу пожеланиям Правительства Финляндской Республики, выразило свою готовность к **сотрудничеству** по вопросам оленеводства в пограничных районах... (Porosopimus, 1965)

On huomionarvoista, että C-korpuksessa yhteistoimintaa tarkoittavan venäläisen *сотрудничество*-sanan suomenkielisenä vastineena käytetään yleisimmin sanaa *yhteistyö*. Taulukossa 5 näkyy, että *yhteistyö*-sanan frekvenssi on viisinkertaistunut ja samalla *yhteistoiminta*-sanan frekvenssi on laskenut selvästi.

**Taulukko 5:** Sanat *yhteistoiminta* ja *yhteistyö* PEST:n eri osakorpuksissa<sup>7</sup>

	<b>A</b>	<b>B</b>	<b>C</b>
yhteistoiminta	3 (0,04)	103 (0,90)	6 (0,11)
yhteistyö	0	241 (2,09)	185 (3,52)

Selittykö tämä ilmiö suomen kielen yleisellä kehityksellä, vai onko sanaa *yhteistoiminta* pidetty niin leimallisena nimenomaan yhteistoiminnalle Neuvostoliiton kanssa, että se on painunut historiaan yhdessä Neuvostoliiton kanssa? Vertailuhaku Suomen Kielipankin korpuksista (korp.csc.fi) osoittaa, että sanan *yhteistoiminta* käyttö suomenkielisissä aineistoissa on tosiaan vähentynyt merkittävästi 1990-luvulta lähtien, mikä tukisi tätä olettamusta. Yksi mahdollinen selitys on myös se, että nykysuomessa sana *yhteistoiminta* on vahvasti assosioitunut työoikeuteen (ja erityisesti negatiivisessa mielessä *yhteistoiminta-* eli *yt-neuvotteluihin*). Ruotsi–suomi-tekstikorpuksen kokoaminen antaa todennäköisesti mielenkiintoista lisätietoa tähänkin kysymykseen.

<sup>7</sup> Suluissa frekvenssi per 1000 sanetta.

### ***Kehittää ja lujittaa***

Neuvostoajan venäjänkielissä asiakirjoissa oli käytössä sellaisia vakioilmauksia kuin *дальнейшее укрепление* ‘edelleen lujittaminen’, *дальнейшее развитие* ‘edelleen kehittäminen’ ja jopa *дальнейшее развитие и укрепление* ‘edelleen kehittäminen ja lujittaminen’ tai *дальнейшее укрепление и развитие* ‘edelleen lujittaminen ja kehittäminen’. Sen lisäksi, että *lujittaa* on sopimuskontekstissa selkeän metaforinen ja ekspressiivisesti latautunut termi, suomen kielessä myös *edelleen*-adverbin käyttö tällaisessa kontekstissa voidaan luokitella tyyllilajille vieraaksi ilmaisutavaksi ja viitteeksi venäjän kielen interferenssistä.

Ilmaisuja *edelleen kehittäminen* tai *edelleen lujittaminen* voidaan pitää esimerkkeinä konstruktivistisista metaforista. Konstruktivistisen metaforateorian mukaan metaforinen ajattelu ei ole pelkästään kielen tason ilmiö vaan monet ihmisen ajattelua ja toimintaa ohjaavat käsitteelliset perustuvat alitajuisesti käyttämiimme metaforisiin järjestelmiin (Lakoff & Johnson 1980: 3). Näin ollen sanan *edelleen* käyttö tässä tarkastelluissa konteksteissa liittyy metaforiseen käsitejärjestelmään ”tulevaisuus on parempi” – ”tulevaisuutta kohti suuntautuva toiminta on parantamista”. Aineistosta on löydettävissä muitakin neuvostoajan metaforia, jotka ovatkin mielenkiintoinen erillisen tutkimuksen aihe.

Oheisen esimerkin lisäksi myös aiemmin annettussa esimerkissä (8) esiintyvät nämä molemmat verbit ja myös edellä mainittu *ystävyyys*.

- (11) ... haluten edistää Suomen ja Neuvostoliiton välille muodostuneiden ystävällisten ja hyvien naapuruussuhteiden **edelleen kehittämistä ja lujittamista**... / ... желая содействовать **дальнейшему развитию и укреплению** сложившихся дружественных и добрососедских отношений между Финляндией и СССР... (Pöytäkirja YVA-sopimuksen pidentämisestä, 1955)

Mainitut ilmaukset ovat vahvasti esillä B-osakorpuksessa, mikä stimuloi mm. *lujittaa*- ja *kehittää*-verbien ja niiden johdannaisten käyttöä. Silti vain *lujittaa*-sana on esillä molemmissa avainsanojen listoissa ja esiintyy B-osakorpuksessa paljon useammin kuin muissa osakorpuksissa. Sana *kehittää* esiintyy C-osakorpuksessa useammin kuin B-osakorpuksessa. Kollokaattien haku paljastaa, että C-osakorpuksessa (taulukko 7) tätä sanaa ei käytetä samoissa ilmaisuissa kuin B-osakorpuksessa (taulukko 6), ja jopa *kehittää* + *yhteistyö* -ilmaus on siinä vähemmän merkittävä.

**Taulukko 6:** *Kehittää*-sanan kollokaatit B-osakorpuksessa

Sana	L5	L4	L3	L2	L1	R1	R2	R3	R4	R5	Summa	LL
yhteis#työ	0	1	3	3	31	2	5	5	4	1	55	27,47
edelleen	0	0	1	0	25	4	0	2	0	0	32	17,58
teollinen	0	12	0	10	0	0	0	0	0	1	23	15,02
syventää	0	0	0	0	0	0	27	0	2	0	29	13,09

olla	0	1	1	0	1	0	1	0	2	1	7	7,52
tieteellis#tekninen	0	0	0	12	0	0	0	0	0	0	12	5,44
haluta	1	0	1	7	4	0	0	0	0	0	13	5,3
lujittaa	0	0	0	1	0	0	9	1	0	0	11	5,27

**Taulukko 7:** *Kehittää*-sanon kollokaatit C-osakorpuksessa

Sana	L5	L4	L3	L2	L1	R1	R2	R3	R4	R5	Summa	LL
ja	5	7	1	9	8	11	7	5	4	3	60	29,4
yhteis#työ	1	0	0	2	3	2	4	5	0	4	21	7,75

Avainsanalistoista käy ilmi, että osakorpusten B ja A välisessä vertailussa *edelleen* on negatiivinen avainsana, kun taas osakorpusten B ja C vertailussa se on positiivinen. Kollokaattien haku osakorpuksista paljastaa, että B-osakorpuksessa *edelleen*-sanon lähikonteksteissa esiintyvät verbit *kehittää*, *lujittaa*, *ehdottaa* ja *pyrkii*. A-osakorpuksesta ei saa yhtä selkeää kuvaa, ja konkordanssien tarkastelu paljastaa, että siinä melkein kaikki *edelleen*-sanon kontekstit ovat rajasopimuksista, joissa sanaa käytetään valtioiden rajan kuvauksessa konkreettisesti merkityksessä ‘eteenpäin’.

- (13) Rajamerkistä N:o III/ 277/ 577 raja kulkee vanhaa Suomen ja Neuvostoliiton välistä rajaa pitkin koilliseen rajamerkkiin k N:o 578 ja **edelleen** rajamerkkiin N:o 775. / От пограничного знака № III/277/577 граница идет по старой Советско-Финляндской границе в северо-западном направлении к пограничному знаку № 578 и далее до пограничного знака № 775. (Rajankuvaus - pöytäkirja valtakunnanrajan kulusta, 1940)

## 5 Päätelmät

Tämän pienimuotoisen leksikaalisen analyysin perusteella voidaan todeta, että valtiosopimusten kieli kertoo monenlaisia asioita. Kielelliset seikat voivat antaa viitteitä sopimusosapuolten suhteista ja toisen osapuolen mahdollisesta dominoivasta vaikutuksesta. Tällainen havainto on esimerkiksi sanan *ystävyyks* runsas esiintyminen Neuvostoliiton ajan valtiosopimuksissa sekä suomenkieliselle viralliselle diskurssille epätyypillisten sanojen ja ilmaisujen, kuten *suotuisat edellytykset* käyttö. Lisäksi voidaan nähdä piirteitä kielen yleisestä kehityksestä, mistä esimerkkinä on *yhteistoiminta*-sanon korvautuminen *yhteistyö*-sanalla.

Sanaston analyysi on vain yksi tapa lähestyä aineistoa; esimerkiksi eri sijamuotojen käyttö, lausepituudet, sivulauseiden käyttö ja sanajärjestys voivat antaa lisää tietoa käännöstoiminnan osuudesta. Myös vertailu Suomien ja Ruotsin välisiin sopimuksiin antaisi lisää tietoa siitä, ovatko tämän artikkelin havainnot spesifejä nimenomaan venäjän- ja suomenkielisille sopimuksille vai liittyvätkö ne yleisesti valtiosopimusgenreen. PEST-korpusta kartutetaan koko ajan ja uusia hakutoimintoja

kehitetään, joten jatkossa tämän artikkelin aiheisiin voidaan vielä palata laajalaisemman analyysin merkeissä.

## Lähteet

### Tutkimusaineisto

PEST, valtiosopimusten rinnakkaistekstikorpus. Tampere: Tampereen yliopisto 2016. Saatavissa: [mustikka.uta.fi/texthammer](http://mustikka.uta.fi/texthammer) (käyttäjätunnus haettavissa Mikhail Mikhailovilta, [mikhail.mikhailov@staff.uta.fi](mailto:mikhail.mikhailov@staff.uta.fi)).

### Kirjallisuuslähteet

Baker, Mona 1996. Corpus-Based Translation Studies: The Challenges that Lie Ahead. Teoksessa: Harold Somers (ed.) *Terminology, LSP and Translation*. Amsterdam: John Benjamins, 175–187.

Baker, Mona 2004. A Corpus-Based View of Similarity and Difference in Translation. *International Journal of Corpus Linguistics* 9:2, 167–193.

Berber Sardinha, Tony 2000. Comparing Corpora with WordSmith Tools: How Large Must the Reference Corpus Be? Teoksessa: *The Workshop on Comparing Corpora*. Hong Kong: Association for Computational Linguistics, 7-13. Saatavissa: <http://www.aclweb.org/anthology/W00-0902> [viitattu 24.2.2017].

Biel, Lucja 2014. *Lost in the Eurofog: The Textual Fit of Translated Law*. Bern: Peter Lang.

Fidler, Masako & Václav Cvrček 2015. A Data-Driven Analysis of Reader Viewpoints: Reconstructing the Historical Reader Using Keyword Analysis. *Journal of Slavic Linguistics* 23:2, 197–239.

Kemppanen, Hannu 2008. *Avainsanoja ja ideologiaa: käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi*. Väitöskirja. Joensuu: Joensuun yliopisto.

Lakoff, George & Mark Johnson 1980. *Metaphors We Live By*. Chicago & London: University of Chicago Press.

Probirskaja, Svetlana 2009. *Rajankäyntiä: Suomen ja Venäjän kahdenväliset valtiosopimukset käännöstieteellisen avainsana-analyysin valossa*. Väitöskirja. Tampere: Tampere University Press.

Ramos, Fernando Prieto 2015. El traductor como redactor de instrumentos jurídicos: el caso de los tratados internacionales. *Journal of Specialised Translation* 15, 200–214 Saatavissa: [http://www.jostrans.org/issue15/issue15\\_toc.php](http://www.jostrans.org/issue15/issue15_toc.php) [viitattu 24.2.2017].

Rayson Paul, Berridge Damon & Francis Brian 2004. Extending the Cochran Rule for the Comparison of Word Frequencies Between Corpora. Teoksessa: G. Purnelle, C. Fairon & A.

Dister (eds.) *Le poids des mots: Proceedings of the 7th International Conference on Statistical Analysis of Textual Data (JADT 2004)*, Louvain-la-Neuve, Belgium, March 10–12, 2004. Presses universitaires de Louvain, vol. 2, 926–936. Saatavissa: [http://ucrel.lancs.ac.uk/people/paul/publications/rbf04\\_jadt.pdf](http://ucrel.lancs.ac.uk/people/paul/publications/rbf04_jadt.pdf) [viitattu 24.2.2017].

Schäffner, Christina 1997. Strategies of Translating Political Texts. Teoksessa: A. Trosborg (toim.) *Text Typology and Translation*. Amsterdam/Philadelphia: John Benjamins, 119–143.

Toury, Gideon 2012. *Descriptive Translation Studies – and Beyond: Revised edition (2)*. Philadelphia: John Benjamins.

Valtiosopimusoikeutta koskeva Wienin yleissopimus. Valtiosopimukset 33/1980.

### **Kirjoittajien esittely ja yhteystiedot**

**Mikhail Mikhailov** on käännöstieteen professori (suomi ja venäjä) Tampereen yliopistossa. Hänen tutkimusintresseihinsä kuuluvat korpuslingvistiikka, erityisesti rinnakkaistekstikorpuukset, ja käännösteknologia.

Sähköpostiosoite: Mikhail.Mikhailov (at) staff.uta.fi

**Miia Santalahti** on kääntämisen ja tulkkauksen yliopisto-opettaja, kääntäjä ja väitöskirjatutkija. Hänen väitöskirjansa aiheena on käännöstoiminnan vaikutus valtiosopimusten kieleen.

Sähköpostiosoite: Miia.Santalahti (at) staff.uta.fi