

## Miksi ettei, mutta ei siksei?

- Partikkelin ja kieltoverbin syntagmaattiset sulaumat ruututeksteissä

Heli-Maaria Luoma  
Tampereen yliopisto  
Kieli-, käännös- ja kirjallisuustieteiden yksikkö  
Monikielisen viestinnän ja käännöstieteen maisteriopinnot  
Pro gradu -tutkielma  
Marraskuu 2017



Tampereen yliopisto  
Monikielisen viestinnän ja käännöstieteen maisteriopinnot  
Kieli-, käännös- ja kirjallisuustieteiden yksikkö

LUOMA, HELI-MAARIA: Miksi ettei, mutta ei siksei? - Partikkelin ja kieltoverbin sulaumat ruututeksteissä

Pro gradu -tutkielma, 81 sivua + 11 liitesivua + englanninkielinen lyhennelmä 10 sivua  
Marraskuu 2017

---

Tämän tutkimuksen tarkoituksena oli tarkastella, miten suomen kieltoverbi *ei* yhdistyy syntagmaattiseksi sulaumaksi konjunktioiden *että*, *jos*, *jotta*, *mutta* ja *vaikka* sekä partikkeleiden *ehkä*, *miksi* ja *siksi* kanssa ruututeksteissä. Syntagmaattiset sulaumat ovat yhdeksi sanaksi hahmotettavia kahden sanavartalon muodostamia kokonaisuuksia, joilla on sanankaltaisia ominaisuuksia. Tutkimuksen päätavoitteena oli tutkia syntagmaattisten sulaumien esiintymistä ruututeksteissä ja vertailla tuloksia alkuperäis-suomessa ja käännös-suomessa ilmeneviin sulaumamuotoihin. Tutkimuksen toissijaisena tavoitteena oli tarkastella, miten paljon ruututeksteissä esiintyy syntagmaattisten sulaumien erilleen kirjoitettuja muotoja verrattuna syntagmaattisiin sulaumiin. Syntagmaattisten sulaumien ilmenemistä ruututeksteissä tarkasteltiin Andrew Chestermanin käännösnormien, ja etenkin odotus- ja suhdenormien kautta. Ruututekstit saatiin YLE-ruututekstikorpukselta lohkaistusta osakorpukselta, joka piti sisällään yli 26 miljoonaa sanaa. Kirjallisen suomen vertailumateriaalina taas toimi Käännös-suomen korpus.

Odotusnormin käsitettä on sovellettu käsittämään suomalaisen katsojakunnan odotusta siitä, että tekstityksissä käytettäisiin samalla tavoin suomea kuin muissakin suomen varianteissa. Mikäli siis ruututeksteissä esiintyisi samoissa määrin sulaumamuotoja kuin muissakin suomen varianteissa, voitaisiin sanoa tekstitysten noudattavan tätä odotusnormia. Mikäli taas sulaumamuotoja ilmenisi ruututeksteissä selkeästi enemmän kuin muissa suomen varianteissa, voitaisiin arvella tekstitysten noudattavan pikemminkin Chestermanin suhdenormia kuin odotusnormia. Tämän tutkimuksen oletuksena oli, että ruututekstejä käännettäessä suhdenormin vaikutus olisi odotusnormia suurempi av-kääntämisen rajoitteiden tähden, eli tutkimuksen hypoteesina oli, että syntagmaattisia sulaumamuotoja esiintyisi enemmän ruututeksteissä kuin muissa suomen varianteissa. Kaikki vertailut suoritettiin laskemalla ensin kielellisten elementtien kaikki esiintymät ja muuttamalla esiintymien määrät suhteellisiksi frekvenssiluvuiksi.

Ruututeksteissä ilmeni selkeästi enemmän syntagmaattisia sulaumia kuin alkuperäis-suomessa ja käännös-suomessa, ja täten kaiken kaikkiaan käännettyssä suomessa supisuomeen verrattuna. Kokonaisuutena tutkielman hypoteesi piti siis paikkaansa ja täten voisi todeta, että syntagmaattisten sulaumien osalta ruututekstien kieli eroaa muista suomen varianteista. Lisäksi sulaumamuotoja voidaan pitää niin sanotun tekstittäjän suomen yhtenä erityispiirteinä. Kuitenkin on syytä muistaa, että osakorpuksen materiaaleista valtaosa on elokuvadialogia, eli puhetta. Sulaumamuoto on voinut valikoitua tekstityksiin myös siksi, että sillä on pyritty saamaan tekstitykseen yleispuhekielen kaltaista puhekielimäisyyttä.

Avainsanat: partikkeli, konjunktio, kieltoverbi, syntagmaattinen sulauma, tekstitys, normi, YLE-ruututekstikorpus, Käännös-suomen korpus



# Sisällysluettelo

Tiivistelmä.....	
1 JOHDANTO.....	1
2 SYNTAGMAATTISET SULAUMAT.....	6
3 RUUTUTEKSTIT JA NORMIT .....	10
3.1 Käännösnormit .....	11
3.2 Ruututekstien rajoitteet ja konventiot .....	15
3.3 Tutkimusasetelma .....	18
4 KORPUSTUTKIMUS .....	20
4.1 Konkordanssiohjelmien toimintoja.....	21
4.2 Yle-ruututekstikorpus audiovisuaalisen kääntämisen tutkimuksessa.....	22
5 AINEISTO JA MENETELMÄT.....	27
5.1 Yle-ruututekstikorpus.....	28
5.2 Käännössuomen korpus.....	30
5.3 Osakorpus.....	31
5.3.1 Osakorpuksen koostaminen .....	31
5.3.2 Osakorpuksen tutkiminen .....	34
5.3.3 Vertailut .....	40
6 TULOKSET.....	44
6.1 Osakorpuksen tulokset.....	45

6.1.1 Syntagmaattiset sulaumat .....	45
6.1.2 Erilleen kirjoitetut muodot .....	55
6.1.3 Osakorpuksen tulosten vertailu .....	58
6.2 Kolmen korpuksen vertailun tulokset.....	62
6.2.1 Syntagmaattiset sulaumat korpuksissa.....	62
6.2.2 Syntagmaattisten sulaumien päätteelliset muodot .....	69
6.2.3 Syntagmaattisten sulaumien vokaalisoinnun muutokset .....	71
7 YHTEENVETO JA LOPPUPÄÄTELMÄT .....	73
LÄHTEET .....	82
LIITE I: SYNTAGMAATTISTEN SULAUMIEN HAKUSANALISTA .....	87
LIITE II: ERILLEEN KIRJOITETTUIJEN MUOTOJEN HAKUSANALISTA .....	89
LIITE III: SYNTAGMAATTISTEN SULAUMIEN FREKVENSSEIT .....	91
LIITE IV: ERILLEEN KIRJOITETTUIJEN MUOTOJEN SUHTEELLISET FREKVENSSEIT .....	93
LIITE V: SYNTAGMAATTISTEN SULAUMIEN PROSENTUAALISET ESIINTYMÄT LYHYISSÄ LAUSEISSA .....	94
LIITE VI: ERILLEEN KIRJOITETTUIJEN MUOTOJEN PROSENTUAALISET ESIINTYMÄT LYHYISSÄ LAUSEISSA .....	96
ENGLISH SUMMARY .....	i



# 1 JOHDANTO

Omassa työssäni av-kääntäjänä olen monesti joutunut miettimään keinoja sanoa tarkoittamani asia lyhyemmin. Pienikin muutos sanoissa ja virkkeissä käytettyjen merkkien määrässä voi usein olla merkittävä ruututekstin kannalta. Merkkimäärän vaihdos saattaa vaikuttaa niin ajan kuin tilankin rajoitteeseen pidentäen tai lyhentäen lukuaikaa tai ehkäpä helpottamalla ohjelmassa sanotun asian mahdollistamista kahdelle tekstiriville. Syntagmaattiset sulaumat, eli yhdeksi sanaksi hahmotettavat kahden sanavartalon muodostamat kokonaisuudet (ISK 2005, 163) on yksi tällainen merkkejä säästävä keino. Syntagmaattiset sulaumat ovat yhdyssanojen kaltaisia yhdistelmiä, joiden eri osilla on lauseessa omat tehtävänsä, esimerkiksi *emmä 'en minä'* (mts.). Itseäni on kuitenkin jäänyt mietityttämään, miten yleisesti syntagmaattisia sulaumia käytetään ruututekstityksissä sekä miten luettavia ne ovat katsojan kannalta. Onhan katsojan ymmärrettävä ruututekstin sanoma ja merkitys vain siinä pienessä hetkessä, jonka ruututeksti on television ruudulla (Vertanen 2008, 153).

Riitta Jääskeläinen (2008, 117-118) huomauttaa, että valtaosa ruututekstejä käsittelevästä tutkimuksesta on keskittynyt vain käyttämään ruututekstejä materiaalina, esimerkiksi kulttuurisidonnaisten ilmaisujen tutkimiseen, eikä varsinaisena tutkimuksen kohteena. Itse olen katsojana ja etenkin työssäni av-kääntäjänä tottunut tarkkailemaan suomen kieltä ja sen oikeakielisiä rakenteita sekä harjaantunut myös havainnoimaan televisiossa näkyviä ruututekstejä ihan sellaisenaan, tavallaan omana erityisenä muotona suomen kielestä, niin sanottuna tekstittäjän suomena. Oma tutkimukseni pyrkii omalta osaltaan vastaamaan kysymykseen ruututekstien kielen eriävyydestä ja erityispiirteistä. Esimerkiksi Sonja Tirkkonen-Conditin ja Jukka Mäkisaloon (2007) tutkimus antaa syytä olettaa, että ruututekstien kieli eroaisi muista suomen varianteista käyttämällä hyväkseen suppeaa ja ytimekästä kielenkäyttöä hyödyntäviä koheesiokeinoja. Lisäksi ruututekstitysten kielen omalaatuisuutta on Suomessa tutkittu vielä verrattain vähän, joten välillä on perusteltua keskittyä tekstityksiin ilman muita multimodaalisen kokonaisuuden osia. Yleensä ottaen tv-tekstityksiä tutkittaessa tulisi kuitenkin kiinnittää huomiota sanan, liikkuvan kuvan ja äänen muodostamaan multimodaaliseen kokonaisuuteen, eli verbaalista ja ei-verbaalista sekä auditiivista ja visuaalista yhdistävään materiaaliin, irrottamatta sitä kontekstistaan ja kenties unohtamatta sen tekijöitä ja vastaanottajia (Gambier 2008, 86). Tässä tutkimuksessa on keskitytty kielen mikrorakenteisiin, joihin oletettavasti kuvallinen informaatio ei niinkään vaikuta. Tällaisista



lähtökohdista nousi vähitellen kielitieteellinen kiinnostus ruututeksteissä ilmeneviä suomen kielen mikrorakenteita kohtaan – tehtyäni siis niistä havaintoja niin katsojana, lukijana kuin kääntäjänäkin.

Tämän tutkimuksen kohteeksi valikoituivat siis syntagmaattiset sulaumat ja nimenomaan kieltoverbin ja partikkelien yhdistymät, kuten *etten*, *mikset* ja *vaikkei*. Tämän tutkimuksen tavoitteena onkin tarkastella, miten kieltoverbi yhdistyy syntagmaattiseksi sulaumaksi konjunktioiden *että*, *jos*, *jotta*, *mutta* ja *vaikka* kanssa sekä ennen kaikkea, miten paljon näitä sulaumamuotoja esiintyy suomenkielisissä ruututeksteissä. Tutkimukseen on otettu mukaan myös partikkelien *ehkä*, *miksi* ja *siksi* vakiintuneet syntagmaattiset sulaumat, sillä yhdistyessään kieltoverbin kanssa ne käyttäytyvät kielellisesti samoin kuin aiemmin mainitut sulaumakonjunktiot (ISK 2005, 164). Lisäksi tässä tutkimuksessa havainnoidaan samojen kielellisten elementtien erilleen kirjoitettujen muotojen, esimerkiksi *että en*, *miksi et* ja *vaikka ei*, ilmenemistä ruututeksteissä ja näitä tuloksia verrataan sulaumamuodoista saataviin tuloksiin. Vertailun tarkoituksena on selvittää, käytetäänkö ruututeksteissä enemmän vai vähemmän sulautuneita muotoja kuin niiden erilleen kirjoitettuja muotoja. Tietävästi tällaista tutkimusta näiden kielen mikrorakenteiden vertailusta ei ole aiemmin ruututekstien yhteydessä toteutettu. Tekstityksistä on toki ennenkin tarkasteltu suomen kielen mikrorakenteita: esimerkiksi aiemmin mainitussa Tirkkonen-Conditin ja Mäkisalonen (2007) tutkimuksessa tarkasteltiin ruututeksteistä suomen kielen koheesiokeinoja, diskurssipartikkeleita sekä liitepartikkeleita (tutkimuksesta tarkemmin luvussa 4.2).

Tutkimuksen lähtökohtana toimivat Andrew Chestermanin käänösnormit, joita esitellään tarkemmin alaluvussa 3.1. Normit ovat sääntöjä ja ohjeita, jotka voivat käännöstieteessä ohjata niin kääntäjää, käännösprosessia kuin käännöstäkin (esimerkiksi Toury 2012 ja Chesterman 1997). Tämän tutkimuksen keskiössä ovat niin Chestermanin odotusnormit kuin suhdenormitkin. Odotusnormit määrittävät käännökselle asetettuja odotuksia ja vaatimuksia sekä ilmentävät sitä, millainen jonkin käännöksen on sen lukijoiden tai tilaajan odotusten mukaan oltava kyseisenä aikana (Chesterman 2004, 345; 2007, 361). Suhdenormien kohdalla ”kääntäjän oletetaan työskentelevän sillä tavoin, että lähdetekstin ja kohdetekstin välillä syntyy ja säilyy tarkoituksenmukainen suhde” (Chesterman 2004, 345), joskin tuo suhde riippuu monesta seikasta ja päätös suhteen tarkoituksenmukaisuudesta jää kääntäjälle (Chesterman 1997, 69).

Chestermanin kahta normia sovelletaan tässä työssä siis syntagmaattisten sulaumien esiintymiseen ruututeksteissä. Odotusnormin käsitettä on sovellettu käsittämään suomalaisen katsojakunnan odotusta siitä, että suomenkielisissä ruututeksteissä käytetään samalla tavoin suomea kuin muissakin suomen varianteissa – onhan suomalaisilla vahvat mielipiteet ruututeksteistä, etenkin mikäli niissä ilmenee heidän mielestään virheitä (esimerkiksi Tuominen 2012 ja Mäyränpää 2010). Mikäli siis ruututeksteissä esiintyy samoissa määrin sulaumamuotoja kuin muissakin suomen varianteissa, voidaan sanoa tekstitysten noudattavan tätä odotusnormia. Mikäli taas sulaumamuotoja ilmenee ruututeksteissä selkeästi enemmän kuin muissa suomen varianteissa, voidaan arvella tekstitysten noudattavan pikemminkin Chestermanin suhdennormia kuin odotusnormia. Onhan av-kääntäjien tehtävä työtään ruututekstikäntämisen rajoitteiden ja konventioiden puitteissa (näitä käsitellään tarkemmin alaluvussa 3.2), ja täten tiivistyskeinojen luulisi nousevan kielellistä vastaavuutta tärkeämmäksi, mikäli vain käytetty kieli on muutoin oikeaoppista, sujuvaa ja ymmärrettävää. Tämän tutkimuksen oletuksena on, että ruututekstejä käännettäessä suhdennormin vaikutus olisi odotusnormia suurempi juurikin av-kääntämisen rajoitteiden tähden, eli tutkimuksen hypoteesina on, että syntagmaattisia sulaumamuotoja esiintyy enemmän ruututeksteissä kuin muissa suomen varianteissa. Yves Gambier (2008, 94) onkin muistuttanut, että tärkeää on myös tarkastella ”vallitsevan kirjallisen normin ja kirjoitetun tekstin konventioiden suhdetta tekstitystekstien kirjoitettuun kieleen”, joskin tässä tutkimuksessa on hyödynnetty normien käsitettä. Normien ja konventioiden suhdetta on käsitelty paremmin luvussa 5. On kuitenkin huomautettava, että nämä kaksi normia eivät tietenkään poissulje toisiaan, mikäli ruututeksteissä ja muissa suomen varianteissa ilmenee yhtäläillä syntagmaattisia sulaumia.

Tämän tutkimuksen varsinainen tutkimusaineisto saatiin Joensuun yliopistossa (nykyisessä Itä-Suomen yliopistossa) koostetusta YLE-ruututekstikorpuksesta (lyhyemmin YLE-korpus), joka sisältää lähes 100 miljoonaa sanaa sekä kaikki Ylen kanavilla esitetyt ruututekstit vuosilta 1991–2004 (Mäkisalo 2006, 18). Ruututekstikorpuksesta koottiin alkujaan kolme erillistä versiota, joista yksi, ruututekstien ajatuskoodit ja oheistiedot sisältävä versio (mp.) valittiin tämän tutkimuksen materiaaliksi. Tutkimuksen alkuperäisenä tarkoituksena oli myös arvioida, onko ruututeksteille varattu aika mahdollisesti vaikuttanut kääntäjän ratkaisuun käyttää tai olla käyttämättä sulaumamuotoa jonkin tietyn ruututekstrivin kohdalla. Tämä ajallisuuden aspekti kuitenkin jätettiin lopulta pois tästä tutkimuksesta sen vaatiman työmäärän tähden. Koska ruututekstikorpuksen sisältämä materiaali on hyvin laaja,

muodostettiin sen vuosien 2000–2004 materiaaleista tätä tutkimusta varten pienempi sivukorpus, johon tässä tutkimuksessa viitataan nimellä *osakorpus*, jotta se erottuisi varsinaisesta ruututestikorpuksesta. Näitä kahta korpusta käsitellään tarkemmin luvussa 5.

YLE-ruututestikorpuksen ensisijainen tutkimuskohde on sen laatijoiden mukaan ruututekstin lingvistinen analysointi. Lisäksi ruututestikorpuksesta alun perin asetettuihin odotuksiin kuului oletus, että korpusta käytetään muun muassa sen tutkimiseen, poikkeavatko ruututekstit muista suomeksi käännettyistä käännösteksteistä ja suoraan suomeksi kirjoitetuista teksteistä jotenkin leksikaalisesti, syntaktisesti tai tekstuaalisesti. (Mäkisalo 2008.) Omalta osaltaan oma tutkimukseni pyrkiikin vastaamaan YLE-korpuksesta asetettuihin kysymyksiin eli lähinnä siihen, onko tekstittäjän suomi todellakin oma muotonsa käännössuomesta, johon vaikuttaa oma normistonsa.

Jotta saataisiin selville tämän tutkimuksen vertailukohta, eli miten syntagmaattisia sulaumia on käytetty muissa suomen varianteissa, on tässä työssä myös tarkasteltu, miten sulaumamuotoja esiintyy kirjoitetussa suomessa. Tähän selvitykseen käytettiin aineistona Käännössuomen korpusta, joka pitää sisällään niin suoraan suomeksi kirjoitettuja tekstejä kuin myös suomeen päin käännettyjä tekstejä sisältäen lähes 10 miljoonaa sanaa (Mauranen 2002). Tämän jälkeen osakorpuksesta saatuja tuloksia verrattiin Käännössuomen korpuksesta löytyneisiin tuloksiin. Käännöskorpuksen alkuperäissuomen ja käännössuomen sekä osakorpuksen vertailun tarkoituksena on toisaalta tarkastella, seuraavatko suomenkieliset ruututekstit kirjallisen suomen konventioita ja normeja, olivat tekstit sitten alun perin suomeksi kirjoitettuja tai suomeen päin käännettyjä tekstejä, ja toisaalta havainnoida, eroavatko käännetyn kielen konventiot tässä kohdin alkuperäissuomesta. Niin vertailu sulautuneiden ja erilleen kirjoitettujen muotojen välillä kuin ruututekstitysten ja kirjoitetun suomen sulaumien välillä suoritettiin laskemalla ensin niiden esiintymät ja muuttamalla esiintymien määrät suhteellisiksi frekvenssiluvuiksi, koska korpukset olivat selkeästi erikokoisia (kts. alaluku 5.3.3).

Tämän tutkimuksen metodina on korpustutkimus. Tutkimuksen apuna käytettiin *WordSmith Tools* -korpusanalyysiohjelman, jonka on kehittänyt Mike Scott (Scott 2008; 2017). Tutkimusaineistosta, eli korpuksista, haettiin korpuseräilyohjelman konkordanssihaulla (ks. 4.1) kaikki syntagmaattiset sulaumat ja niiden taivutusmuodot sekä erilleen kirjoitettujen muotojen taivutusmuodot, jotka kaikki käytiin läpi tapaus tapaukselta, jotta materiaalista saatiin

seulottua pois tähän tutkimukseen kuulumattomat tapaukset. Materiaalin puhdistuksen jälkeen muotojen lukumäärät kirjattiin ylös ja niille laskettiin suhteelliset frekvenssiluvut miljoonaa sanaa kohden, eli kuinka monta kertaa kukin kielellinen elementti esiintyy miljoonan sanan joukkiossa korpuksessa. Samoin tehtiin sulaumamuotojen kanssa vertailumateriaalista. Tutkimuksen työvaiheista kerrotaan tarkemmin alaluvussa 5.3.

Kuten ylempänä on todettu, vastaavanlaista tutkimusta on ruututekstitysten osalta toteutettu Suomessa varsin vähän. Yhtenä perusteluna tämän tutkimuksen tarpeelle onkin aiemman tutkimuksen vähyys. Ennen tätä tutkimusta Yle-korpuksista on tarkasteltu mm. ruututekstien rivijaon konventioiden rikkomuksia, persoonapronominien käyttöä sekä sanaston laajuutta. Näitä tutkimuksia on käsitelty lyhyesti alaluvussa 4.2. Ainoa aihepiiriltään tätä tutkimusta lähellä oleva selvitys on ollut Sonja Tirkkonen-Conditin ja Jukka Mäkisaloon (2007) tutkimus tiettyjen suomen koheesiokeinojen, diskurssipartikkelien ja liitepartikkelien esiintymisestä ruututekstityksissä. Heidän tutkimuksessaan on samanlainen kontrastiivinen tutkimusasetelma kuin tässä tutkimuksessa, lisäksi aineistoina on käytetty samoja materiaaleja, joskin ruututekstikorpuksista on käytetty tuossa tutkimuksessa suppeampaa aineistoa (mts. 223).

Tämän tutkimuksen päätavoitteena on siis tutkia partikkelin ja kieltoverbin syntagmaattisten sulaumien esiintymistä ruututeksteissä ja vertailla sitä muussa kirjoitetussa suomessa ilmeneviin sulaumamuotoihin sekä arvioida, ovatko odotus- vai suhdenormit tärkeämpiä ruututekstejä laatiessa. Tutkimuksen toissijaisena tavoitteena on tarkastella, miten paljon ruututeksteissä esiintyy syntagmaattisten sulaumien erilleen kirjoitettuja muotoja verrattuna sulaumamuotoihin. Seuraava luku keskittyy tässä tutkimuksessa tarkasteltuihin kielellisiin elementteihin: syntagmaattisiin sulaumiin ja niiden erilleen kirjoitettuihin muotoihin. Luku 3 taas käsittelee tämän tutkimuksen teoreettista pohjaa, eli käänösnormeja sekä ruututekstikäntämisen rajoitteita, sekä tarkastelee tutkimusasetelmaa tarkemmin. Luku 4 esittelee, millaisia korpusohjelmien toimintoja on käytetty tätä tutkimusta tehdessä sekä miten YLE-ruututekstikorpuksista on hyödynnetty av-kääntämistä tutkittaessa. Luku 5 käsittelee tutkimuksen tekemisessä hyödynnettyjä tutkimusaineistoja ja -menetelmiä. Luku 6 esittelee tämän tutkimuksen tuloksia yksityiskohtaisesti käsitellen erikseen jokaista kielellistä elementtiä sekä kahden korpuksen välisen vertailun tuloksia. Viimeinen luku keskittyy saatujen tutkimustulosten analysointiin.

## 2 SYNTAGMAATTISET SULAUMAT

Kuten jo aiemmin on mainittu, tämän tutkimuksen tarkoituksena on tarkastella, miten suomen kieltoverbi *ei* yhdistyy partikkeleiden *ehkä*, *että*, *jos*, *jotta*, *miksi*, *mutta*, *siksi* ja *vaikka* kanssa. Yleiskonjunktio *että*, rinnastuskonjunktio *mutta* sekä adverbiaalikonjunktiot *jos*, *jotta* ja *vaikka* ovat itsessään taipumattomia sanoja, joilla on kiinteä paikka lauseessa välittömästi toisten kytkettävän elementin edessä. Niillä on kuitenkin taipumus yhdistyä kieltoverbiin ja täten muuttua persoonissa taipuviksi konjunktioiksi. (ISK 2005, 784–785). On olemassa myös muita konjunktion ja kieltoverbin sulautuneita muotoja (muun muassa *jossei*, *koskei* ja *kunnei*), mutta ne on jätetty pois tästä tutkimuksesta, sillä ne esiintyvät suomen kielessä vain joissakin murteissa ja yleispuhekielessä, kuten Ilkka Savijärvi (1988, 81) huomauttaa. Tämän tutkimuksen osaksi on hyväksytty vain muodot, jotka löytyvät suomen kielen kieliopista ja ovat vakiintuneet kirjoitettuun yleiskieleen (ISK 2005, 164). Modaalipartikkeli *ehkä* sekä adverbit *miksi* ja *siksi* on otettu mukaan tutkimukseen, sillä ne käyttäytyvät kielellisesti samoin kuin aiemmin mainitut sulaumakonjunktiot yhdistyessään kieltoverbin kanssa (mp). Kaikki tutkimuksessa mukana olleet partikkelit taivutusmuotoineen ovat nähtävillä taulukossa 1 (lista kaikista sulaumista ja niiden esiintymistä löytyy liitteestä I).

Taulukko 1. Tutkimuksessa mukana olleet syntagmaattiset sulaumat taivutusmuotoineen

ell		
ett		
joll		
jott		
mutt		
vaikk		
ehk		
miks		
siks		

+

en
et
ei
emme
ette
eivät ~ eivat

Tietävästi tässä työssä tutkitulle partikkelin ja kieltoverbin yhdistymiselle ei ole suomen kielessä vakiintunutta nimitystä. Näitä yhdistelmämuotoja on kutsuttu niin kieltolauseen konjunktioiksi tai konjunktion ja kieltosanan yhtymiksi (Ikola 1966, 9), kieltosanan ja konjunktion muodostamiksi kiinteiksi sandhiyhtymiksi tai epäorganisisiksi yhdyssanoiksi (Savijärvi 1988, 81–82) sekä syntagmaattisiksi sulaumiksi, eli partikkelin ja kieltoverbin sulaumiksi (ISK 2005, 164–165). Tässä tutkimuksessa on käytetty *Iso Suomen kieliopissa* (lyhyemmin ISK) esitettyä termiä ja sen määritelmää, vaikka sekään ei varsinaisesti ole

vakiintunut suomen kieleen: ”Syntagmaattiset sulaumat ovat yhdeksi sanaksi hahmotettavia kahden sanavartalon muodostamia kokonaisuuksia, jotka eivät ole yhdyssanoja (esimerkiksi *ettei, miksette, tajuutsä* ’tajuatko sinä’)” (mts. 163). Syntagmaattisissa sulaumissa konjunktio on menettänyt loppuvokaalinsa elision myötä ja konjunktio ja kieltoverbi ovat liittyneet yhdeksi fonologiseksi sanaksi. Sulaumien osilla on myös lauseessa omat tehtävänsä, mikä erottaa ne yhdyssanojen osista. (Mts. 164–165.)

Vaikka syntagmaattisilla sulaumilla ei ole vakiintunutta nimitystä, eivät kieltoverbin ja konjunktin sulaumat ole mikään tuore ilmiö, vaan niitä on ollut olemassa siitä lähtien, kun Mikael Agricola alkoi kehittää suomen kielen kielioppia. Ilkka Savijärven (1988) mukaan Agricola käytti yhtymiä *ehkei* (tosin nykykielen merkityksessä ’vaikkei’), *ellei, ettei, jollei* sekä *miksei*. Muissa tapauksissa Agricolan kielessä kieltoverbi ja konjunktio oli kirjoitettu erilleen. (Mts. 81–82.) Vaikka Savijärven (mp.) mukaan Agricola vieroksui monikollisia yhtymiä, eivätkä nämä yhtymät yleensä kongruoineet, eli mukautuneet taivutukseltaan, persoonan tai monikon mukaan, käytti Agricola itse asiassa muun muassa sanoja *etten, ettei* ja *miksei* enemmän kuin samojen sanojen erilleen kirjoitettuja muotoja. Mielenkiintoista kyllä, ISK:n (2005, 165) mukaan nykykielessä sulaumat ovat kaiken kaikkiaan yleisemmin käytössä kuin konjunktiot ja kieltoverbi erilleen kirjoitettuina muotoina. Täten onkin kiinnostavaa verrata, miten ruututeksteistä saadut löydökset vertautuvat ISK:n tuloksiin.

Vaikka syntagmaattiset sulaumat koostuvatkin erillisistä sanayksiköistä, on niiden yksiköillä siis omat tehtävänsä lauseessa, esimerkiksi sulaumamuodossa *mutten* alun *mutt-* toimii konjunktiona ja sanan loppu *-en* taas persoonamuotoisena kieltoverbinä ja täten osana predikaattiverbiä (ISK 2005, 164). Lisäksi sulaumilla on muitakin sanamaisia ominaisuuksia. Ensinnäkin ne äännetään ja ymmärretään yhdeksi sanamuodoksi, yhdeksi fonologiseksi sanaksi. Sulaumissa tapahtuu myös vokaalisoinnun muutoksia, sillä ”takavokaaliseen konjunktioon sulautuneeseen kieltoverbiin voi liittyä monikon 3. persoonan tunnus joko etu- (*-eivät*) tai takavokaalisena (*-eivat*)” (mts. 165), esimerkkinä muodot *jolleivät* ja *jolleivat*. ISK:n (mp.) mukaan etuvokaaliset muodot ovat yleisempiä kuin takavokaaliset muodot. Eräs sulaumien sanamaineinen piirre on se, ettei sulaumien sisään voi sijoittaa liitepartikkelia. Syntagmaattisten sulaumien kohdalla liitepartikkeli liittyy sulaumamuodon perään (esim. *mikseipä*), kun taas sulaumien vastineissa, erilleen kirjoitetuissa kaksisanaisissa ilmauksissa, liitepartikkeli voi asettua niin kieltoverbiin (esim. *mutta eipä*) kuin konjunktin (esim.

*vaikkakaan ei*). (Mp.) Yksi sulaumien sanamainen ominaisuus on myös se, että konjunktioon sulautunut kieltoverbi on tunnusmerkillisesti eli luonnollisimmin lauseen tai virkkeen alussa. Sulauma saattaa esiintyä lauseen alussa silloinkin, kun kieltoverbi ei itsenäisenä lauseenosana esiintyisi suoraan konjunktin jälkeen (mts. 166). Esimerkiksi lauseessa ”Ajattelimme, *ettei* sota loppuisi koskaan” konjunktioon sulautunut kieltoverbi asettuu sivulauseen alkuun, kun taas erilleen kirjoitettuna lauseen subjekti sijoittuu konjunktin ja kieltoverbin väliin, kuten lauseessa ”Ajattelimme, *että* sota *ei* loppuisi koskaan”. *Iso suomen kieliopin* mukaan (mp.) tämä ilmiö on merkki siitä, että ”partikkelin ja kieltoverbin sulaumat ovat kiteytyneitä kokonaisuuksia, joiden sijaintia lauseessa säätelevät eri ehdot kuin niitä osia, joista se koostuu”. Myös partikkeli *ehkä* esiintyy lauseenalkuisessa asemassa yhdistyessään kieltoverbin kanssa sulaumaksi (mp.), kuten esimerkiksi ”*Ehkei* esitys ala vielä” verrattuna lauseeseen ”*Ehkä* esitys *ei* ala vielä”. Kuitenkin *ehkä* voi yhdistyessään kieltoverbin kanssa muista sulaumapartikkeleista poiketen esiintyä lauseessa myös kieltoverbin jälkeen, kuten lauseessa ”Chris *ei ehkä* huomannut meitä”.

Kuten taulukosta 1 näkyy, syntagmaattisten sulaumien tutkimuksessa on mukana monikon 3. persoonassa joistakin sulaumamuodoista kaksi varianttia, sillä kuten jo aiemmin huomautettiin, ”monikon 3. persoonassa voi takavokaalisten konjunktoiden kieltoverbi mukautua vokaalisointuun: *jotteivat*, *vaikkeivat*” (ISK 2005, 164). Huomattavaa on myös, että konjunktio *jos* voi muodostaa syntagmaattisen sulauman kahdella tapaa yhdistyessään kieltoverbin *ei* kanssa, joko sulaumaksi *ellei* tai *jollei*, tosin näiden sulaumamuotojen alkuosat eivät esiinny yleiskielessä itsenäisenä sananaan (mp.). Nämä kummatkin muodot taivutusmuotoineen ovat mukana tässä tutkimuksessa, kuten käy myös ilmi taulukosta 2.

Pelkkiä ruututekstejä tutkittaessa syntagmaattisten sulaumien vertailukohtana pidetään tässä tutkimuksessa samojen kielellisten elementtien erilleen kirjoitettuja muotoja, joissa siis konjunktio tai partikkeli ei ole sulautuneena osaksi kieltoverbiä. Oheisessa taulukossa 2 on lueteltu kaikki tässä tutkimuksessa mukana olleet konjunktiot ja partikkelit taivutusmuotoineen, kun ne ovat kirjoitettu erilleen kahdeksi sanaksi (kattava listaus erilleen kirjoitettujen muotojen esiintymistä liitteessä II).

Taulukko 2. Tutkimuksessa mukana olleet partikkelit taiputusmuotoineen

että	+	en
jos		et
jotta		ei
mutta		emme
vaikka		ette
ehkä		eivät
miksi		
siksi		

Tämän tutkimuksen varsinaisena tutkimuskohteena ovat siis syntagmaattiset sulaumat ja nimenomaan yleiskieleen vakiintuneet sekä persoonittain taipuvat kieltoverbin ja partikkelin yhdistymät (ISK 2005, 164). Kun tässä tutkimuksessa tarkasteltiin juurikin syntagmaattisia sulaumia, otettiin tarkastelussa huomioon myös yhdistelmien päätteelliset muodot, esimerkkeinä *etteikö*, *mikseihän* ja *mutteipä*. Kun taas aineistoa verrattiin vastaaviin kaksisanaisiin ilmauksiin, jätettiin liitepartikkelin sisältämät muodot pois, sillä erilleen kirjoitettujen ilmausten kohdalla liitepartikkelit olisivat nostaneet tutkimusmateriaalin määrää merkittävästi. Kuten tässä tutkimuksessa on jo aiemmin mainittu, tiettävästi tällaista tutkimusta juuri näistä kielen mikrorakenteista ei ole ruututekstien yhteydessä toteutettu suomen kielessä. Toki kielentutkimuksessa on esimerkiksi otettu jo pitkään kantaa *etteikö*-muodon oikeakielisyyteen (tarkemmin aiheesta mm. Matti Larjavaara 1992; Sari Maamies 1993). Mielenkiintoista olisikin lisätutkimuksena tarkastella, miten *etteikö*-muotojen käyttö tekstityksissä vertautuu aihealueen aiempien tutkimusten tuloksiin.



### 3 RUUTUTEKSTIT JA NORMIT

Audiovisuaalisen kuva- ja äänimateriaalin muuntaminen alkuperäiskielisestä auditiivisesta muodosta suomenkieliseksi ruututeksteiksi unohtamatta kuvallisen informaation vaikutusta on erityislaatuinen kääntämisen laji. Toisaalta se on yhtä lailla kääntämistä siinä missä mikä tahansa muukin kääntämisen laji, toisaalta sitä taas säätelevät erityislaatuiset rajoitukset ja konventiot, ehkäpä myös normit, kuten esimerkiksi Jan Pedersen (2011) esittää. Kielitoimiston sanakirja kertoo *normin* tarkoittavan sääntöä, ohjetta tai mallia (MOT 2017), käännöstieteessä normit voivat ohjata niin kääntäjää, käännösprosessia kuin itse käännöstäkin (kts. esimerkiksi Toury 2012 ja Chesterman 1997). Käännöstieteessä normin käsite kuuluu olennaisena osana deskriptiiviseen kääntämisentutkimukseen, jossa painoarvo on käännöksen ja kohdekirjallisuuden suhteella sekä sillä, ”millaisten (ideologisten ym.) tekijöiden ohjaamana ja millaisilla käännösratkaisuilla käännös saadaan sulautumaan samanlaisten, alun perin kohdekielisten tekstien joukkoon” (Puurtinen 2004, 83). Normien avulla on mahdollista myös tutkia, miten säännönmukaisia käännökset ovat keskenään (Aaltonen 2004, 399).

Yksi ruututekstikäntämisen erityispiirteistä on se, että siinä korostuu moniin muihin kääntämisen lajeihin verrattuna käännöksen vastaanottajat eli katsojat. Yleisökirjoitukset lehdissä ja sosiaalisessa mediassa antavat ymmärtää, että suomalaisilla on vahvat mielipiteet käännöksistä, samaa todistavat myös vastaanottotutkimukset (esim. Tuominen 2012 ja Mäyränpää 2010). Kuten Esko Vertanen (2008, 149) huomauttaa, suomalainen katsojakunta on tottunut lukemaan ruututekstejä, lukevathan he arviolta 150 paksun romaanin verran tekstityksiä vuodessa. Kuitenkin esimerkiksi Satu Katajamäen (2009) tutkimus osoittaa, että suomalaiset ovat huonoja tunnistamaan ruututekstien konventioita, vaikka tekstitysten ”virheisiin” helposti tartutaankin. Toisaalta siis suomalaiset osaavat vaatia ruututekstikäännöksiltä paljon, mutta toisaalta he eivät useinkaan tiedä tai ymmärrä, miten tekstitykset saavat muotonsa.

Tässä tutkimuksessa en keskity antamaan ylimalkaista kuvausta av-kääntämisestä, vaan tässä luvussa paneudutaan tämän tutkimuksen viitekehykseen. Lisää informaatiota audiovisuaalisesta kääntämisestä ja sen tutkimuksesta esimerkiksi Agost 2011; Gambier 2007; Díaz Cintas & Remael 2006; de Linde & Kay 1999; Ivarsson & Carroll 1998b. Tämän luvun ensimmäinen osa keskittyy käännösnormeihin ja ennen kaikkea Andrew Chestermanin käsitteisiin *odotus-* ja *suhdenormi*. Toinen osa taas käsittelee ruututekstien rajoitusten ja

konventioiden kautta tekstitysten ymmärrettävyyttä ja luettavuutta. Luvun jälkimmäinen osa nivoo näiden kahden ensimmäisen osan asiat yhteen ja esittelee tutkimuskysymyksen tarkemmin.

### 3.1 Käännösnormit

Käännöstieteeseen normin käsite on tullut pitkälti Gideon Touryn ansiosta (Hermans 1996, 25). Toury itse sanoo, että normit ovat jonkin yhteisön yleisten arvojen tai käsitysten perusteella muovautuneita niin kutsuttuja toimintaohjeita. Nämä toimintaohjeet täsmentävät sitä, mikä on säänneltyä ja kiellettyä, mikä taas suvaittua ja sallittua. (Toury 2012, 63.) Andrew Chesterman (2007, 357) taas kiteyttää normin merkityksen seuraavasti: ”Normi on jokin yhteiskunnan tai yhteisön käsitys korrektiudesta eli oikeasta tavasta toimia tai oikeanlaisesta tuotteesta. Norminmukainen toiminta tai tuote on siis sellainen, joka on kyseisen yhteiskunnan enemmistön mielestä korrekti tai standardi.” Chesterman on selkeyttänyt Renate Bartschin (1987)<sup>1</sup> näkemystä normien olemassaolosta seuraavasti:

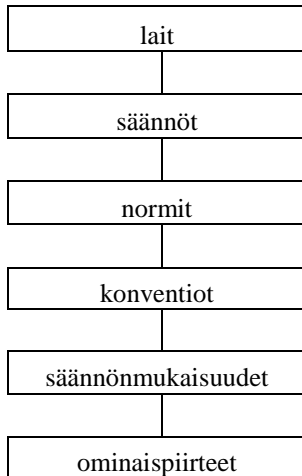
”Hänen mukaansa jokin normi (sanokaamme normi N) on olemassa, jos kolme ehtoa täyttyy. (a) Tietyn yhteiskunnan tai väestöryhmän enemmistön jäsenet toimivat N:n mukaan säännöllisesti. (b) Jos joku ei toimi sen mukaan, muut yhteiskunnan jäsenet voivat oikeutetusti arvostella normin rikkojaa. (c) Yhteiskunnan jäsenet puhuvat N:stä käyttäen sanontoja kuten ’tietyissä tilanteissa pitää tehdä näin’ tai ’oikea tapa toimia on näin’. Näin he perustelevat omaa norminmukaista toimintaansa.” (Chesterman 2007, 357; kts. myös Chesterman 1997, 54–55.)

Chesterman (1997, 55) asettaa normit konventioiden ja lakien väliin. Lait ovat ehdottomia ja objektiivisia sekä jonkin auktoriteetin asettamia, ja lisäksi niiden rikkomisesta seuraa rangaistus. Normit taas ovat heikompia kuin lait: niitä voidaan rikkoa, mutta normien rikkomisesta seuraa kritiikkiä (Chesterman 1997, 55; 2007, 357–358.) Normien rikkomuksista saattaa seurata sanktioita, mutta ne eivät välttämättä ole todellisia, vaan vain mahdollisia (Aaltonen 2004, 400). Konventiot taas ovat näistä kolmesta lievimpiä, niiden rikkomisesta ei yleensä edes koidu kritiikkiä (Chesterman 1997, 55; 2007, 358). Lisäksi, kuten kaikki muukin, myös normit muuttuvat aikojen saatossa, ohjasivat ne sitten meidän elämäämme tai käännöksiä (Mäkinen 2004, 412). Kaiken kaikkiaan normit ovat terminologisesti kuin veteen piirretty viiva, joka muutoinkin vaihtelee tutkijoiden mukaan (Eskola 2002). Esimerkiksi Jan

---

<sup>1</sup> Bartsch, Renate 1987. *Norms of Language - Theoretical and Practical Aspects*. New York: Longman Inc.

Pedersen (2011, 30) yhdistää Chestermanin, Touryn ja Theo Hermansin näkemyksiä ja lisää normien ja lakien väliin vielä säännöt, kuten on nähtävissä kuvasta 1. Myös termien välisiä eroja on hankala määrittää, esimerkiksi missä kohdin tarkalleen konventiot muuttuvat normeiksi (mp.). Pedersenin (mts. 31) mukaan konventiot ilmentävät sitä, mitä ihmisillä on tapana tehdä, kun taas normit määrittävät, mitä pitäisi tehdä, sillä se on muiden toivoma tapa.



Kuva 1. Käännöstoiminnan termien liukuma (Pedersen 2011)

Gideon Toury jakaa normit pääosin ennakkonormeihin, alkuunormeihin ja toimintanormeihin. Ennakkonormit vaikuttavat yleiseen käännöspolitiikkaan, siihen mitä on sopivaa kääntää, miksi käännetään, käännetäänkö välikielen kautta vai suoraan alkukielestä, kuka on sopiva kääntäjäksi. Alkunormit sen sijaan vaikuttavat kääntäjän valintoihin, siihen onko käännös kotouttava vai vieraannuttava, tarkka vai vapaa. (Chesterman 2007, 359.) ”Alkunormin valinta riippuu mm. käännöksen tarkoituksesta, tekstityypistä, asiakkaan toivomuksesta ja vallitsevasta käännösperinteestä” (Chesterman 2007, 360). Toimintanormit taas vaikuttavat kääntäjän päätöksiin käännösprosessin aikana. Toury esittelee myös tekstuaalis-lingvistisen normien käsitteen. Hänen mukaansa ne ohjaavat käännökseen päätyvien kielellisten aineksien valintaa ja ne voivat olla enemmän tai vähemmän yleisluontoisia. Lisäksi jotkin niistä voivat olla samankaltaisia kuin ne normit, jotka säätelevät saman kulttuurin alkuperäiskielellä kirjoitettuja tekstejä. (Toury 2012, 83.)

Kuten on jo aiemmin mainittu, tässä työssä teoreettinen viitekehys perustuu Andrew Chestermanin normijärjestelmään, sillä se perustuu ennemminkin normien alkuunpanijoihin kun käännösprosessiin itseensä (Pedersen 2011, 35). Chestermanin normijärjestelmä jakautuu kahteen päähaaraa, tuote- ja prosessinormeihin. Tuotenormit määrittävät niitä vaatimuksia ja

odotuksia, joita tuotteille on asetettu, esimerkiksi käännosten osalta odotukset, jotka liittyvät esimerkiksi tekstilajiin tai -tyyppiin, tekstien kielellisiin elementteihin tai rakenteeseen (Chesterman 2004, 345). Koska tuotenormit ilmentävät etenkin sitä, millainen jonkin käännoksen on sen asiakkaiden/lukijoiden odotusten mukaan oltava tuona kyseisenä aikana (Chesterman 2007, 361), kutsuu Chesterman tällaisia normeja odotusnormeiksi (Chesterman 1997, 64). Odotusnormit siis ovat käännosten lukijakunnan asettamia (Pedersen 2011, 35). Lukijoiden odotukset perustuvat osittain kohdekulttuurin käännostraditioihin ja osittain siihen, millaisia kyseisten käännosten rinnakkaistekstit ovat tuossa kulttuurissa. Heillä saattaa olla odotuksia niin tekstin tyyppiin ja tyyliin liittyvistä seikoista kuin lauseiden rakenteista ja sanastollisista valinnoista. (Chesterman 1997, 64.) Toisaalta kääntäjillä on useampia keinoja toteuttaa näitä odotusnormeja, eivätkä normit tarkoita, että vain yksi ratkaisu olisi se yksi ja ainoa oikea (Chesterman 1997, 64). Odotusnormien avulla on mahdollista arvioida, miten paljon käännookset noudattavat näitä normeja (mts. 65). Lukijat saattavat odottaa toisten käännoosten sekoittuvan täysin suoraan alkukielellä kirjoitettuihin teksteihin ja tällaisten ”piilokäännoosten” paljastuessa käännoöksiksi niitä saatetaan pitää huonoina kääntämättömiin teksteihin verrattuna, vaikkei syytä tähän edes keksittäisi. Esimerkiksi ruututekstikäännoosten voidaan sanoa olevan katsojan kannalta parhaimmillaan silloin, kun katsoja ei huomaa lukevansa tekstityksiä, vaan saa illuusion siitä, että hän ymmärtää lähtökielistä puhetta (Vertanen 2008, 151). Toisaalta käännoöksissä saattaa olla piirteitä, jotka ovat avoimesti vieraannuttavia ja tuovat esiin, ettei lähdeteksti välttämättä kuulu samaan kulttuuri- ja kielipiiriin kuin kohdekielinen käännoös (Chesterman 1997, 65). Esimerkiksi brittiläisessä poliittisessa satiirissa ei ruututeksteihin voi muuntaa ohjelman puolueita suomalaisiksi puolueiksi.

Chestermanin prosessinormit taas vaikuttavat käännoösprosessiin ja ”kuvaavat niitä periaatteita, joita jossain tietyssä prosessissa pitäisi noudattaa” (Chesterman 2004, 345). Prosessinormit ilmentävät lähde- ja kohdetekstin ihanteellista suhdetta omassa ajassaan, sekä millainen viestintätehtävä käännoöksillä on ja miten kääntäjän pitäisi eettisesti toimia. Chestermanin mukaan kääntäjän näkökulmasta tarkasteltuna prosessinormit ovat alisteisia odotusnormeille; määrittäähän lopputuote eli käännoös sitä, miten tähän lopputulokseen päädytään. (Chesterman 1997, 67.) Vaikka normit sinänsä ovat ohjailevia, eli kääntäjillä on taipumus tehdä niin kuin he uskovat, että heidän tulisi toimia, pyrkivät Chestermanin prosessinormit kuitenkin olevaan kuvailevia, eli lähinnä kertomaan, millaisia normeja on

olemassa (Chesterman 1997, 68). Kääntämisen kohdalla prosessinormeista tärkeimmät ovat suhdenormit, kommunikointinormi sekä eettinen normi. Chesterman kiteyttää näiden erot seuraavasti:

- **suhdenormi:** kääntäjän oletetaan työskentelevän sillä tavoin, että lähdetekstin ja kohdetekstin välillä syntyy ja säilyy tarkoituksenmukainen suhde
- **kommunikaationormi:** kääntäjän oletetaan työskentelevän niin, että teksti viestittää sanomansa mahdollisimman hyvin
- **eettinen normi:** kääntäjän oletetaan toimivan eettisesti vastuuntuntoisella tavalla (Chesterman 2004, 345.)

Suhdenormien kohdalla jää kääntäjän vastuulle päättää, millainen suhde on tarkoituksenmukainen kussakin tapauksessa tekstityypistä, asiakkaan toiveista, alkuperäisen materiaalin tarkoituksista ja mahdollisten lukijoiden tarpeista riippuen. Suhdenormit tuovat esiin sen, että lähde- ja kohdeteksteillä voi olla vastata toisiaan hyvinkin erilaisilla tavoilla. Esimerkiksi auktorisoiduissa käännöksissä lauseiden keskinäinen vastaavuus on tärkeää, kun taas runokäännöksessä saatetaan korostaa tyylin tärkeyttä (Chesterman 1997, 69.) Ruututekstitysten kohdalla tärkeintä ei ole sanatarkka vastaavuus, vaan viestistä poimitaan olennaisin sanoma ja pyritään välittämään sen tyyli ja tunnelma katsojalle (Vertanen 2008, 150–152).

Omassa tutkimuksessaan Jan Pedersen (2011) on tutkinut ruotsin- ja tanskankielisiä televisiotekstityksiä ja vertaillut niiden eroja. Tutkittuaan näistä kulttuuriviittauksia hän laati myös normeja ja konventioita, jotka ovat hänen mukaansa päteviä ainakin ruotsin- ja tanskankielisten ruututekstitysten osalta, mutta jotka todennäköisesti pätevät myös monissa muissa tekstityksissä suosivissa maissa. Mitä enemmän tuloksia jokin piirre, sitä voimakkaampi on niiden vaikutus, josta riippuen Pedersen kutsui niitä joko konventioiksi tai normeiksi (Pedersen 2011, 186–187). Seuraavaksi on lyhyesti esitelty muutamia niistä, jotka todennäköisimmin pitävät paikkansa myös suomalaisissa ruututekstityksissä. Tietävästi näitä termejä ei ole suomennettu aikaisemmin, joten normien nimet ovat omia käännöksiäni.

Pedersenin suuntautuvuusnormin mukaan englanninkielisten elokuvien ja televisiosarjojen ruututeksteissä hyödynnetään lähdepainotteisia käännösratkaisuja (Pedersen 2011, 194). Kotouttamisen normi taas pätee yleensä vain silloin, kun kulttuuriviittaukset viittaavat virastojen nimiin, koulutukseen, hallitukseen, titteleihin ja ruokiin, joskin Pedersen huomauttaa, että tällä normilla on hyvin rajattu käyttötarkoitus (mts. 195). Ylikulttuurisen

normin mukaan kääntäjät käyttävät kulttuurien rajat ylittävien kulttuuriviittausten kohdalla ennalta muodostettuja kohdekielisiä ratkaisuja, mikäli ne vain mahtuvat tekstitykseen. Paino- ja mittanormin mukaan taas ruotsin- ja tanskankielisissä ruututeksteissä käytetään lähinnä virallista vastinetta, eli metrisen järjestelmän vastineita. (Mts. 201–202.) Valuuttanormin mukaan rahayksiköitä koskevat viittaukset jätetään kyseisen ohjelman mukaisiksi, eli niitä ei kotouteta kohdekulttuurin valuuttayksikköön (mts. 204). Virallisen vastineen normin mukaan tekstityksissä käytetään ensisijaisesti virallista vastinetta, vaikkapa jonkin ohjelman suomennettua nimeä, esimerkkinä *Kauniit ja rohkeat* (mts. 205). Säilytysnormin mukaan kulttuuriviittaukset säilytetään yleisimmin sellaisenaan ruotsalaisissa ja tanskalaisissa ruututeksteissä, paitsi jos ne riitelevät paino- ja mittanormia tai virallisen vastineen normia vastaan (mts. 194).

### 3.2 Ruututekstien rajoitteet ja konventiot

Audiovisuaaliseen kääntämiseen vaikuttavat monet tekijät. Osaa niistä kutsutaan konventioiksi, osaa rajoitteiksi ja terminologia vaihtelee kirjoittajasta toiseen. Osaa niistä voisi jopa kutsua normeiksi, ovathan ne laajalti käytössä olevia ja niiden käyttöä opetetaan tuleville kääntäjille ruututekstityksen sääntöinä. Nämä normit ja konventiot vaihtelevat ruututekstien välineen (esim. televisio, elokuvateatteri, DVD) mukaan, kuten Yves Gambier muistuttaa (2008, 87), mutta myös yksittäisen välineen sisällä, sillä eri tv-kanavilla on erilaiset säännökset ja ohjeet. Sen lisäksi ne myös vaihtelevat paljon maittain, myös Euroopan sisällä. Jan Ivarsson ja Mary Carroll (Ivarsson & Carroll 1998a) ovat tehneet yleistetyn yhteenvedon Euroopassa käytetyistä tekstityskonventioista (kts. myös Díaz Cintas & Remael 2006). Tässä alaluvussa keskitytään lähinnä Suomen televisiotekstityksien rajoituksiin ja konventioihin, sillä onhan tutkimuksen materiaali niistä peräisin, mutta valtaosa niistä pätee myös muissa maissa.

Rajoitteista tärkeimmät ovat ruututekstille annettu aika ja tila. Kuten tässä työssä on aiemmin mainittu, pienikin muutos ruututekstissä käytettyjen merkkien määrässä saattaa vaikuttaa tekstitykseen merkittävästikin. Yksi tekstitysten suurimmista rajoitteista onkin siis tila. Aiemmin Ylen tekstityksissä yhden tekstitysriville mahtui noin 30 merkkiä, nykyisin taas 33 (Vertanen 2008, 151). Maksukanaville tekemissäni töissä yhden rivin merkkimäärä on ollut 34–37 lyöntiä. Tämä tieto kuitenkin pätee lähinnä tv-tekstitykseen, josta tässä tutkimuksessa käytetty materiaali on peräisin. Kuten Yves Gambier (2008, 87) huomauttaa, esimerkiksi

elokuvateatteri- ja DVD-kääntämiseen pätevät eri merkkirajoitukset, eikä niissä usein kuvien esitysnopeuskaan ole sama. Merkkimäärien rajoitus perustuu siihen, ettei yhdelle tekstitysriville voi laittaa määräänsä enempää merkkejä ja harvemmin käytössä on enemmän kuin kaksi tekstitysriviä (Pedersen 2011, 19), joskin erikoistapauksissa on käytettävissä kahdeksan riviä, esimerkiksi tapauksissa, joissa alarivien repliikkien lisäksi ruutuun tarvitsee saada myös yläriville puhujan nimi tai jonkin muu planssi (Vertanen 2008, 155).

Tilanrajoitteeseen liittyy olennaisesti myös ajan rajoite (Pedersen 2011, 19), joka onkin kenties ruututekstien rajoitteista tärkein. Yleensä kaksiriviset tekstitysurepliikit ovat ruudussa neljästä viiteen sekuntia, kun taas yksiriviset vievät vähemmän luku-aikaa ja niiden lukemiseen tarvitaan täysipitkinä kahdesta kolmeen sekuntia. Kuitenkin yksittäisen repliikin kesto voi vaihdella 1–30 sekuntiin. (Vertanen 2008, 151.) Tärkeintä on, että repliikki pysyy ruudussa juuri saman ajan, jonka filmimateriaalin puhuja käyttää puhumiseen (mts. 152). Ajalliseen rajoitteeseen ja repliikkien pituuteen vaikuttava tekijä on myös lukunopeus. Katsojan on ehdittävä lukemaan repliikit niin, että hän ehtii ymmärtää niiden merkityksen ja hän on koko ajan tietoinen, kenen puheesta on kyse (Vertanen 2008, 151). Lisäksi monimutkaisen sanaston ja lauserakenteiden käyttäminen vaatii pidemmän lukuajan, että katsoja ehtii ymmärtää tekstityksen. Lisäksi se, mitä enemmän ruudulla tapahtuu, vaikuttaa lukunopeuteen, kun katsojan pitää keskittyä myös kuvaruudun tapahtumiin ja silmän on ehdittävä liikkua kuvan ja tekstityksen väliä (Ivarsson 1992, 38). Myös kuvanleikkauksen yli jatkuvat tekstitykset saattavat häiritä katsojaa (de Linde & Kay 1999, 16) ja vaikuttaa häiritsevästi myös ruudulla näkyvän tekstin lukemiseen. Toki on muistettava, että vaikka tekstityksen ymmärrettävyys ja helppolukuisuus on ruututekstin koostamisen tärkeä kriteeri, vaikuttaa näihin kahteen myös tekstin sommittelu. Siinä missä sommittelulla on vaikutus katsojakokemukseen, esimerkiksi se peittää kuvasta jotakin tärkeää informaatiota tai henkilöahmon kasvoja, voi se myös oikeasti hankaloittaa ymmärtämistä, mikäli esimerkiksi attribuutti erotetaan pääsanastaan toiselle riville tai jopa toiseen repliikkiin. (Vertanen 2008, 154–156.)

Rajoittavana tekijänä voisi pitää myös ohjelman katsojakuntaa, eli tekstityksen vastaanottajia. Esimerkiksi lastenohjelmissa pitää ottaa huomioon, että ohjelman tarkoitettu vastaanottajakunta tarvitsee niin sanottua normaalikatsojaa enemmän luku-aikaa. (Pedersen 2011, 20). Myös vanhempi väestö tarvitsee enemmän luku-aikaa. Ruotsin tekstityshistorian

alussa oli normina, että kääntäjän itsensä piti ehtiä lukea repliikki kolmeen kertaan sinä aikana, jona sen on tarkoitus pysyä ruudussa, jotta tekstitys olisi soveltuva yli 60-vuotiaalle katselijakunnalle (Ivarsson 1992, 44).

Ruututekstien rajoitusten tähden kaikkea ohjelmassa sanottua ei voida välittää tekstityksissä, vaan jotain pitää jättää pois. Av-kääntämisen kohdalla voidaankin osittain puhua tiivistämisestä kääntämisen ohella, sillä ideaalitapauksessa liian pitkä lähtökielinen viesti muokataan välittämään sama viesti kohdekielellä tiivistetympinä (Pedersen 2001, 20). Pois jätetään ensisijaisesti henkilöiden nimet ja tittelit, adjektiiviattribuutit, paikannimet ja aikamääreet. Mikäli lukuaikaa riittää, tiivistä käännettyyn viestiin voidaan lisätä näitä tarkentavia ja täydentäviä elementtejä antamaan sille syvyyttä (Vertanen 2008, 153).

Ruututekstien konventioihin taas kuuluu se, että yleensä ottaen ruudun alareunassa on korkeintaan kaksi tekstitysriviä. Suurimmilla kanavilla tekstitykset on tasattu vasemmalle, mutta joillakin pienemmillä kanavilla ja esimerkiksi elokuvateattereissa ruututekstit on keskitetty. Television puolella taas planssit keskitetään. Lisäksi tavuviivalla yleisimmin merkitään puheen jatkumista seuraan repliikkiin, kun taas kolme pistettä yleensä tarkoittaa puheen katkeavan kesken. Repliikkiviivalla erotetaan puhujat toisistaan – pääkanavalla vain jälkimmäisen puhujan repliikin eteen tulee tavuviiva, kun taas joillakin maksukanavilla ja varsinkin DVD-käännöksissä molempien puhujien eteen tulee viiva. Kursiivia taas voidaan käyttää silloin, kun puhuja ei näy ruudussa, takaumissa, lainauksissa tai vieraskielisiä sanoja tai lauseita käytettäessä. (Vertanen 2008; BTI 2011; SDI Media 2012.)

Ruututekstitykseen vaikuttaa tietenkin sen alkuperäinen materiaali eli lähdekieli. Ruututeksteissä suullinen ja kirjallinen viestintä sekoittuvat siinä mielessä, että ruututekstit ilmentävät puhuttua kieltä kirjoitetussa muodossa (de Linde & Kay 1999, 26). Esko Vertasen (2008, 150) mukaan kääntäjän tulisi olla lähtökielen ilmaisulle uskollinen ja yrittää saada sen tyyli ja tunnelma välitettyä ruututeksteihin. Ruututekstien ymmärrettävyyteen vaikuttaa toki sekin, että tekstityksen on tarkoitus seurata mahdollisimman tarkasti puheen rytmiä ja olla lähinnä puhujan puheen ajan ruudussa (Vertanen 2008, 150–151). Kun tekstitys toimii sopusoinnussa ohjelman kuvan ja äänen kanssa, syntyy katsojan mieleen kuvitelma siitä, että hän ymmärtää ohjelmassa puhuttua kieltä sen sijaan, että lukisi omankielistä ruututekstitystä (mts. 151). Lisäksi on todettu, että mikäli tekstitykset seuraavat puheen ja kuvan rytmejä



kunnolla, ne eivät häiritse katsojaa käsittelemästä ruudun kuvainformaatiota (Lång, Mäkisalo, Gowases & Piettinen 2013, 72). Hyvä käänös siis auttaa katsojaa tukien ohjelman katsomiskokemusta (Hyvärinen 2007).

Ymmärrettävyyteen ja samalla luettavuuteen vaikuttaa tietenkin se, miten tekstitys on rakennettu. Vertasen (2005, 154) mukaan yksittäisen ruututekstin pitäisi olla järkevä kokonaisuus, jonka katsoja ymmärtää nopealla lukaisulla. Suositeltavinta olisi, jos yksi repliikki sisältäisi kokonaisen virkkeen tai lauseen. Mikäli virke jää kesken, se jaetaan useampaan repliikkiin järkevinä kokonaisuuksina. (Mp.) Ylen kääntäjien mukaan selkeä lauserakenne ja ymmärrettävyys ovat ruututeksteissä tärkeitä. (Hyvärinen 2007.) Kääntäjän tavoitteena tulisi aina olla mahdollisimman lyhyt ja ytimekäs ilmaus kohdekielellä (Vertanen 2008, 170). Esimerkiksi kiilalauseet, pitkät virkkeet ja pelkät pronominit saattavat haitata ymmärtämistä (Hyvärinen 2007). Myös Tiina Tuomisen (2012, 10) vastaanottotutkimuksen mukaan lukijoille oli tärkeää ”tekstityksen sujuvuus, konventioiden mukaisuus ja audiovisuaaliseen kokonaisuuteen sopivuus”. Ivarssonin ja Carrollin (1998, 157) mukaan tekstitysten kielen tulisi olla kielipiillisesti ”oikeaa”, toimivathan ruututekstit luku- ja kirjoitustaidon mallina.

### 3.3 Tutkimusasetelma

Tässä työssä on siis hyödynnetty Chestermanin käänösnormeja, ja erityisesti odotus- ja suhdenormien käsitteitä. Chestermanin normeja voidaan hyödyntää ruututekstejä tarkasteltaessa, sillä ne keskittyvät käänösten lukijoihin, ja ruututekstien kääntämisessähän lukijoiden/katsojien asema korostuu (Pedersen 2011, 37), sillä elinkaareltaan lyhytaikaiset tekstitysrepliikit on suunnattu jollekin tietylle yleisölle sinä tiettyinä aikana. Ensinnäkin, vaikka suomalaiset lukevat vuodessa valtavan määrän ruututekstejä (Vertanen 2008, 149, he eivät välttämättä ymmärrä, mitä vaatii vieraskielisen ohjelman puheen muuttamisesta kotimaisen kielen tekstitykseksi, eivätkä täten tiedosta sen rajoitteita ja konventioita. Esimerkiksi Satu Katajamäen (2009) tutkimus osoittaa, että suomalaiset eivät ole hyviä tunnistamaan ruututekstien konventioita, vaikka tekstitysten ”virheisiin” tartutaankin helposti. Täten on syytä olettaa, että suomalaisen katsojakunnan yhtenä odotusnormina on se, että suomenkielisissä ruututeksteissä käytetään samalla tavoin samaa kieltä, jota muissakin suomeksi kirjoitetussa teksteissä käytetään, olivat ne sitten käännettyjä tai suoraan suomeksi kirjoitettuja. Tämän tutkimuksen puitteissa odotusnormi koskee tietenkin muussa kirjoitetussa

suomessa käytettyjä syntagmaattisia sulaumia. Ovathan tutut, yksinkertaiset sanat helpompia katsojien ymmärtää, sekä ne ovat nopeammin luettavissa kuin oudommat muodot (Ivarsson & Carroll 1998, 89). Lisäksi Chesterman muistuttaa, että odotusnormit saattavat koskea hyvinkin pieniä tekstin osasia ja sanavalintoja (Chesterman 1997, 64). Mikäli siis ruututeksteissä esiintyy samoissa määrin sulaumamuotoja kuin muissakin suomen varianteissa, voidaan sanoa tekstitysten noudattavan tätä odotusnormia.

Toiseksi, vieraskielisen puheen kääntäminen suomenkielisiksi ruututeksteiksi vaatii ruututekstikäntämisen rajoitteiden ja konventioiden tähden usein paljon tiivistämistä. Täten voisi olettaa av-kääntäjien käyttävän hyväkseen pieniäkin keinoja, joilla voi helpottaa viestin mahtumista repliikkeihin, vaikka nämä keinot eivät niin yleisiä olisikaan muissa käänöksissä tai alun perin suomeksi kirjoitetuissa teksteissä. Syntagmaattiset sulaumat soveltuvat hyvin tällaiseksi keinoksi. Mikäli siis sulaumamuotoja ilmenee ruututeksteissä selkeästi enemmän kuin muissa suomen varianteissa, voidaan arvella tekstitysten noudattavan pikemminkin Chestermanin suhdenormia kuin odotusnormia. Tällöin on oletettavaa, että kääntäjät ovat pitäneet tärkeämpänä välittää ohjelmassa sanotun viestin ruutuun niillä keinoilla, joilla vain pystyvät. Ovathan syntagmaattiset sulaumamuodot kuitenkin kieliopillisesti oikeaoppisia muotoja, vaikkakin niitä ilmenisikin vähemmän muissa suomen varianteissa.

Tämän tutkimuksen tarkoituksena on siis tutkia, kumpi näistä normeista vaikuttaa enemmän ruututeksteissä. Pyrkimyksenä ei suinkaan ole arvottaa toista normia paremmaksi. Tutkimuksen oletuksena on, että ruututekstejä käännettäessä vaikuttaisi niiden suhdenormi enemmän kuin odotusnormi juurikin av-kääntämisen rajoitteiden tähden, eli tutkimuksen hypoteesina on, että syntagmaattisia sulaumamuotoja esiintyy enemmän ruututeksteissä kuin muissa suomen varianteissa.

## 4 KORPUSTUTKIMUS

Korpuslingvistiikka on metodologia, jonka avulla tutkitaan kielenkäyttöä. Korpuslingvistiikka on alettu hyödyntää käännösten ja kääntäjien hyväksi 1980-luvulta lähtien, ja korpuksia on siitä alkaen käytetty tutkimaan, poikkeako käännösten kieli jotenkin suoraan alkuperäiskielellä kirjoitetuista teksteistä (Anderman & Rogers 2008b, 13-15). Maeve Olohan (2004, 16) listaa tiivistetysti, millaisia tutkimuskohteita korpuslingvistiikan avulla on tutkittu käännöstieteessä:

- käännösten ominaisuuksien deskriptiivinen tutkimus
- käännöksissä käytetty kieli
- käännöksille ilmeiset ja tyypilliset sekä samalla epätyypilliset piirteet
- kvantitatiiviset ja kvalitatiiviset tutkimukset, joiden kohteena saattaa olla sanasto, syntaksi tai puheen ominaisuudet
- korpusmetodologian soveltaminen erilaisiin käännöstyyppeihin

Alkujaan korpuksilla on tarkoitettu laajoja tekstikokoelmia, mutta vuosien ja teknologisen kehityksen saatossa on korpuslingvistiikan 'korpus' tullut tarkoittamaan koneen avulla luettavia tekstimassoja (Anderman & Rogers 2008a, 5–15). Nykyisin *Oxford Dictionariesin* (2017) määritelmän mukaan *korpus* on kirjoitettujen tai puhuttujen materiaalien kokoelma, joka on konelukuisessa muodossa ja koottu yhteen kielellistä tutkimusta varten. Kielitoimiston sanakirja kuvaa korpusta sanoilla ”tutkimuksen aineistoksi valittu kokoelma kieliainesta” (MOT 2017). Tästä lähtien, kun tässä työssä puhutaan korpuksista, viitataan tekstissä juurikin elektronisiin korpuksiin. Koska korpuslingvistiikka on tämän tutkimuksen metodi, eikä varsinainen tutkimuskohde, ei tämä luku pyri esittämään kattavaa kuvausta korpuslingvistiikasta ja -tutkimuksesta, vaan antamaan ylimalkaisen kuvauksen siitä, miten tässä tutkimuksessa korpuslingvistiikkaa on käytetty tutkimuksen tulosten saamiseksi.

Tätä tutkimusta tehdessä korpuksia käsiteltiin *WordSmith Tools* -hakuohjelmalla (versio 4.0, julkaistu 2004), jonka kehitti Mike Scott vuonna 1996 (Scott 2004; 2008; kts. myös 2017). Seuraavassa alaluvussa käsitellään nimenomaan tässä työssä hyödynnettyjä *WordSmith Toolsin* tarjoamia lisätyökaluja, joskin samanlaisia toimintoja löytyy myös muista samankaltaisista kaupallisista korpusohjelmista sekä ilmaishakuohjelmista. Tällaisia ovat mm.

MonoConc, MicroConcord, Longman MiniConcordancer sekä internetissä vapaasti käytettävissä oleva Corpus Wizard (kts. Wilkinson 2006, 9; Kujamäki ja Jääskeläinen 2001, 251).

#### 4.1 Konkordanssiohjelmien toimintoja

**Sanalista** (WordSmith Toolsissa *WordList*) on korpusohjelmien toiminto, joka nimensä mukaisesti luo sanalistan kaikista sanoista ja sanaryhmistä, jotka esiintyvät analyysiä varten valituissa teksteissä (kts. Austermühl 2001, 130). WordSmith Toolsissa tämä toiminto luo automaattisesti listan korpuksen sanoista niin aakkosjärjestyksessä kuin frekvenssien mukaan yleisyysjärjestyksessä sekä tunnistaa kaikkein yleisimmät sanarykelmät (Scott 2004, 110).

Sanalista-toiminto myös antaa kyseessä olevan korpuksen tilastolliset tiedot sisältäen tiedostojen koot, tekstissä esiintyvien *saneiden (token)* eli juoksevien sanojen määrän sekä *sanojen (type)* eli erillisten sanamuotojen määrän, sanojen ja lauseiden pituudet sekä useita muita tietoja (Scott 2017). Sanalista-toiminnon avulla on myös mahdollista tarkastaa, millainen on tutkitun tekstimassan saneiden ja sanojen suhde tuhatta sanaa kohti (Bowker & Pearson 2002, 110). Tämä ns. *type/token*-suhde (englanniksi *type/token ratio*) antaa kuvan tekstin sanastollisesta rikkaudesta, sillä mitä suurempi tuo suhdeluku on, sitä useampia sanoja on tekstimassassa käytetty (Mäkisalo 2008). Tätä tutkimusta tehdessä sanalista-toimintoa hyödynnettiin erityisesti sen tilastollisen informaation luomiseen, jota tarvittiin korpuksesta saatuja tietoja vertaillaessa, esimerkiksi laskettaessa suhteellisia frekvenssilukuja (lisää alaluvussa 5.3.3).

**Konkordanssi** on toiminto, joka tutkii tekstejä tai tekstiryhmiä ja tuottaa tuloksenaan halutun merkkijonon esiintymiä (Scott 2017). Konkordanssi (WordSmith Tools -korpusohjelmassa *concord*) on siis työkalu, jolla voi hakea käyttäjän määrittämästä materiaalista jotakin tiettyä hakusanaa tai -sanoja (Wilkinson 2006, 9). Eritoten konkordanssilla tarkoitetaan halutun merkkijonon esiintymistä kontekstinsa kanssa (Olohan 2004, 63). Perusidealtaan konkordanssitoiminto toimii kuten mikä tahansa verkossa toimiva hakukone: se etsii sille määritellystä materiaalista kaikki käyttäjän syöttämän hakusanan tai -sanojen esiintymät.

Kun jokin tietty hakusana tai -avain on syötetty hakuohjelmaan, ohjelma tuottaa tuloksista esiintymälistan, eli konkordanssin, jossa haettu sana tai fraasi näkyy luettelon keskellä korostettuna. Sanan ympärillä näkyy käyttäjän itsensä määrittämä määrä hakusanan ympärillä

olevaa kotekstiä, eli tarkasteltavana olevaa kieliainesta ympäröivää tekstiä (kts. Kujamäki & Jääskeläinen 2001, 243–244; Wilkinson 2006, 9). Kotekstiä voidaan myös kutsua hakusanan verbaaliseksi kontekstiksi tai verbaaliseksi ympäristöksi (Sinclair 2003, 174). Tällainen hakuluettelo eli konkordanssi tunnetaan paremmin sen englanninkielisellä nimellä Key Word in Context (suom. ”hakusana kontekstissään”, tästä lähin viitattu nimellä KWIC-haku). KWIC-hakua voidaan käsitellä monilla tavoin, jotta saadaan selville hakuavaimelle tyypilliset kielelliset kaavat. (Wilkinson 2008.) Esimerkiksi WordSmith Tools -ohjelmassa hakua pystytään muokkaamaan niin, että ohjelma järjestelee hakutulokset käyttämällä L1:tä pääasiallisena lajitteluperusteena – eli ohjelma järjestää tulokset aakkoselliseen järjestykseen sen sanan perusteella, joka on ensimmäisenä vasemmalla hakusanasta tai -fraasista. Mikäli jotkin L1-sanoista vastaisivat täydellisesti toisiaan, voi ohjelman asettaa lajittelemaan tulokset toissijaisesti R1:sen, eli ensimmäisenä hakuavaimen oikealla puolella olevan sanan, mukaan. (Scott 2004, 89–90; 2017.) Tämän tutkimuksen tuloksia tarkasteltaessa käytettiin sekä L1-että R1-sanoja lajitteluperusteena (aiheesta tarkempi selostus alaluvussa 5.3.2).

**Kollokaatit** ovat sanoja, jotka esiintyvät usein jonkun toisen sanan kanssa, korpustutkimuksessa juurikin hakusanan läheisyydessä (Scott 2017). Kuten Kujamäki ja Jääskeläinen (2001, 244) esittävät saman asian, ”tietyssä merkityksessä käytettyyn sanaan liittyy syntaktisten erityispiirteiden lisäksi taipumus esiintyä tietynlaisten toisten sanojen, ns. läheisten kumppaneidensa, kanssa”. Kollokaatit ovat kääntäjille tärkeitä etenkin käännettäessä vieraaseen kieleen päin (mp.), varsinkin käännettäessä jotakin erikoisalan kieltä (Bowker & Pearson 2002, 32). Kollokaattien listasta näkyy myös, miten paljon sanaa esiintyy ja missäkin asemassa, eli kuinka kaukana hakusanasta kyseinen sana esiintyy, esimerkiksi tässä aineistossa *ole*-sana esiintyy *ettei*-sulaumapartikkelin kanssa yleisimmin kahden sanan päässä oikealla ja persoonapronomini *hän* taas yleisimmin suoraan hakusanaa seuraavana sanana. Kollokaattilistaa käytettiin tässä tutkimuksessa hyväksi, kun tarkasteltiin ruututeksteissä esiintyvien syntagmaattisten sulaumien lähimpiä sanoja.

## 4.2 Yle-ruututekstikorpus audiovisuaalisen kääntämisen tutkimuksessa

Tämän alaluvun tarkoituksena on esitellä lyhyesti, miten korpuksia on hyödynnetty audiovisuaalista kääntämistä tutkittaessa. Ulkomailla on jonkin verran tutkittu av-kääntämistä käyttäen apuna suuria korpusaineistoja, joista merkittävin lienee Italiassa Bolognan yliopistossa koottu Forlì Corpus of Screen Translation (tai *Forlìxt 1*), joka sisältää yksitoista

italialaista ja yhdeksän saksalaista elokuvaa sekä täydelliset transkriptiot jokaisen elokuvan dialogista sekä elokuvien ruututextitykset ja/tai dubbaukset (lisää esimerkiksi Valentini 2006a; 2006b; 2008). Tässä työssä on kuitenkin haluttu keskittyä Suomessa tehtyyn tutkimukseen, ovathan työssä tutkitut syntagmaattiset sulaumat nimenomaan suomen kielen mikrorakenteita. Lisäksi tässä koosteessa keskitytään juurikin YLE-ruututextikorpuksesta tehtyihin tutkimuksiin.

Tutkimusprojekti, jossa YLE-ruututextikorpusta koottiin, toteutettiin entisessä Joensuun yliopistossa (nykyinen Itä-Suomen yliopisto), joten luonnollisesti valtaosa sitä hyödyntävästä tutkimuksesta on kyseisen yliopiston opiskelijoiden ja tutkijoiden tekemiä. Tähän mennessä ruututextikorpuksen aineistoa on hyödynnetty seitsemässä gradussa vuosilta 2008–2013, joista kolmeen en ikävä kyllä päässyt käsiksi tätä tutkimusta tehdessäni. Lisäksi Sonja Tirkkonen-Condit (2007) ja myöhemmin myös Jukka Mäkisalo (Tirkkonen-Condit & Mäkisalo 2007) ovat tutkineet tekstitysten erityispiirteitä YLE-ruutukorpusta hyödyntämällä. Näiden tutkimusten lisäksi Jukka Mäkisalo on opetustarkoitusta varten tarkastellut esimerkiksi ruututexttien yleisintä sanastoa, fonologisia piirteitä ja tyylillisesti värityntä sanastoa (Mäkisalo 2008). Tietävästi YLE-korpusta ei ole hyödynnetty uudemmissa tutkimuksissa (Mäkisalo 2017).

Ensimmäisenä opiskelijana YLE-ruututextikorpusta on hyödyntänyt Tiia Pelkonen (2008) omassa pro gradu -tutkielmassaan. Pelkonen tutki rivinjaon konventioiden rikkomuksia, eli tapauksia, joissa kieliopin mukaan yhteen kuuluvat sanat (määrite ja pääsana) on ruututextityksessä erotettu eri riveille. Ideaalitapauksessa repliikin pitäisi koostua virkkeestä, eikä etenkään predikaattia ja objektia saisi erottaa toisistaan rivijaossa (Vertanen 2008, 156). Pelkonen tarkasteli konventioiden rikkomuksia muun muassa demonstratiiviattribuuttien, genetiivi- ja adjektiiviattribuuttien sekä erilaisten sanaliittojen osalta, ja hänen tutkimusmateriaalinaan oli yksi jakso kymmenestä tv-sarjasta, ja tutkimuksen tuloksena oli, että kaikista kaksirivisistä repliikeistä rivinjaon rikkomuksia esiintyi 5,5 prosentissa ja näistä rikkomuksista 21,5 prosentissa rivinjaon rikkomus ei ollut välttämätön tilanpuutteen tähden. (Pelkonen 2008.) Pelkosen tutkimuksessa YLE-korpusta oli siis käytetty niin materiaalin saamiseen kuin tekstityskonventioiden noudattamisen tutkimiseen.

Heli Teitto (2010) taas on tarkastellut omassa pro gradu -tutkielmassaan ruututextikorpuksen materiaalien avulla persoonapronominien ja erisnimien käyttöä

ruututeksteissä vertailemalla niiden käyttöä muuhun käännösuomeen sekä supisuomeen. Teitto tutki lisäksi myös käännösuomen ja supisuomen välisiä eroja olettaen, että ruututekstit eroavat jossain määrin muusta käännösuomesta ja supisuomesta. Hänen hypoteesinaan oli, että ruututekstien rajoitteiden (lukuaika ja tila) tähden ruututeksteissä suosittaisiin ennemminkin erisnimiä kuin monitulkintaisia 3. persoonan pronomineja (mts. 3), ja täten erisnimet olisivat ruututeksteissä yleisempiä kuin muussa käännösuomessa tai supisuomessa. Teiton lopputuloksena oli, että persoonapronominit ovat ruututekstissä harvinaisempia kuin käännösuomessa ja supisuomessa, eritoten dialogia tarkastellessa (mts. 33-34). Kaiken kaikkiaan ruututekstit ja muu käännösuomi sisälsivät Teiton materiaalissa miltei yhtä suuren määrän persoonapronomineja ollen selvästi yleisempiä niissä kuin supisuomessa. Erisnimiä Teiton materiaalissa esiintyi eniten supisuomessa ja vähiten ruututeksteissä. (Teitto 2010.)

Anu-Irina Martikainen (2013) havainnoi pro gradu -tutkielmassaan YLE-korpuksen englannista suomennettujen ohjelmien tekstitysten avulla eri ohjelmatyyppien sanastoa, tarkemmin sanottuna sanaston laajuutta, sanojen ja virkkeiden pituuksia sekä persoonapronominien esiintymiä ja käyttöä. Martikaisen osakorpus piti sisällään tekstityksiä genreistä fakta, fiktio ja viihde. Tutkielman lopputuloksena oli, että ruututekstien sanastollinen laajuus oli suurimmillaan hyvin informatiivisissa ohjelmatyypeissä kuten dokumenteissa. Lisäksi tällaisissa ohjelmissa oli yleensä ottaen pitkiä sanoja ja virkkeitä sekä valtaosin niissä käytettiin kolmatta persoonaa. Suppeimmillaan ruututekstien sanasto taas oli viihteellisissä ja fiktiivisissä ohjelmatyypeissä, samoin niiden sanat ja virkkeet olivat enimmäkseen lyhyempiä ja niissä käytettiin eniten yksikön 1. ja 2. persoonaa. Martikaisen sanoin ”tutkielmassa tarkasteltujen piirteiden osalta vaikuttaa siltä, että ohjelman tyyli ja tarkoitus, eli onko kyse informatiivisesta vaiko fiktiivisestä tai viihdyttävästä ohjelmasta, vaikuttaa hyvin paljon ruututekstien sanastoon”. (Martikainen 2013.)

Taina Ukkola (2014) tutki pro gradu -tutkielmassaan kahta versiota *Star Wars Episode IV: A New Hope* -elokuvan ruututeksteistä. Hän tarkasteli DVD-version ja ruututekstikorpuksesta saadun Ylen kääntäjän suomennoksen samankaltaisuuksia ja eroja verraten näitä kahta alkuperäiseen dialogiin. Ukkolan tarkoituksena oli selvittää, pystyisikö propositioiden avulla vertailla tekstityksiä (mts. 1). Ukkola löysi materiaaleista enemmän yhtäläisyyksiä kuin hän alun perin oletti, mutta kaiken kaikkiaan Ylen materiaalissa oli enemmän propositioita kuin

DVD-version käännöksessä, tosin Ylenkin materiaalissa oli selvästi vähemmän propositioita kuin alkuperäisessä materiaalissa. (Ukkola 2014.)

YLE-ruututekstikorpuista hyödynsi omassa tutkimustyössään myös Sonja Tirkkonen-Condit (2007), joka laajensi tutkimustaan yhdessä Jukka Mäkisaloon kanssa (Tirkkonen-Condit & Mäkisalo 2007). He tekivät kontrastiivista tutkimusta aineistonaan YLE-korpus ja Käännössuomen korpus. *Kääntäjä*-lehden artikkelissaan Tirkkonen-Condit (2007) esittelee tarkastelleensa ruututekstien koheesiokeinoja, diskurssipartikkeleita (esim. *muka* ja *entä*) sekä liitepartikkeleita (*-kin* ja *-han*) ja aineistonaan tässä tutkimuksessa hän on käyttänyt ruututekstityskorpuksesta vuonna 2004 lähetettyjä englannista suomennettuja tekstityksiä ja verrannut tuloksia Käännössuomen korpuksen materiaaleihin. Tirkkonen-Conditin ja Mäkisaloon (2007, 223) yhteistutkimuksessa tutkittiin edeltävän tutkimuksen materiaaleista seuraavia koheesiokeinoja: *kuitenkin/kuitenkaan*, *mutta*, *myös*, *entä*, *-kin*, *-han*, *toki*, *tosin*, *kunhan* sekä *muka*. He tutkivat näiden konjunktioiden ja diskurssipartikkelien esiintymistä niin alkuperäissuomen kuin käännetyn suomenkin teksteissä sekä ruututekstityksissä. Nämä koheesiokeinot valikoituivat tutkimuksen kohteeksi, sillä tutkijoiden mukaan aiemmat tutkimukset antavat olettaa ns. tekstittäjän suomen hyödyntävän suppeutta ja ytimekkyyttä ilmentäviä koheesiokeinoja. (Mts. 221–223).

Vertailun tarkoituksena oli selvittää, eroaako ruututeksteissä käytetty suomi jotenkin muusta kirjoitetusta suomesta, eritoten onko tekstittäjän suomella omia erityispiirteitä muuhun käännössuomeen verrattuna (Tirkkonen-Condit 2007) tai piirteitä, jotka esiintyvät tekstityksissä selkeästi vähemmän tai enemmän kuin muissa suomen varianteissa (Tirkkonen-Condit & Mäkisalo 2007, 222). Kuten itsellänikin tätä tutkimusta tehdessä, oli Tirkkonen-Conditin ja Mäkisaloon hypoteesina, että ruututekstien dialoginen luonne sekä vaatimus ytimekkästä kielenkäytöstä tuottaisivat tekstityksiin suhteellisesti suuremman määrän lyhyitä konjunktioita ja diskurssipartikkeleita kuin muissa suomen kielen varianteissa. Lisäksi tutkijoiden mukaan aiemmat tutkimukset osoittavat, että suomen kielen ns. uniikkiaineiksia on vähemmän käännettyssä suomessa kuin alun perin suomeksi kirjoitetuissa teksteissä ja myös tästä syystä ruututekstitysten kieli on hedelmällinen tutkimuskohde kielen koheesiokeinoista. (Mts. 224.) Tirkkonen-Conditin kirjoituksissa termillä 'uniikkiaines' viitataan sanoihin tai kieliopillisiin elementteihin, ”joille ei ole yksiselitteisiä muunkielisiä vastineita”, esimerkkeinä liitepartikkelit *-kin* ja *-han* (2007; kts. myös 2005).



Tirkkonen-Conditin ja Mäkisaloon (2007, 228) tutkimuksessa tuloksia vertailtiin laskemalla eri koheesiokeinojen esiintymille suhteelliset frekvenssiluvut tuhatta sanaa kohden. Tutkimuksen tuloksena oli, että ruututekstitysten suomi käyttää hyväkseen lyhyitä ja dialogisuutta korostavia koheesiokeinoja, mutta ruututeksteissä myös ilmeni enemmän uniikkiaineita kuin kaiken kaikkiaan ilmeni käännetyssä suomessa. Toki ruututekstien kääntäjät ovat käyttäneet ytimekkäitä ilmaisuja, mutta he ovat käyttäneet tällaisia lyhyitä koheesiokeinoja paikoitellen jopa enemmän kuin alkuperäis-suomessa on käytetty. (Mts. 229.)

Tirkkonen-Condit ja Mäkisalo (2007, 229) tiivistävät tutkimuksensa tulokset kolmeen hypoteesiin: (1) mikäli jokin koheesiokeino on lyhyt, dialoginen ja uniikki, se todennäköisesti esiintyy ruututeksteissä enemmän kuin muissa käännöskielen varianteissa, (2) mikäli jokin koheesiokeino on lyhyt, dialoginen ja uniikki, se saattaa ilmetä ruututeksteissä jopa runsaslukuisempaan kuin alkuperäis-suomessa ja (3) mikäli jokin koheesiokeino on lyhyt, dialoginen, muttei uniikki, se esiintyy luultavasti ruututeksteissä useammin kuin käännösuomessa ja ns. supisuomessa. Tutkimuksen loppupäätelmänä oli, että ns. tekstittäjän suomi on omanlaisensa kielivariantti, sillä se toisaalta se korostaa alkuperäis-suomen erityispiirteitä ja toisaalta taas käännösuomen erityispiirteitä. (Mts. 228–229.) Tirkkonen-Conditin (2007) mukaan tekstittäjän suomi hyödyntää tehokkaasti juurikin niitä kielen elementtejä, joita hän on tähän asti pitänyt alkuperäis-suomeksi kirjoitettujen tekstien ominaisuuksina, supisuomen uniikkeina piirteinä. Esimerkiksi liitepartikkeli *-han* esiintyi tietyssä ruututekstien otannassa neljä kertaa tiheämmin kuin vastaavasti käännösuomessa ja alkuperäis-suomessa (mp.). Omalta osaltaan oma tutkimukseni pyrkii vastaamaan samankaltaiseen kysymykseen tekstittäjän suomen erityispiirteistä, joskin tarkastelemalla asiaa eri koheesiokeinojen kautta.

## 5 AINEISTO JA MENETELMÄT

Tämän luvun ensimmäinen osa keskittyy tutkimusmateriaaliin, josta tässä tutkimuksessa käsitellyt kielelliset elementit on kerätty. Luvun jälkimmäisessä osassa taas tarkastellaan, millä tutkimuskeinoilla ja millaisia aineistoja tämän tutkimuksen tulosten saamiseen on hyödynnetty.

Tämän tutkimuksen tutkimusmateriaaliksi valikoitui entisessä Joensuun yliopistossa (nykyisin Itä-Suomen yliopisto) Suomen Yleisradion (vastedes Yle) kanavien tv-tekstityksistä koottu ruututekstikorpus. Jukka Mäkisaloon mukaan (2006, 18) kyseinen korpus kuuluu maailman suurimpien korpusten joukkoon ja tällä hetkellä se on tietävästi maailman suurin ammattikäntäjien tekstityksiä sisältävä korpus. Sitä suurempi on toki fanitekstityksistä koostettu OpenSubtitles2016-korpus (kts. lisää Lison & Tiedemann 2016). Yle-ruututekstikorpusten materiaali sisältää lukuisten eri käntäjien tekemiä tv-tekstityksiä, joten tällöin tiettyjen kielellisten elementtien esiintymiseen ei ole vaikuttamassa yksittäisen käntäjän tekemät käänösratkaisut. Kuten Tirkkonen-Condit ja Mäkisalo (2007, 230) huomauttavat, Ylen käntäjät ovat koulutettuja ammattilaisia, jotka asettavat työlleen korkeat laatutavoitteet. Täten on myös oletettavaa, että Ylen käntäjät kiinnittävät sanavalintoihinsa enemmän huomiota kuin amatööriharrastajat. Esimerkiksi Taina Ukkolan (2014) tutkimuksen mukaan ruututekstikorpusten sisältämissä Ylen tekstityksissä esiintyi enemmän propositioita kuin toisessa saman aineiston ruututeksteissä. Lisäksi Ylen voidaan ajatella toimivan yhtenä tekstitysnormien auktoriteettina (Chesterman 2007, 364).

Tutkimuksen vertailumateriaalina toimii myöskin Joensuun yliopistossa koottu Käänössuomen korpus. Se valikoitui tarkastelun kohteeksi, koska se on sisällöltään erityisen sopiva vertailukohde ruututekstikorpukselle. Kaikki Käänössuomen korpuksessa esiintyvät materiaalit ovat ammattikäntäjien tekemiä julkaistuja teoksia (Mauranen 2002). Hannu Kemppasen (2007, 12) mukaan kielenkäytön normeja tutkittaessa referenssikorpusten, eli vertailuaineiston, on oltava kyllin laaja, sisältäen miljoonia tai peräti kymmeniä miljoonia sanoja, jotta siitä saatua vertailutietoa voidaan pitää kyllin luotettavana. Tähän vaatimukseen Käänössuomen korpus vastaakin, sillä se sisältää lähes 10 miljoonaa sanaa (Mauranen 2002), kun taas ruututekstikorpus sisältää lähes 100 miljoonaa sanaa (Mäkisalo 2008). Yhtenä vertailumateriaalin valintaperusteena toimi myös se, että Käänössuomen korpuksen materiaalit ovat suunnilleen samalta aikakaudelta kuin ruututekstikorpusten aineisto. Tällöin

on syytä olettaa, että niissä käytetty kieli on tyyllillisesti samankaltaista ja antaa kattavan katsauksen suomen kielestä ja sen käytöstä.

## 5.1 Yle-ruututekstikorpus

Vuonna 2005 Ylen käännöstoiminnasta vastaava YLE Import lahjoitti Joensuun yliopistolle tutkimuskäyttöön digitaalisen ruututekstiarkistonsa, eli kaikki tekstitykset, joita oli käytetty yhtiön neljällä eri kanavalla vuosien 1991–2004 aikana. Yliopiston sivukampuksella Savonlinnassa tekstitysarkisto, jossa jokainen ohjelma oli omana tiedostonaan, koottiin Yle-ruututekstikorpuksi, joka koostuu miltei 52 500 televisio-ohjelman ja elokuvan tekstityksistä sisältäen yli 99 miljoonaa sanaa. Rakenteeltaan korpus on mahdollisimman samanlainen kuin sen alkuperällä, YLE Importin arkistolla, eli jokainen vuosi on jaoteltu omaan kansioon. (Mäkisalo 2006, 18; 2008.) Tässä tutkimuksessa alkuperäisestä Yle-ruututekstikorpukselta on käytetty nimeä myös nimeä ruututekstikorpus tai Yle-korpus ja sanalla *osakorpus* taas viitataan juurikin tätä tutkimusta varten alkuperäisestä korpukselta johdettuun sivukorpukseseen. Osakorpus pitää sisällään viiden vuoden tekstitykset, vuosilta 2000–2004, sisältäen noin neljänneksen Yle-korpuksen materiaaleista. Osakorpusta käsitellään tarkemmin alaluvussa 5.3. Yle-ruututekstikorpukselta tehtyjä aiempia tutkimuksia on kuvattu tässä työssä alaluvussa 4.2.

Valtaosa Yle-ruututekstikorpuksen sisältämistä tekstitystiedostoista on käännösuomea, mutta se sisältää tiedostoja myös ruotsiksi, saameksi, englanniksi, saksaksi, ranskaksi sekä espanjaksi. Lisäksi korpukselta löytyy monikielisiä tekstitystiedostoja, jotka sisältävät lähinnä kahta kotimaista kieltä, suomea ja ruotsia. (Mäkisalo 2006, 18.) Lähtökieliä ruututekstikorpuksen materiaaleilla on kymmenittäin, aina arabiasta vietnamiin, vaikkakin englanti on niistä hallitsevin (Mäkisalo 2006, 18; 2008). Koska ruututekstikorpus pitää sisällään useammalla kielellä kirjoitettuja tekstejä, on se tyyppiltään monikielinen verrannollinen korpus, vaikkakin valtaosa sen tiedostoista on käännösuomea. Sonja Tirkkonen-Condit ja Jukka Mäkisalo (2007, 223) huomauttavat, että koska Suomessa valtaosa ulkomaisista tv-ohjelmista tekstitetään dubbaamisen sijaan, on täten korpuksen ohjelmisto genreltään hyvin laaja. Korpus sisältää niin tv-sarjojen, elokuvien, uutislähetysten, dokumenttien, urheiluohjelmien yms. tekstityksiä (mp.). Valtaosa materiaalista tosin koostuu dialogista, etenkin fiktiivisten ohjelmien osalta (Mäkisalo 2008).

Yle-korpuksesta koottiin yliopistolla kolme erillistä versiota: ensimmäinen sisältää sellaisenaan YLE Importin alkuperäiset 890-formaatin tekstitystiedostot, joita voidaan tarkastella yhdessä kyseisen televisio-ohjelman tai elokuvan kuvanauhan kanssa – joskin varsinaiset kuvamateriaalit eivät sisälly korpukseen. Toinen versio sisältää tekstimuotoon konvertoidut tekstitystiedostot, joissa näkyvät ruututekstien lisäksi kyseisen ohjelman tunnistetiedot sekä ajastuskoodit. Kolmas versio sen sijaan pitää sisällään ainoastaan yksittäisten ohjelmien repliikkitekstit, siis pelkästään sen mitä ohjelman aikana on televisioruudulla näkynyt. (Mäkisalo 2006, 18.) Tämä korpusversio soveltuukin Mäkisaloon (mp.) mukaan parhaiten kielentutkimukseen, koska sen avulla on mahdollista suorittaa laajamittaisia hakuja tai täsmähakuja niin, ettei oheistieto haittaa varsinaisiin repliikkeihin kohdistuvia hakuja. (Mäkisalo 2006, 18; 2008.)

Itä-Suomen yliopiston hankesivulla sanotaan Yle-ruututekstikorpukselta seuraavasti:

Varsinaisen tieteellisen tutkimuksen kohteena ovat sekä ruututekstien omanlaisensa käännöskieli että audiovisuaalisen eli av-kääntämisen tekniset haasteet. Tutkimuksen kauaskantoisempi tähtäin on av-kääntämisen kielellis-teknisessä ja ohjelmallisessa kehittämisessä, mm. av-kääntämiseen soveltuvien digitaalisten sanakirjojen ja käännösmuistiohjelmien kehityksessä. (Tirkkonen-Condit 2017.)

Lisäksi ruututekstikorpukselle alun perin asetettuihin odotuksiin kuului oletus, että sitä käytetään muun muassa sen tutkimiseen, poikkeavatko ruututekstit muista suomeksi käännettyistä käännösteksteistä ja suoraan suomeksi kirjoitetuista teksteistä jotenkin leksikaalisesti, syntaktisesti tai tekstuaalisesti. (Mäkisalo 2008.) Omalta osaltaan tämä tutkimus pyrkiikin vastaamaan korpukselle asetettuihin kysymyksiin juurikin ruututekstitysten uniikkiainesten ja ns. tekstittäjän suomen kannalta. Kuten tässä tutkimuksessa on jo aiemmin mainittu, ruututekstikorpuksen ensisijainen tutkimuskohde on ruututekstien lingvistinen analysointi ja Mäkisalo (2006, 18) huomauttaakin, että toisen version tekstitystiedostot ovat tekstimuotoisina, jotta tiedostoja olisi helppo käsitellä haku- ja analyysiohjelmilla, kuten WordSmith Toolsin kaltaisilla korpusanalyysityökaluilla. Toista versiota onkin Mäkisaloon (mp.) mukaan tarkoitus käyttää esimerkiksi tekstitysten loogisten ja kielellisten konventioiden tutkimiseen. Tietoturvasyistä korpus sijaitsee ainoastaan yliopiston serverillä, joten sen käyttämistä varten on anottava erikseen lupa.

## 5.2 Käännössuomen korpus

Samoin kuin Yle-ruututekstikorpus, Käännössuomen korpus koottiin entisessä Joensuun yliopistossa, Savonlinnan kampuksella vuosien 1997 ja 2000 välisenä aikana käännöstieteen laitoksen tutkimushankkeen ”Käännössuomi ja kääntämisen universaalit: tutkimus korpusaineistolla” tutkimustavoitteita varten (Mauranen 2004, 72). Käännössuomen korpus on yksikielinen verrannollinen korpus ja tietävästi yksi suurimmista sellaisista (mts. 73). Se sisältää niin alkuperäiskielisiä, eli siis suomeksi alun perin tuotettuja julkaistuja teoksia, kuin käännösprosessin tuotoksena syntyneitä suomennoksia eri lähdekielistä. Nämä teokset ovat kokonaisia, keskenään samoista tekstilajeista ja samalta ajalta. (Jantunen & Eskola 2002, 186.)

Käännössuomen korpus koostuu neljästä osakorpuksesta: (1) alkuperäissuomen korpus, joka sisältää suoraan suomeksi tuotettuja teoksia, (2) käännössuomen korpus, jonka lähteinä on kymmenen eri kieltä (niin indoeurooppalaisia kuin suomalais-ugrilaisia kieliä) sekä (3) käännössuomen korpus, joka sisältää ainoastaan englannista käännettyjä tekstejä (Jantunen & Eskola 2002, 186.) Tässä tutkimuksessa ensimmäistä korpuksesta on käytetty vertailuaineistona niin ruututekstien osakorpuksesta kuin toisen, kolmannen ja neljännen osakorpuksen yhdistelmälle, jota kutsun tässä tutkimuksessa *käännössuomen sivukorpukseksi* erottamaan sitä Käännössuomen korpuksen kokonaisuudesta. Kaiken kaikkiaan Käännössuomen korpus sisältää lähes 10 miljoonaa sanaa, joista noin 4 miljoonaa sanaa on alkuperäissuomella ja noin 6 miljoonaa sanaa, jotka on käännetty suomeen lähinnä englannista ja venäjältä, muut lähdekielet ovat espanja, hollanti, norja, ranska, ruotsi, saksa, unkari ja viro (Mauranen 2002; Jantunen 2004, 40), joista kaikista löytyy korpuksesta kaksi suomennosta. Käännössuomea sisältäviä tekstejä on korpuksessa enemmän kuin alkuperäissuomeksi kirjoitettuja, jotta eri lähdekielisiä suomennoksia voitaisiin kerrata alkujaan suomeksi kirjoitettuihin teksteihin. (Mauranen 2004, 73.)

Käännössuomen korpus sisältää samoja tekstilajeja niin käännettyksi suomeksi kuin alun perin suomeksi kirjoitettuna. Korpus pitää sisällään tekstejä seuraavista genreistä: kaunokirjalliset tekstit, akateemiset kirjoitukset, lastenkirjallisuus, kertoma- ja tietokirjallisuus, viihdekirjallisuus sekä elämäkertakirjallisuus (Mauranen 2004, 74). Käännössuomen korpuksesta on alkuperäissuomelle ja käännössuomelle omat kansionsa, joissa taasen on omat alakansionsa teksteille genrejen mukaan jaoteltuina.

Kaikki korpuksen sisältämät tekstit on kerätty 1990-luvun jälkipuoliskolla, käännökset ovat vuosilta 1995–1999 (Mauranen 2002; 2004). Anna Maurasen (Mauranen 2002) mukaan aineistoon on valittu käännettyjä teoksia sen perusteella, mitä Suomessa paljon luetaan ja käännetään sekä millaisia tekstejä täällä arvostetaan. Näin aineisto pitää sisällään kulttuurillemme tärkeitä käännöksiä ja sen voidaan ”katsoa edustavan ’prototyyppejä’ käännöksiä”. Käännössuomen korpus koottiin tarkoituksena selvittää, millaiset kielelliset piirteet ovat ominaisia käännössuomelle erottaen sen alkuperäissuomesta. Toissijaisena tavoitteena oli tutkia niin sanottuja käännösuniversaaleja, eli käännösten yleisiä ominaisuuksia, jotka eivät ole riippuvaisia lähde- ja kohdekielestä, sekä kehittää tutkimusmetodeja niiden tarkastelemiseen. (Jantunen 2004, 5.)

## 5.3 Osakorpus

### 5.3.1 Osakorpuksen koostaminen

Kuten ylempänä on jo mainittu, tutkimuksen ensisijainen tutkimusmateriaali kerättiin Joensuun yliopiston Yle-ruututekstikorpukselta. Tutkimusmateriaalia valitessa piti päättää, mikä versio alkuperäisestä korpukselta tulisi tutkimuksen käyttöön. Ylempänä on selitetty, että ruututekstikorpukselta koottiin kolme erillistä versiota. Yhtenä tämän tutkimuksen alkuperäisenä tarkoituksena oli siis arvioida, onko ruututekstille varattu aikamäärä mahdollisesti vaikuttanut kääntäjän ratkaisuun käyttää tai olla käyttämättä syntagmaattista sulaumaa jonkin tietyn ruututekstirivin kohdalla. Jotta tutkimuksessa voitaisiin tarkastella, onko ajallinen aspekti vaikuttanut kääntäjän päätökseen joko käyttää tai olla käyttämättä sulaumamuotoja, valikoitui tutkimusmateriaaliksi alkuperäisen Yle-korpuksen toinen versio, jonka tiedostot sisältävät siis paitsi varsinaiset ruututekstit, myös kyseisen ohjelman ajastuskoodit sekä tunnistetiedot (Mäkisalo 2006, 18). Tämä ajallisuuden aspekti kuitenkin jätettiin myöhemmin pois tästä tutkimuksesta sen vaatiman työmäärän tähden. Silloiselta Joensuun yliopistolta anottu käyttöoikeuslupa koski kuitenkin ainoastaan ruututekstikorpuksen toista versiota, joka sisältää siis ruututekstien lisäksi myös tekstitysten ajastuskoodit sekä oheistiedot. Toisen version käytöstä aiheutui lopulta paljon ylimääräistä työtä, sillä tiedostojen sisältämät oheistiedot häirtasivat varsinaisiin repliikkeihin kohdistuvia hakuja. Korpusanalyysityökalun hakuohjelma ei osaa erotella toisistaan tekstin eri osioita, esimerkiksi sanoja ja ajastuskoodeja. Lisäksi ohjelma ei osannut laskea oikein korpuksen sanamäärää, jonka avulla suhteelliset frekvenssiluvut oli tarkoitus laskea, vaan laski

oheistiedot, ajatuskoodit ja varsinaiset repliikkitekstit kaikki yhteen. Tästä syystä ohjelmaa sotkevat tiedot oli poistettava manuaalisesti. Tätä työvaihetta on kuvattu tarkemmin myöhemmin tässä samassa luvussa.

Alkuperäinen Yle-ruututekstikorpus sisältää siis yli 52 000 tekstitystiedostoa (Mäkisalo 2008). Jottei tämän tutkimuksen vaatima työmäärä kasvaisi liian suureksi tutkimuksen laajuuteen nähden, oli tutkittavan korpuksen kokoa pienennettävä. Tästä syystä alkuperäisestä Yle-korpuksesta valittiin ainoastaan osa tutkimusmateriaaliksi. Valintakriteeriksi valikoitui aika, eli ruututekstikorpukselta valittiin tuoreimmat tekstitykset vuosilta 2000–2004. Alkujaan tämä viisi vuotta kattava osakorpus sisälsi lähes 24 000 tiedostoa mukaan lukien uusintanäytökset ja monikieliset ruututekstitykset. Alkuperäisen ruututekstikorpuksen toisesta versiosta ja tarkemmin tuon version viimeisten viiden vuoden tekstitystiedostoista muodostettiin siis tämän tutkimuksen tutkimusmateriaalina käytetty sivukorpus, jota kutsutaan vastedes nimellä *osakorpus*, jottei sitä sekoitettaisi alkuperäiseen Yle-ruututekstikorpuksen.

Kuten tässä tutkimuksessa on jo aiemmin mainittu, Yle-ruututekstikorpus sisältää kaikki Ylen kanavilla vuosien 1991–2004 aikana esitetyt ruututekstitykset, tv-ohjelmien uusintaesitykset mukaan lukien. Koska korpuksen sisältämät tekstitystiedostot on nimetty ohjelmanumeroiden mukaan ja jaettu kansioihin esitysvuosien perusteella, yksittäinen tekstitystiedosto saattaa löytyä koko korpuksesta useitakin kertoja uusintanäytösten tähden. Tämän tutkimuksen tarkoitusta varten, eli erityisten kielellisten elementtien tutkimista ja tilastollista laskentaa varten, osakorpusta kootessa rikottiin alkuperäisen ruututekstikorpuksen kansiorakenne ja kaikki tekstitystiedostot koottiin samaan kansioon. Samalla osakorpuksesta poistettiin ylimääräiset tekstitystiedostot niistä ohjelmista, joita löytyi useampia samannimisenä, eli yksittäisen ohjelman uusinnat. Tällöin saman ohjelman useat tekstitykset eivät jääneet aiheuttamaan päällekkäisiä osumia tutkittaessa yksittäisten kielellisten elementtien esiintymiä.

Sen jälkeen, kun kaikki uusintoja sisältävät tekstitystiedostot oli poistettu, ilmeni uusi ongelma. Suomenkielisten ruututekstirivien lisäksi tiedostot saattoivat sisältää ruotsia, saamea, englantia, saksaa, espanjaa ja ranskaa. Vieraskieliset ruututekstitykset piti poistaa osakorpuksesta sen varmistamiseksi, etteivät nämä tiedostot ja niiden sisältämä informaatio häiritsisi osakorpuksesta haettavia tilastollisia tietoja, lähinnä sanojen määrää. Näitä tilastollisia tietoja tarvittiin osakorpuksen ja Käännössuomen korpuksen välisessä vertailussa

lähinnä suhteellisten frekvenssilukujen laskennassa. Suurin osa näistä häiritsevistä muunkielisistä tiedostoista oli helppo poistaa korpusanalyysityökalun avulla. Mikäli vieraskielinen tekstitystiedosto sisälsi enemmän kuin kolme tekstitysriviä suomea, sisällytettiin tiedosto mukaan osakorpukseen niin, että siitä poistettiin vieraskieliset rivit. Myös suomenkielisistä tekstitystiedostoista poistettiin ruotsinkieliset tekstitysrivit, sillä jotkin tiedostot sisälsivät rinnakkain molempia kotimaisia kieliä. Näin osakorpus päättyi sisältämään ainoastaan suomenkielisiä tekstityksiä.

Valmis, niin sanotusti puhdistettu, osakorpus sisälsi tässä vaiheessa viiden vuoden yksittäiset tekstitystiedostot ohjelmatietoineen ja ajastuskoodeineen. Alustavissa tilastollisissa testeissä kävi kuitenkin ilmi, että ohjelmatietojen sisältämät oheistiedot ja varsinaiset ajastuskoodit haittasivat korpusanalyysityökalun toimintaa, etenkin tilastotieteellisiä tietoja, joista vertailuissa käytetyt suhteelliset frekvenssiluvut laskettiin. Loppujen lopuksi siis, ajallisen aspektin jäätyä pois tutkimusasetelmasta, olisi ruututekstikorpuksen kolmas versio sopinut paremmin tämän tutkimuksen tutkimusmateriaalin lähteeksi. Silloiselta Joensuun yliopistolta haettu käyttöoikeuslupa kuitenkin koski ainoastaan ruututekstikorpuksen toista versiota. Tästä syystä osakorpuksen sisältämistä tiedostoista piti poistaa ohjelmatiedot ja ajastuskoodit manuaalisesti. Tämän viimeisen vaiheen jälkeen viimeistelty osakorpus, jota käytettiin ensisijaisena tutkimusmateriaalina, pitää sisällään 26 454 148 sanaa ja 14 212 tiedostoa.

Vertailumateriaalina tässä tutkimuksessa käytetty Käännössuomen korpus ei vaatinut lainkaan niin suurta työmäärää tutkimustyön aloittamista varten. Kuten ylempänä tässä luvussa on mainittu, Käännössuomen korpuksessa on omat kansionsa alkuperäissuomeksi kirjoitetuille teksteille ja omansa käännössuomen teksteille. Näiden pääkansioiden alla taas tekstit on jaettu alakansioihin seitsemän tekstilajin mukaisesti. Koska korpusanalyysityökalun hakuohjelman voi määrätä etsimään materiaaleja myös alakansioista, tarvitsi hakuja tehdessä asettaa materiaalikansioksi joko alkuperäissuomen tai käännössuomen kansio ja tulokset koskivat automaattisesti koko tuota sivukorpusta. Toisin kuin monikielinen Yle-ruututekstikorpus, Käännössuomen korpus on siis yksikielinen korpus, joten korpuksesta ei myöskään tarvinnut etsiä ja poistaa ”vääränkielisiä” tekstirivejä.

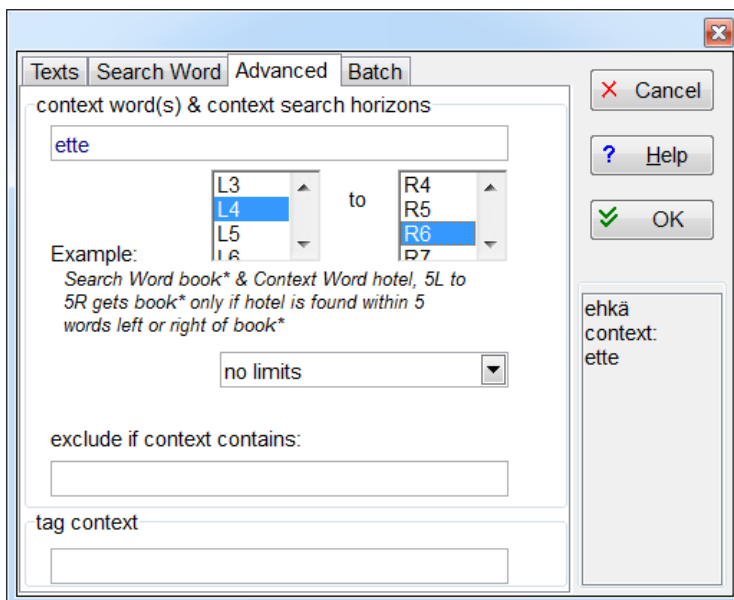


### 5.3.2 Osakorpuksen tutkiminen

Kuten aiemmin on mainittu, tämän tutkimuksen ensisijaisena tarkoituksena oli tarkastella, miten syntagmaattiset sulaumat esiintyvät ruututeksteissä. Tämä alaluku keskittyy kuvaamaan niitä menetelmiä, joilla osakorpukselta saatuja materiaaleja on karsittu ja muokattu, jotta varsinaista tutkimusmateriaalia saatettiin tarkastella. Tässä alaluvussa käsitellään kuitenkin vain ruututekstien syntagmaattisten sulaumien tutkimista. Seuraava alaluku 5.3.3 taas keskittyy vertailuihin, joissa osakorpuksen sulaumamuotojen esiintymistä verrattiin saman korpuksen erillisinä sanoina kirjoitettuihin partikkelin tai konjunktion ja kieltoverbin yhdistelmiin sekä Käännösuomen korpuksessa esiintyviin syntagmaattisiin sulaumiin.

Ennen kuin varsinainen tutkimustyö saattoi alkaa, piti tätä tutkimusta varten koostaa hakusanalista (Liite I & II) ja hakuavainlista, jotka sisältävät kaikki tutkimuskohteena käytetyt kielelliset elementit – kyseiset konjunktiot ja partikkelit sekä niiden sulaumamuodot on jo aiemmin esitelty luvussa 2. Hakusanalista pitää siis sisällään jokaisen yhdeksän sulaumamuotoa sekä jokaisen sulaumamuodon eri persoonamuodot. Esimerkiksi kieltoverbin ja partikkelin *ehkä* sulaumia esiintyy hakusanalistassa kuusi (*ehken, ehket, ehkei, ehkemme, ehkette, ehkeivät*). Kuten luvussa 2 on jo mainittu, syntagmaattisissa sulaumissa monikon 3. persoonan tunnus voi liittyä takavokaaliseen konjunktion joko etu- tai takavokaalisena (ISK 2005, 165). Täten hakusanalistasta löytyy neljä sulaumaa, joissa on seitsemän taivutusmuotoa (*jollei, jottei, muttei* sekä *vaikkei*), sillä näissä neljässä sulaumassa oli mahdollista löytää kaksi erilaista monikon 3. persoonan taivutusmuotoa (esimerkkinä *jolleivat* ja *jolleivät*). Hakuavainlista taas sisältää nimensä mukaisesti hakuavaimet, joilla eri syntagmaattisia sulaumia ja niiden erilleen kirjoitettuja muotoja taivutusmuotoineen haetaan tutkimusmateriaaleista. Esimerkiksi sanaa *ehkemme* on haettu avaimella *ehkemme\**, jotta materiaaleista löytyvät mahdolliset päätteelliset muodot päätyisivät mukaan tuloksiin (hakuavaimista lisää alempana). Lisäksi sulaumamuotojen ja sulautumattomien muotojen vertailua varten hakusanalistassa löytyy omat kohtansa kahtena erillisenä sanana esiintyville muodoille, joissa taas persoonamuotoisuus on liittynyt kieltoverbiin, esimerkkinä partikkelin *ehkä* muodot: *ehkä + en, ehkä + et, ehkä + ei, ehkä + emme, ehkä + ette, ehkä + eivät*. Kattava lista näistä erilleen kirjoitettujen muotojen esiintymistä löytyy liitteestä II.

Hakusanalistan ja hakuavainten toimivuutta testattiin käyttämällä aiempaa tutkimusta varten koottua korpusta (lisää tietoa kts. Koivikko 2008), joka sisältää erään amerikkalaisen romaanin suomennoksen. Koeasetelma paljasti hakusanalistassa muutamia puutteita, jotka korjattiin vastaamaan osakorpuksen tarpeita. Kiinnostavaa kyllä, tässä koetutkimuksessa sulaumamuodot olivat harvinaisempia kuin sulautumattomat muodot. Kirjallisessa materiaalissa tämän yhden romaanin osalta sulaumat olivat harvinaisempia kuin kahtena erillisenä sanana esiintyvät muodot, mutta yhden kääntäjän tekemät subjektiiviset ratkaisut vaikuttivat tässä tietenkin ratkaisevasti toisin kuin YLE-ruututekstikorpuksessa ja Käännössuomen korpuksessa, jotka molemmat sisältävät useiden kääntäjien kääntämiä tekstejä.



Kuva 2. Kontekstisanan määrittäminen WordSmith Toolsissa

Hakusanalistan sulaumamuodot haettiin aineistosta yksi kerrallaan jokaista persoonamuotoa myöden ja samalla tuloksista luotiin erilliset konkordanssilistat, esimerkiksi yhdellä listalla on kaikki *miksen*-sulauman esiintymät ja toisella taas vastineen *miksi + en* tulokset. Koska hakuohjelmat eivät oikeasti ymmärrä sanoja, vaan lukevat merkkijonoja (Kujamäki & Jääskeläinen 2001, 251), on näissä aineistohauissa käytetty niin sanottua villiä korttia, eli asteriskiä (\*). Esimerkiksi sulaumaa *ehken* haettiin hakuavaimella *ehken\**, jolloin hakuohjelma ottaa huomioon myös liitepartikkelin sisältävät muodot, kuten *ehkenpä*. Kaksisanaiset erilleen kirjoitetut muodot haettiin aineistosta laittamalla konkordanssiohjelman hakuavaimeksi haluttu partikkeli ja persoonamuotoinen kieltoverbi taas niin sanotuksi

kontekstisanaksi, jota ohjelma hakee hakusanan läheisyydestä – tässä tutkimuksessa neljä sanaa vasemmalle ja kuusi sanaa oikealle hakusanasta. Kuvassa 2 näkyy tilanne, jossa hakusanaksi on asetettu partikkeli *ehkä* ja kontekstisanaksi persoonamuotoinen kieltoverbi *ette* niin, että kontekstisana esiintyy hakutuloksissa korkeintaan neljä sanaa vasemmalle hakusanasta katsottuna (L4) tai korkeintaan kuusi sanaa oikealle hakusanasta (R6).

Kontekstisanan hakukriteeriksi valikoituivat asemat L4 ja R6, sillä yllämainitussa koetustauksessa tuli ilmi, että partikkeli *ehkä* saattoi esiintyä lauseissa kieltoverbin edellä, ja kaiken kaikkiaan kahtena erillisenä sanana kirjoitettujen muotojen kohdalla persoonamuotoinen kieltoverbi saattoi sijaita lauseessa hyvinkin kaukana partikkelin jäljessä, kuten esimerkiksi lauseessa ”Pitää toivoa, *että* tien ali kulkevia isoja kaapeleita *ei* kaiveta poikki”, jossa kontekstisana *ei* on asemassa R6. Vaikka *ehkä* oli ainoa partikkeleista, joka saattoi esiintyä selkeästi ennen kieltoverbiä, käytettiin kaikissa erilleen kirjoitettujen sanojen hauissa samaa hakukriteeriä, jotta tutkimusasetelma pysyisi samana. On toki mahdollista, että aineistossa esiintyy myös tapauksia, joissa partikkeli *ehkä* esiintyy kauempana kuin neljäntenä sanana kieltoverbin edellä, tai että varsinkin paljon määritteitä sisältävissä lauseissa persoonamuotoinen kieltoverbi esiintyy jopa kauempana kuin kuuden sanan päässä partikkelista, mutta hakukriteeri R7 tai jopa R8 olisi taas tuottanut lisää tämän tutkimuksen kannalta epäolennaisia löydöksiä. Kuitenkin nykyisilläkin hakukriteereillä 68 prosenttia kaksisanaisten erilleen kirjoitettujen muotojen osumista oli turhia KWIC-rivejä (tarkemmin selitetty alaluvussa 4.1), jotka piti manuaalisesti poistaa tutkimustuloksista, sillä hakuohjelma ei tunnista materiaalista, mikäli virkkeen raja kulkee tällaisen hakukriteerin rajan sisällä.

Jokainen konkordanssilista käytiin siis tarkasti läpi ja varmistettiin, ettei lopulliseen tutkimusmateriaaliin pääse mukaan muotoja, jotka eivät tue tämän tutkimuksen tarkoituksia. Tämä oli tarpeellinen vaihe, sillä osa löydetyistä esiintymistä johtui kirjoitusvirheistä (esimerkiksi *elletten*), muutoin virheellisistä muodoista tai villin kortin käytöstä, esimerkiksi liitepartikkelien tähden sulaumaa *ehket* haettiin hakusanalla *ehket\**, jolloin kyseisiin esiintymiin tulivat mukaan myös monikon 2. persoonan muodot. Kuten ylempänä mainittiin, monet esiintymät olivat epäolennaisia tämän tutkimuksen kannalta, varsinkin erilleen kirjoitettujen muotojen osalta, ja täten ne piti tarpeettomina poistaa tutkimusmateriaalista. Usein poistot ovat johtuneet siitä, että kontekstisana on esiintynyt eri virkkeessä kuin hakusana.

Kuten yllä on kerrottu, valtaosa varsinaisesta tutkimusmateriaalista poistetuista löydöksistä johtui siitä, että hakusana ja kontekstisana olivat materiaalissa eri virkkeissä. Lisäksi materiaalista poistettiin ne KWIC-rivit, joissa hakusana ja kontekstisana ovat olleet eri repliikeissä, eli ne ovat näkyneet eri aikaan ruudussa (esimerkkinä lause ”...**että** Myanmar - **ei** voi olla välittämättä...”). Tällaisten poistojen perusteluna on se, että näissä tekstitysrepliikeissä ymmärrettävyys kärsisi, mikäli partikkeli ja kieltoverbi muutettaisiin sulaumamuodoksi – kuuluuhan kieltosana olennaisesti predikaattiverbin oheen. Kaksisanaisten erilleen kirjoitettujen muotojen kohdalla materiaalista on siis poistettu kaikki sellaiset esiintymät, jotka eivät vastaa sulaumamuotojen esiintymiä, eli tapaukset, joissa erilleen kirjoitettuja muotoja ei voisi yhtä hyvin muuttaa sulaumamuodoiksi esimerkiksi sen tähden, että niiden merkitys muuttuisi, kuten lauseessa ”**Eivät** he minua **siksi** raiskanneet”. Myöskin tapaukset, joissa sanajärjestyksestä olisi pitänyt muuttaa merkittävästi, on jätetty pois materiaalista, esimerkiksi lause ”Teatterit **eivät** kenties **siksi** uskalla panostaa siihen”. Lisäksi aineistosta on poistettu ne tapaukset, joissa joko haku- tai kontekstisana esiintyi sitaatissa ilman toista, kuten lauseessa ”Hän vastasi, **että** **ei** tarpeeksi kuumassa”. Niin ikään selkeät runokäännökset on jätetty pois materiaalista, sillä niissä sanavalintojen voi katsoa olevan tärkeämpiä esimerkiksi soinnillisuuden kannalta, jolloin sulaumamuodot ja erilleen kirjoitetut muodot eivät automaattisesti vastaa toisiaan. Aiemmin mainittujen tapausten ohella materiaalista poistettiin myös tapaukset, joissa *ei*-kieltoverbin käytöllä on painoarvo tai se on muutoin tarpeellinen, kuten lauseissa ”**Miksi** sanoit **ei**?”, ”**Jos ei**, niin ei” ja ”Ehkä, **ehkä en**”.

Hankalimman tapausten luokan muodostivat lyhyet lauseet, esimerkiksi sivulause virkkeessä ”Ronny hyväksyi Carolinen joukkoon, **mutta** Jenny **ei**”. Toki sivulause olisi mahdollista sanoa muodossa ”**muttei** Jenny”, mutta aineistoa käsitellessä jäin miettimään, kärsiikö lyhyiden lauseiden luettavuus, mikäli kaksisanaisten ilmaus muutettaisiin näissä tapauksissa sulaumamuodoiksi. Tuolloinhan lauseessa olisi vain sulaumapartikkeli tai -konjunktio ja yksi muu sana. Koska sulaumamuotoja tutkittaessa oli kuitenkin havaittu, että sulaumat saattoivat esiintyä yksinään, joten tällaiset lyhyet lauseet jätettiin myös erillisinä sanoina kirjoitettujen muotojen konkordanssilistoihin. Pelkät lyhyet lauseet, jotka sisälsivät 2–3 sanaa eli korkeintaan yhden muun sanan konjunktion tai partikkelin ja kieltoverbin lisäksi, kuitenkin erotettiin omiksi konkordanssilistoikseen lisätutkimuksia varten.

Kun materiaalista käytiin läpi poistettavia KWIC-rivejä, tuli myös vastaan rivejä, jotka sisällytettiin kahteen konkordanssilistaan. Kun partikkeli *ehkä* esiintyi samassa lauseessa kuin jokin muista tutkituista konjunktioista tai partikkeleista, jäi tulkinnanvaraiseksi kumman kanssa kieltoverbi yhdistyisi, mikäli kaksisanainen erilleen kirjoitettu muoto muunnettaisiin syntagmaattiseksi sulaumaksi. Esimerkiksi lauseessa ”Ajattelin, *että* *ehkä* ne *eivät* tule enää” on kieltoverbin mahdollista sulautua niin konjunktioon *että* kuin partikkeliin *ehkä* ilman, että lauseen merkitys muuttuu (vrt. ”*etteivät* ne *ehkä* tule enää” sekä ”että *ehkeivät* ne tule enää”). Tällaiset monitulkintaiset tapaukset on sisällytetty siis kummankin partikkelin esiintymiin (tässä esimerkissä konkordanssilistaan *ehkä* + *eivät* sekä *että* + *eivät*). Toki kyseisten tapauksien sisällyttäminen kahden eri partikkelin alle vääristää hieman tuloksia, mutta määrällisesti tällaisia tapauksia löytyi vain kourallinen.

Siinä missä sulaumamuotojen haut tuottivat vain suhteellisen vähän tämän tutkimuksen kannalta epäolennaisia löydöksiä (esimerkiksi hakusanalla *ellen* löytyi nimeen Ellen viittaavia osumia), tuottivat kahtena sanana erilleen kirjoitetut muodot valtavasti ylimääräisiä löydöksiä (esimerkiksi osakorpuksesta löytyi haulle *mutta* + *ei* 30 988 osumaa, joista 17 864 osumaa oli relevantteja tälle tutkimukselle), sillä kuten ylempänä on selitetty, korpusanalyysityökalu ei ymmärrä virkerajoja, ellei niitä ole erikseen materiaaliin määritetty. Kaiken kaikkiaan erilleen kirjoitettujen muotojen esiintymistä vain 32 prosenttia oli tämän tutkimuksen aihepiiriin kuuluvia löydöksiä, kun taas syntagmaattisten sulaumien kohdalla sama luku oli 91 prosenttia. Sulaumamuotojen osalta valtaosa hukkaosumista aiheutui villin kortin käytöstä, esimerkiksi haku *jottet*\* tuottaa mukanaan myös osumat muodolle *jottette*.

Kun epäolennaiset rivit oli poistettu, jokainen konkordanssilista järjestettiin uudelleen kolmeen kertaan käyttäen apuna kolmea eri kriteeriä. Ensimmäiseksi konkordanssilista järjestettiin käyttämällä hakusanaa pääasiallisena lajittelukeinona, jolloin listassa näkyi keskimmäisenä hakusanan eri muodot. Tämän lajittelun tarkoituksena oli erotella hakusanan perusmuodot ja päätteelliset muodot toisistaan, jonka jälkeen nämä tallennettiin erillisiksi listoikseen, jotta eritoten hakusanan perusmuotoja saatettiin tarkastella tarkemmin. Perusmuodot ja päätteelliset muodot piti erottaa toisistaan, sillä suhteelliset frekvenssiluvut laskettiin pelkistä perusmuodoista. Erilleen kirjoitetuissa ilmauksissahan liitepartikkeli voi asettua niin kieltoverbin kuin konjunktioinkin perään, toisin kuin sulaumapartikkeleissa, joissa liitepartikkeli liittyy sulaumamuodon perään (ISK 2005, 164). Kaksisanaiset sulautumattomat

muodot siis haettiin aineistosta ainoastaan niiden perusmuodossa, sillä kaksisanaisissa vastineissa villin kortin käyttäminen olisi aiheuttanut liian suuren lisäyksen tarkistettaviin konkordanssiriveihin tämän tutkimuksen kokoon nähden. Esimerkiksi haku *vaikka + en* tuotti tuloksena 1 977 esiintymää, joista 345 oli varsinaiseen tutkimusaineistoon sopivia esiintymiä, kun taas haku *vaikka\* + en* toi esiin alkujaan 2 042 tulosta ja haku *vaikka + en\** 3 289 esiintymää. Toki liitepartikkelilliset muodot olisivat tuoneet oman lisänsä vertailuun, mutta tämän tutkimuksen puitteissa tähän ei ollut mahdollisuuksia.

Toisessa konkordanssilistojen uudelleen järjestelyssä listat lajiteltiin käyttämällä ensisijaisena lajittelukeinona R1:tä (suoraan hakusanaa seuraava sana), toissijaisena R2 (toinen sana oikealle hakusanasta) ja kolmantena keinona R3 (kolmas sana oikealle hakusanasta). Tilanne nähtävissä kuvassa 3, kun hakusanana on *jotta* ja kontekstisanana *eivät*. Tämän lajittelun tarkoituksena oli muokata konkordanssinäkymää niin, että esiintymien aakkosellista listaa voitiin käyttää tutkimaan hakusanojen läheisimpiä sanoja, tässä tapauksessa siis hakusanan jäljessä tulevia sanoja, etenkin verbejä ja seuraako hakusanaa mahdollisesti persoonapronomini. Siis kaiken kaikkiaan etsittiin kollokaatteja, eli hakusanan lähimpiä sanoja.



Kuva 2. KWIC-näkymän lajittelukeinona R1–R3

Kolmannessa vaiheessa lista järjestettiin uudelleen toistamalla toisen vaiheen lajitteluita, mutta tällä kertaa käyttäen L1:tä (suoraan hakusanaa edeltävä sana) pääasiallisena lajittelukeinona ja L2:ta (toinen sana vasemmalle hakusanasta) toissijaisena lajittelukeinona,

samoin L3:tta (kolmas sana vasemmalle hakusanasta) kolmantena lajittelukriteerinä. Tämän lajittelun tarkoituksena oli tutkia tarkemmin, miten sulaumat sijoittuvat lauseissa ja virkkeissä, etenkin partikkeli *ehkä*, joka voi muista tutkituista konjunktioista ja partikkeleista poiketen esiintyä kieltoverbin jälkeen.

### 5.3.3 Vertailut

**Osakorpuksen sisällä.** Jotta syntagmaattisten sulaumien esiintymistä television ruututeksteissä voitiin arvioida, piti sulaumista saatuja tuloksia tietenkin verrata kahtena erillisenä sanana kirjoitettuihin muotoihin. Kuten tässä tutkimuksessa on jo aiemmin mainittu, näitä erillisinä sanoina esiintyviä muotoja tutkittiin vain osakorpuksen materiaaleista. Samojen kielellisten elementtien tutkiminen Käännössuomen korpuksen materiaaleista ja vertailu osakorpuksen vastaaviin tuloksiin olisi varmasti ollut hedelmällistä, mutta tätä tutkimusta varten työmäärä olisi noussut liian laajaksi. Osakorpuksen sisäistä vertailuasetelmaa varten luotiin osakorpuksesta korpusanalyysityökalun avulla sanalista. Tällä toiminnolla saatiin esiin korpuksen tilastollisia tietoja, esimerkiksi sen sanamäärä. Samaa tietoa hyödynnettiin myös, kun osakorpuksen tuloksia syntagmaattisten sulaumien osalta verrattiin Käännössuomen sivukorpusten vastaaviin tuloksiin.

Koska painotus syntagmaattisten sulaumien ja niiden erilleen kirjoitettujen muotojen vertailussa oli kvantitatiivinen, eli löydösten suhteelliseen esiintymiseen perustuva, aloitettiin varsinainen vertailu laskemalla kaikkien hakusanojen ja niiden taivutusmuotojen esiintymistä suhteelliset frekvenssiluvut. Kuten tässä tutkimuksessa on aiemminkin todettu, ei sulaumamuotojen liitepartikkelillisiä muotoja otettu huomioon suhteellisia frekvenssilukuja laskettaessa, sillä näitä muotoja ei haettu lainkaan erilleen kahtena sanana kirjoitetuista sulautumattomista muodoista, jottei tutkimuksen laajuus kasvaisi liian suureksi. Ensin siis kunkin sulauman ja sen taivutusmuotojen esiintymistä poistettiin päätteelliset muodot, jotka tallennettiin omiksi hakulistoikseen, ja eri konkordanssilistojen esiintymien lukumäärät kirjattiin ylös. Syntagmaattisten sulaumien päätteettömien muotojen sekä kahtena sanana esiintyvien muotojen esiintymismääristä laskettiin suhteelliset frekvenssiluvut kertomalla kunkin taivutusmuodon esiintymismäärä 1 000 000:lla ja jakamalla tulo osakorpuksen sanamäärällä (kaava nähtävissä alla kuvassa 4). Suhteutetun sanajoukkion lukemaksi valittiin miljoona, koska esimerkiksi 10 000 sanan joukkiossa harvinaisempien syntagmaattisten sulaumamuotojen suhteelliseksi frekvenssiluvuksi olisi tullut alle 0,00.

### Kuva 3. Suhteellisten frekvenssilukujen laskentakaava

Näitä suhteellisia frekvenssilukuja vertaamalla oli siis mahdollista tarkastella, miten monta kertaa mikäkin kielellinen elementti esiintyy miljoonan sanan joukkiossa osakorpuksen ruututeksteissä. Koska tässä kohdin molemmat tutkitut kielelliset elementit kerättiin samasta korpuksesta, olisi vertailu niiden välillä voitu suorittaa myös ilman suhteellisten frekvenssien laskemista. Kolmen erikokoisen korpuksen vertailua varten suhteelliset frekvenssiluvut oli kuitenkin tarpeen laskea, joten myös osakorpuksen sisäisessä vertailussa haluttiin käyttää samoja suhdelukuja tutkimuksen yhdenmukaisuuden tähden. Tarkat listaukset suhteellisista frekvenssiluvuista löytyvät kokonaisuudessaan liitteistä III–IV. Frekvenssilukujen laskemisen jälkeen esiintymien määrää tarkasteltiin taulukoiden avulla. Vertailun tuloksista kerrotaan tarkemmin tuloksia käsittelevässä luvussa 6.

Tutkimuksessa otettiin huomioon myös seikka, joka tuli esiin, kun samasta materiaalista etsittiin erilleen kahtena sanana kirjoitettuja muotoja ja näiden kielellisten elementtien konkordanssilistoja puhdistettiin epäolennaisista KWIC-riveistä – nimittäin lyhyet lauseet (esimerkkinä seuraavan virkkeen sivulause: ”Pelkäsitkö, *että et* jaksaisi?”). Lisäksi kahtena sanana kirjoitettavien muotojen löydöksissä ilmeni paljon lauseita, jotka sisälsivät vain hakusanan ja kontekstisanana toimineen persoonamuotoisen kieltoverbin, kuten kysymyksessä ”*Miksi et?*” Koska erillisinä sanoina esiintyvien muotojen muuttaminen syntagmaattiseksi sulaumiksi tällaisissa lyhyissä virkkeissä mietitytti materiaalin karsintavaiheessa, haluttiin tutkimuksessa tarkastella erikseen näitä lyhyitä lauseita. Täten syntagmaattisten sulaumien esiintymistä tutkittiin yksi- tai kaksisanaisia lauseita muista esiintymistä erillään ja niiden tuloksia verrattiin kahtena erillisenä sanana esiintyvien muotojen vastaaviin lyhyisiin lauseisiin, jotka sisälsivät korkeintaan yhden muun sanan sulaumapartikkelin ja kieltoverbin lisäksi, eli erilleen kirjoitettujen muotojen osalta kaksi- tai kolmisanaanisiin virkkeisiin ja lauseisiin. Sekä syntagmaattisten sulaumien että kahtena erillisenä sanana esiintyvien muotojen esiintymistä erotettiin omiksi konkordanssilistoikseen lyhyissä lauseissa ilmenevät esiintymät ja näistä listoista vielä erikseen yksin esiintyvät tapaukset, eli syntagmaattisten sulaumien osalta yksisanaiset lauseet, jotka sisälsivät vain sulaumamuodon, ja erilleen kirjoitettujen muotojen osalta lauseet, jotka sisälsivät ainoastaan hakusanan ja kontekstisanana toimineen persoonamuotoisen kieltoverbin välimerkkien



osuutta laskematta. Esimerkiksi virke ”*Ellei* sitten...” sisällytettiin lyhyiden lauseiden tuloksiin. Näiden hakujen tulokset taulukoitiin ja niistä laskettiin prosenttiosuudet verraten lyhyiden lauseiden esiintymistä kaikkiin kyseisen kielellisen elementin osumiin sekä yksin esiintyvien muotojen esiintymistä lyhyiden lauseiden osumiin. Lisää tietoa tämän vertailun tuloksista alaluvussa 6.1.3 ja kaikki taivutusmuodot kattavat laskelmat liitteistä V–VI.

***Kolme korpusta.*** Jotta osakorpuksen ruututeksteistä löytyviä syntagmaattisia sulaumia ja niiden käytön yleisyyttä voitiin arvioida, piti tutkia myös, miten samaiset kielelliset elementit esiintyvät alun perin suomeksi kirjoitetuissa teksteissä sekä muussa käännetyissä suomessa. Kuten aiemmin on huomautettu, kolmen korpuksen vertailussa tarkasteltiin vain tämän tutkimuksen pääkohdetta, eli syntagmaattisia sulaumia taivutusmuotoineen. Ensimmäinen vertailuun tarvittava tieto oli tilastollinen, eli korpusten sanamäärät. Osakorpuksen sanamäärä oli selvitetty jo aiemmassa vaiheessa (tätä vaihetta on kuvattu edeltävässä alaluvussa); nyt korpusohjelman avulla luotiin Käännössuomen sivukorpuksille sanalistat, joista selvisi korpusten sanamäärät.

Samoin kuin syntagmaattisten sulaumien ja niiden erilleen kirjoitettujen muotojen vertailussa, myös kolmen korpuksen vertailussa painopiste oli kvantitatiivisessa tutkimuksessa, eli tutkimuksessa painotettiin sulaumamuotojen suhteellista esiintymistiheyttä. Toisin kuin osakorpuksen sisäisessä tarkastelussa, kolmen korpuksen vertailussa suhteellisten frekvenssilukujen laskeminen oli hyvinkin tarpeen, sillä korpusten sanamäärissä oli suuri ero, ja täten pelkkä esiintymien lukumäärä ei riitä vertailujen tekemiseen. Alkuperäissuomen korpuksessahan sanamäärä on vajaa neljä miljoonaa, kun taas osakorpuksessa sanoja löytyy miltei 26,5 miljoonaa. Koska korpusten sanamäärissä on niin merkittävä ero, ei esiintymien lukumääriä voinut vertailla keskenään ilman suhteellisten frekvenssilukujen laskemista. Frekvenssiluvut on siis suhteutettu siihen, miten monta kertaa kukin kielellinen elementti esiintyy korpuksessaan miljoonan sanan joukkiossa (kaava löytyy aiempaa kuvasta 4). Suhteellisia frekvenssilukuja laskettaessa ei otettu huomioon syntagmaattisia sulaumia, jotka päättyivät liitepartikkeliin. Vaikka kolmen korpuksen vertailussa oli mukana vain syntagmaattiset sulaumamuodot, oli osakorpuksen sisäisessä vertailussa mukana kahtena sanana erilleen kirjoitettavat muodot, joista ei niiden vaatiman suuren työmäärän tähden haettu lainkaan liitepartikkeliin päättyviä muotoja, ja näin tutkimusasetelma haluttiin pitää

yhdenmuotoisena. Tarkempi listaus eri korpusten suhteellisista frekvenssiluvuista on nähtävissä kokonaisuudessaan liitteestä III.

Varsinainen kolmen korpuksen vertailu aloitettiin hyödyntämällä aiemmassa vaiheessa koottuja hakusanalistoja (vaihe selitetty edeltävässä alaluvussa), ja Käännösuomen korpuksen sekä alkuperäissuomen että käännösuomen sivukorpuksista haettiin kaikki hakusanalistoissa esiintyvät syntagmaattiset sulaumat sekä niiden taivutusmuodot, jonka jälkeen hakuohjelmalla saadut tulokset tallennettiin erillisiksi konkordanssilistoiksi. Lisäksi tulokset taulukoitiin osakorpuksen tulosten rinnalle, ja vertailut suoritettiin näistä taulukoista. Kuten tässä luvussa on aiemmin mainittu, Käännösuomen korpus on yksikielinen, verrannollinen korpus, josta haettiin vain syntagmaattisia sulaumia, eikä lainkaan niiden kahtena erillisenä sanana kirjoitettuja muotoja, joten konkordanssilistoihin ei tarvinnut Käännösuomen korpuksen osalta tehdä juurikaan poistoja.

Kolmen korpuksen vertailussa keskityttiin ensisijaisesti suhteellisiin frekvenssilukuihin, mutta tutkailun alle päätyivät myös tuloksista erityisesti esille nousevat seikat. Esiintymismäärien lisäksi tässä tutkimuksessa tarkasteltiin syntagmaattisten sulaumien päätteellisten muotojen käyttöä, sulaumamuodoissa esiintyviä vokaalisoinnun muutoksia sekä sulaumamuotojen persoonamuotoisuutta, eli miten eri korpuksissa on käytetty yksikkö- ja monikkomuotoisia sulaumapartikkeleita ja -konjunktioita. Samoja tutkimuskohteita tarkasteltiin niin korpuskohtaisesti kuin yksittäisten syntagmaattisten sulaumien ja niiden taivutusmuotojen osalta. Tarkka kuvaus kolmen korpuksen vertailun tuloksista löytyy seuraavan luvun alaluvusta 6.2. Vertailut toteutettiin samoin kuin osakorpuksen sisäinen vertailu, eli tarkasteltavat asiat tallennettiin omiksi konkordanssilistokseen – kunkin syntagmaattisen sulauman jokainen taivutusmuoto erikseen, mikäli kyseinen tutkittava asia löytyi niiden tuloksista. Sitten tulokset taulukoitiin ja tuloksille laskettiin suhteelliset frekvenssiluvut, joista varsinaiset vertailut suoritettiin.

## 6 TULOKSET

Kuten tämän tutkimuksen johdannossa on jo linjattu, tutkimuksen tavoitteena oli tarkastella, miten suomen kieltöverbi *ei* yhdistyy syntagmaattiseksi sulaumaksi partikkeleiden *ehkä, että, jotta, jos, miksi, mutta, siksi* ja *vaikka* kanssa television ruututeksteissä. Tämän tutkimuksen ensisijainen tavoite taas oli tutkia, miten paljon syntagmaattisia sulaumia esiintyy ruututeksteissä ja verrata tuloksia muussa käännösuomessa sekä alun perin suomeksi kirjoitetuissa teksteissä, eli supisuomessa, ilmenevien sulaumamuotojen määrään. Vertailun tarkoituksena oli tarkastella, seuraavatko ruututekstit alkuperäissuomen tai muun käännetyn suomen konventioita ja normeja, ja toisaalta havainnoida, eroavatko kaiken kaikkiaan käännetyn suomen konventiot syntagmaattisten sulaumien osalta alkuperäissuomen konventioista. Oletuksena oli, että alkuperäissuomessa ja muussa käännetyissä suomessa esiintyisi vähemmän syntagmaattisia sulaumia kuin ruututeksteissä, joihin taas ruututekstikäntämisen rajoitteet vaikuttavat ensisijaisesti.

Toissijaisena tavoitteena oli tutkia, miten paljon syntagmaattisia esiintyy tekstityksissä, sekä vertailla, miten paljon samassa materiaalissa esiintyy sulaumien erilleen kahtena sanana kirjoitettuja muotoja, esimerkkinä *siksi en*. Kiinnostavaa olisi toki vertailla, miten paljon muussa käännetyissä suomessa sekä alkuperäissuomessa esiintyy erilleen kirjoitettuja muotoja samoista kielellisistä elementeistä, eli miten Käännösuomen korpuksessa syntagmaattisten sulaumien käyttö vertautuu erilleen kirjoitettuihin muotoihin, mutta se olisi tuottanut liikaa työtä tähän tutkimukseen.

Koska molemmissa tavoitteissa oli suurena osana mukana vertailuasetelma, on tässä luvussa esitelty tulokset tutkielman tavoitteiden mukaisesti. Koska ensin on haluttu käsitellä pelkän osakorpuksen tuloksia, on tulokset esitetty edempänä tavoitteiden suhteen päinvastaisessa järjestyksessä. Ensimmäinen alaluku käsittelee sitä, miten ruututeksteissä on käytetty syntagmaattisia sulaumia, ensin yleisluontoisesti ja sitten konjunktio- ja partikkelikohtaisia tuloksia. Ensimmäisen alaluvun viimeinen osa tarkastelee osakorpuksen sisäistä vertailuasetelmaa, eli syntagmaattisten sulaumien ja samojen kielellisten elementtien erilleen kahtena sanana kirjoitettujen muotojen esiintymistä ruututeksteissä. Kuten aiemmin on jo mainittu, näitä erillisinä sanoina esiintyviä muotoja tutkittiin vain osakorpuksesta vertailuna syntagmaattisten sulaumien käyttöön. Jälkimmäinen alaluku tarkastelee vertailuasetelmaa, miten syntagmaattisten sulaumien käyttö ruututeksteissä vertautuu niiden käyttöön muussa

käännössuomessa ja alkuperäissuomessa. Tarkat listaukset osakorpuksen ja Käännössuomen korpuksen tuloksista löytyvät tämän työn liitteistä.

## 6.1 Osakorpuksen tulokset

### 6.1.1 Syntagmaattiset sulaumat

*Iso Suomen kieliopin* (2005, 165) mukaan partikkelien ja kieltoverbin syntagmaattiset sulaumat ovat yleiskielessä yleisempiä kuin samojen kielellisten elementtien erilleen kahtena sanana kirjoitetut muodot. Täten tässä tutkimuksessa lähdettiin liikkeelle laskemalla kaikkien syntagmaattisten sulaumien esiintymät kaikissa niiden taivutusmuodoissa. Kaiken kaikkiaan osakorpukselta löytyi 67 085 esiintymää. Yli 26 miljoonaa sanaa käsittävässä osakorpuksessa syntagmaattiset sulaumat kaikkine taivutusmuotoineen muodostivat noin 0,25 prosenttia koko korpuksen sanastosta. Kuten on nähtävissä taulukosta 3, eniten esiintymiä löytyi sulaumamuotojen *ettei*, *miksei* ja *ellei* taivutusmuodoista, harvinaisimpia olivat sulaumien *siksei* ja *jottei* taivutusmuodot. Tämä tulos tukee ISK:n toteamusta, jonka mukaan ”sulaumakonjunktioista selvästi yleisin on *ettei* taivutusmuotoineen” (mts.).

Taulukko 3. Osakorpuksen kaikki sulaumapartikkelit

SYNTAGMAT	OK
ehkei	802
ellei	6 347
ettei	39 923
jollei	2 020
jottei	534
miksei	11 558
muttei	3 886
siksei	17
vaikkei	1 998
<b>yhteensä</b>	<b>67 085</b>

Taulukossa 3 on esillä partikkelikohtaiset tulokset kaikista niiden esiintymistä (taivutusmuotojen tulokset eritelty liitteessä I). Lisäksi esiintymien määrät laskettiin erikseen ilman liitepartikkelin sisältämiä tuloksia, joita oli yhteensä 66 808. Näistä päätteettömistä muodoista taas laskettiin esiintymille suhteelliset frekvenssiluvut sen mukaan, miten ne esiintyivät miljoonan sanan joukkiossa. Suhteellisista frekvenssiluvuista puhutaan tarkemmin

osakorpuksen sisäisen vertailun kohdalla alaluvussa 6.1.3 ja erityisesti, kun osakorpuksen materiaaleja on verrattu Käännösuomen aineistoon alaluvussa 6.2.

Kun syntagmaattisten sulaumien eriteltyjä taulukoita tarkasteltiin, tuli selkeästi esiin, että ruututekstityksistä löytyi sulaumamuodosta riippumatta enemmän yksiköllisiä persoonamuotoja kuin monikollisia. Kaiken kaikkiaan osakorpuksen hakutuloksista miltei 89 prosenttia esiintyi yksiköllisessä persoonamuodossa. Yleensä ottaen yksikön 3. persoona oli sulaumamuotojen yleisin persoonamuoto, joka kaikkien sulaumamuotojen yhteistuloksissa kattoi noin 55 prosenttia löydetyistä tapauksista. Persoonamuotoista syntagmaattista sulaumaa harvemmin kuitenkaan seurasi ruututeksteissä persoonapronomini; yhteensä vain kahdeksassa prosentissa esiintymistä sulaumamuodon jäljessä esiintyi persoonapronomini. Yleisimmin persoonapronomini löytyi monikon 3. persoonan sulaumamuotojen vierestä, etenkin etuvokaalisten (-eivat) muotojen seurana. Syntagmaattisten sulaumien esiintymistä tutkittiin myös niiden lauseenalkuisuutta. Koska suurin osa tutkimuksessa tarkastelluista partikkeleista ovat konjunktioita (paitsi *ehkä*, *miksi* ja *siksi*), sijaitsivat aineiston syntagmaattiset sulaumat pääosin virkkeen tai lauseen ensimmäisenä sanana, joskin muutamia poikkeuksiakin löytyi. Lauseenalkuisuuden tuloksia on käsitelty tarkemmin partikkelikohtaisissa tuloksissa edempänä.

Kuten aiemmin on jo tullut ilmi, vaikka syntagmaattiset sulaumat itsessään koostuvat partikkelista ja persoonamuotoisesta kieltoverbistä, saattavat ne silti ottaa peräänsä liitepartikkelin. Vaikka liitepartikkeliin päättyviä muotoja ei otettukaan osaksi osakorpuksen sisäistä vertailuasetelmaa, eli syntagmaattisten sulaumien ja samojen kielellisten elementtien erilleen kirjoitettujen muotojen vertailua keskenään, tarkasteltiin liitepartikkelin sisältäviä sulaumamuotoja erikseen, osana tutkimusta siitä, miten syntagmaattiset sulaumat esiintyvät ruututeksteissä. Kaikista sulaumamuodoista löytyi yhteensä 279 liitepartikkeliin päättyvää muotoa, joten vain murto-osa sulaumista oli liittynyt liitepartikkeliin. Lisäksi päätteitä löytyi ainoastaan sulaumapartikkelien *ettei*, *miksei*, *muttei* ja *vaikkei* tuloksista, ja niistäkin vajaa 98 prosenttia koostui *ettei*-sulauman esiintymistä. Lisäksi ehdottomasti yleisin pääte oli liitepartikkeli *-kö*, jonka osuus osakorpuksen materiaaleissa esiintyneistä liitepartikkeleista oli noin 97 prosenttia.

Aiemmin, luvussa 2 on jo selitetty, että syntagmaattisissa sulaumissa tapahtuu vokaalisoinnun muutoksia, kun takavokaaliseen konjunktioon liittyessään kieltoverbi voi ottaa joko

etuvokaalisen tai takavokaalisen muodon monikon 3. persoonan muodoissa (ISK 2005, 165). Takavokaalisen *-eivat*-päätteen voivat siis ottaa sulaumakonjunktioit *jollei, jottei, muttei* sekä *vaikkei*. Osakorpuksen materiaaleissa etuvokaaliset (*-eivät*) variantit olivat monikon 3. persoonan esiintymissä selkeästi suuremmassa asemassa kuin takavokaaliset. Etuvokaaliset variantit kattoivat 82 prosenttia kaikista monikon 3. persoonan esiintymistä. Eri sulaumakonjunktioiden välillä oli kuitenkin jonkin verran eroa vokaalisointujen muutosten osalta. Pienin ero etu- ja takavokaalisten varianttien esiintymämäärien välillä oli sulaumakonjunktio *jollei* tuloksissa, suurin ero taas *muttei*-sulaumamuodon varianttien välillä. Tarkemmat tiedot näiden varianttien käytöstä löytyvät edempää partikkelikohtaisista tuloksista. Vokaalisoinnun muutoksia tarkasteltiin myös kolmen korpuksen välisessä vertailussa (tulokset esillä alaluvussa 6.2.3).

***Sulaumat lyhyissä virkkeissä ja lauseissa.*** Kuten edeltävässä luvussa on selitetty, syntagmaattisten sulaumien tuloksista eriteltiin erilleen tapaukset, jotka esiintyvät ns. lyhyissä lauseissa, eli silloin, kun sulaumamuodot esiintyvät virkeissä tai lauseissa joko yksin tai korkeintaan yhden muun sanan kanssa (esimerkkinä kysymys ”***Mikset*** haluaisi?”). Lisäksi syntagmaattisista sulaumamuodoista tutkittiin lyhyiden lauseiden osalta erikseen, miten usein ne esiintyvät yksin virkkeissä tai lauseissa. Puhuessani tässä *virkkeistä*, tarkoitan täysin yksinään yhtenä lausekokonaisuutena esiintyviä sulaumamuotoja (esimerkiksi ”***Siksen.***”), kun taas *lauseista* puhuessani tarkoitan tapausta, jossa sulaumamuoto on pää- tai sivulauseessa joko yksin tai mahdollisesti jonkin toisen sanan kanssa (kuten ”Minähän sanoin, ***ettei*** kannata!” tai ”Tiedän, ***mutten...***”). Kuten jo edeltävässä luvussa on selitetty, näihin yksittäin esiintyvien tapausten joukkoon on laskettu myös jälkimmäisen esimerkin tapauksen, joissa lause katkeaa kolmeen pisteeseen.

Taulukossa 4 on esillä partikkelikohtaiset tulokset siitä, miten syntagmaattiset sulaumat esiintyvät miljoonaan sanaan suhteutettuna tässä aineistossa lyhyissä lauseissa: vasemmassa sarakkeessa näkyy sulaumamuotojen esiintyminen lyhyissä lauseissa, jotka sisälsivät korkeintaan yhden muun sanan syntagmaattisen sulauman lisäksi, keskimmaisessä sarakkeessa taas on nähtävissä syntagmaattisten sulaumamuotojen esiintyminen yksinään pää- tai sivulauseessa. Yksinään esiintyvät muodot sisältyvät lyhyiden lauseiden esiintymiin. Näitä suhteellisia frekvenssilukuja on käsitelty tarkemmin alaluvussa 6.1.3. Oikeanpuoleisesta sarakkeesta selviää, miten suuri osa lyhyiden lauseiden esiintymistä ilmeni yksin lauseessa.

Taulukossa 4 on esitetty tulokset partikkelikohtaisesti, tarkempi listaus eri taivutusmuotojen suhteellisista frekvenssiluvuista löytyy liitteestä V. Kaiken kaikkiaan kaikista syntagmaattisten sulaumien esiintymistä 13 prosenttia muodostui yksi tai kaksisanaisista virkkeistä tai lauseista. Yksittäisten sulaumien ja niiden taivutusmuotojen osalta tuloksissa oli kuitenkin suuria eroja toisiinsa nähden. *Siksei*-sulauman osalta yli 80 prosenttia esiintymistä löytyi yksi- tai kaksisanaisista lauseista, kun taas *miksei*- ja *muttei*-sulaumamuotojen kohdalla sama prosenttiluku oli 30. Harvinaisin kaiken kaikkiaan lyhyissä lauseissa esiintynyt sulaumakonjunktio oli *jottei* ja yleisin taas *miksei*.

Taulukko 4. Syntagmaattiset sulaumat lyhyissä lauseissa

	SS:t lyhyissä lauseissa (suht. F)	SS:t yksin lauseissa (suht. F)	SS:t yksin lauseissa (%:a lyhyistä lauseista)
ehkei	4,1	0,2	5 %
ellei	24,9	2,3	9 %
ettei	120,4	2,6	2 %
jollei	10	0,3	3 %
jottei	0,8	-	-
miksei	129,2	73,0	57 %
muttei	43,4	0,2	0 %
siksei	0,5	0,5	100 %
vaikkei	7,8	0,1	1 %
<b>kaikki</b>	<b>341,1</b>	<b>79,2</b>	<b>23 %</b>

Syntagmaattisten sulaumien osalta lyhyitä lauseita tarkasteltiin siis myös erikseen, milloin sulaumamuodot muodostivat yksin joko kokonaisen virkkeen tai pää- tai sivulauseen (partikkelikohtaiset tulokset miljoonaan sanaan suhteutettuna esillä taulukon 4 keskimmaisessä sarakkeessa). Yhteensä 23 prosenttia kaikista lyhyissä lauseissa löytyneistä syntagmaattisista sulaumista esiintyi yksinään, omana virkkeenään tai lauseenaan. Valtaosa lyhyissä lauseissa esiintyvistä sulaumamuodoista siis ilmeni kaksisanaisissa lauseissa. Yksinäisinä lauseina tai virkkeinä esiintyvien syntagmaattisten sulaumien tuloksista nousevat esiin etenkin *miksei*- ja *siksei*-sulaumapartikkelit, vaikka jälkimmäisen tulokset miljoonaan sanaan suhteutettuna eivät vaikutakaan merkittävilä. Siinä, missä muiden sulaumamuotojen kohdalla kaksisanaisten lauseiden osuus lyhyiden lauseiden esiintymistä oli 91–100 prosenttia, olikin *siksei*-sulaumapartikkelin kohdalla yksin esiintyvien muotojen osuus 100 prosenttia. *Miksei*-sulaumapartikkelin taivutusmuotojen kesken yksittäisinä lauseina

esiintyvien muotojen prosenttiosuus lyhyissä lauseissa esiintyvistä muodoista vaihteli yksikön 2. persoonan 43 prosentista yksikön 3. persoonan 74 prosenttiin. *Jottei*-sulaumakonjunktio taas erottuu muiden sulaumien tuloksista sillä, että sen esiintymissä ei ilmennyt yhtäkään yksittäin esiintyvää sulaumamuotoa.

Alla tuloksia on esitelty tarkemmin partikkelikohtaisesti eriteltynä. Jokaisesta sulaumapartikkelista on tutkittu samoja asioita, joskin jaotelluissa tuloksissa on myös tuotu esiin asioita, jotka ovat tulleet ilmi vain kyseisen sulaumamuodon ja sen taivutusmuotojen tuloksissa. Ellei persoonamuotoa ole tuloksissa erikseen mainittu, kyseisen sulauman tulokset koskevat kaikkia sen taivutusmuotoja.

***Ehkei.*** Sulaumapartikkeli *ehkei* esiintyi osakorpuksen aineistossa taivutusmuotoineen yli 800 kertaa, joskin yli 80 prosenttia näistä esiintymistä löytyi yksikön 3. persoonan muodossa. Muutoinkin yksikön muodot kattoivat lähes 94 prosenttia kaikista *ehkei*-partikkelin esiintymistä. Valtaosassa kaikista sen esiintymistä *ehkei* oli lauseenalkuisessa asemassa, eli se aloitti virkkeen tai lauseen, jossa se esiintyi, pääasiassa tässä aineistossa virkkeen – se siis oli yleensä virkkeen ensimmäinen sana. Mikäli *ehkei* ei aloittanut lausetta, sitä edelsi jokin vahvistussana (esim. *no*) tai rinnastuskonjunktio (esim. *ja*, *tai*, *mutta*), esimerkkinä lause ”Minut siis paikattiin nyt, mutta ***ehkei*** enää ensi kerralla”. Kiinnostavaa kyllä, tällaisissa tapauksissa kieltoverbi olisi myös voitu yhdistää konjunktioon *mutta*. Tämän sulaumapartikkelin yleisin kollokaatti, eli läheisin sana, oli *olla*-verbi preesensmuodossa. Huomion arvioista on se, että varsinkin yksikön 1. persoonan ja monikon 2. persoonan muotoja seurasi usein sana *enää*.

Lisäksi verbien päätteinä esiintyi *ehkei*-partikkelin sisältävissä lauseissa jonkin verran *-kAAn*-fokuspartikkelia. Yleisimmin *ehkei* esiintyi tässä materiaalissa ilman persoonapronominia, sillä vain 18 prosentissa esiintymistä seurasi persoonapronomini hakusanaa. Eniten persoonapronomineja löytyi monikon 3. persoonasta (46 %:ssa sen esiintymistä). Lyhyiden, yksi- tai kaksisanaisten lauseiden osalta 13 prosenttia kaikista *ehkei*-sulaumapartikkelin osumista esiintyi tällaisissa lyhyissä lauseissa. Näistä esiintymistä taas 95 prosenttia esiintyi kahden sanan lauseissa. Yksin yhden sanan virkkeinä esiintyviä sulaumamuotoja löytyi ainoastaan yksikön 3. persoonan muodoista.



**Ellei.** Sulaumakonjunktio *ellei* löytyi tutkitusta materiaalista taivutusmuotoineen yli 6 300 kertaa. Yksikön 3. persoonan muodot kattoivat yli 54 prosenttia *ellei*-partikkelin esiintymistä ja kaiken kaikkiaan yksikölliset muodot lähes 84 prosenttia kaikista sen osumista. Lähes kaikissa esiintymissään *ellei* ja sen taivutusmuodot esiintyivät lauseenalkuisessa asemassa. Vain kourallisessa löydöksistä *ellei* ilmeni lauseen toisena sanana, yleensä kysymyspartikkelin *entä* jälkeen. Myös tämän sulaumamuodon yleisin kollokaatti oli preesensmuotoinen *olla*-verbi, joskin myös konditionaalimuotoinen *olla*-verbi korostui jokaisessa persoonamuodossa – yksikön 1. persoonan muodoissa se oli jopa yleisin kollokaatti. Myös *haluta*-verbiä esiintyi *ellei*-sulauman kollokaattina suurin määrin.

Kymmenessä prosentissa kaikista *ellei*-konjunktin esiintymistä sulaumamuotoa seurasi persoonapronomini. Etenkin monikon 3. persoonan löydöksissä hakusanan oikealta puolelta löytyi persoonapronomini (22 %:ssa muodon esiintymistä). Kaikista *ellei*-sulauman osumista kymmenen prosenttia esiintyi lyhyissä, yhden tai kahden sanan lauseissa. Näistä esiintymistä 91 prosenttia esiintyi kahden sanan lauseissa. Kuitenkin yhden sanan virkkeinä esiintyviä sulaumia löytyi kaikista muista taivutusmuodoista, paitsi yksikön ja monikon 1. persoonista. Huomattavaa on, että yksikön 3. persoonan kohdalla valtaosa yksittäin esiintyvistä muodoista ilmeni vaillinaiseksi jäävänä lauseena, kuten esimerkissä ”Edistystä ei tapahdu, *ellei*...”

**Ettei.** Sulaumakonjunktio *ettei* ilmeni taivutusmuotoineen osakorpuksessa lähes 40 000 kertaa kattaen yli puolet kaikista syntagmaattisten sulaumien esiintymistä osakorpuksen tekstityksistä. Reilu 59 prosenttia *ettei*-konjunktin sulaumamuodoista esiintyi tässä materiaalissa yksikön 3. persoonassa, ja yhteensä 89 prosenttia näistä löydöksistä oli yksiköllisessä muodossa. *Ettei*-sulaumamuodot esiintyivät vain harvoin lauseessa toisena sanana ja tuolloinkin niitä useimmiten edelsi rinnastuskonjunktio *ja*. Selkeästi yleisimmät kollokaatit asemassa R1 olivat *olla*- ja *voida*-verbit. Liitepartikkelien lukumäärän osalta *ettei*-sulauman tuloksissa löytyi merkittävästi enemmän esiintymiä kuin minkään muun sulaumapartikkelin osumista. Niistä lähes 67 prosenttia ilmeni yksikön 3. persoonan esiintymissä, ja selkeästi yleisin päätte oli liitepartikkeli *-kö*.

Kahdeksassa prosentissa *ettei*-konjunktin tuloksista sulaumamuotoa seurasi persoonapronomini, eritoten monikon 3. persoonan löydöksissä (21 %:ssa sen osumista). Kaikista *ettei*-sulauman osumista kahdeksan prosenttia löytyi yhden tai kahden sanan lauseista. Näistä lyhyiden lauseiden esiintymistä 98 prosenttia ilmeni kaksisanaisissa

lauseissa. Tästä huolimatta yhden sanan lauseina toimivia sulaumamuotoja löytyi kaikista muista paitsi monikon 1. ja 2. persoonan muodoista, joskin valtaosa näistä yksittäin esiintyvistä muodoista edusti kesken katkeavaa sivulausetta, kuten esimerkiksi ”Sanoin jo, *etten...*”

**Jollei.** Sulaumakonjunktio *jollei* löytyi taivutusmuotoineen tutkimusaineistoista hieman yli 2 000 kertaa. Sen yleisin taivutusmuoto oli yksikön 3. persoona, joka sisälsi yli 48 prosenttia sulaumakonjunktio kaikkien esiintymistä. Kaiken kaikkiaan lähes 84 prosenttia sen esiintymistä löytyi yksiköllisessä muodossa. *Jollei*-sulaumamuodot esiintyivät yllättävän usein lauseensa toisena sanana, ja yleisimmin niitä edelsivät alistuskonjunktio *että* tai rinnastuskonjunktio *entä*, esimerkkinä lauseet ”Sanoimme, että *jollei* hän auta meitä...” ja ”Entä *jolleivat* he maksa?” Yleisin kollokaatti *jollei*-sulaumille oli konditionaalimuotoinen *olla*-verbi.

Lisäksi kymmenessä prosentissa *jollei*-konjunktio tuloksista ensimmäisenä hakusanan oikealla puolella oli persoonapronomini. Tämänkin sulaumakonjunktio kohdalla monikon 3. persoonan muodot erottuivat selkeästi muista persoonamuodoista sillä, että niiden rinnalla esiintyi useammin persoonapronomini (*jolleivat* 38 %, *jolleivät* 23 %). *Jollei*-sulauman kaikkien esiintymistä 13 prosenttia löytyi lyhyistä lauseista, joissa oli korkeintaan kaksi sanaa. Näistä löydöksistä 97 prosenttia esiintyi kahden sanan lauseissa, kun taas yhden sanan lauseina *jollei*-sulaumaa löytyi vain yksikön persoonista. Vokaalisoinnun muutosten osalta *jollei*-sulaumakonjunktio monikon 3. persoonan esiintymisissä löytyi vähiten eroa etu- ja takavokaalisten varianttien välillä muihin muutoksia ottaviin konjunktioihin verrattuna. Etuvokaalisia variantteja esiintyi 66 prosentissa *jollei*-sulauman monikon 3. persoonan muodoista.

**Jottei.** Sulaumakonjunktio *jottei* ja sen taivutusmuodoista tuli osakorpuksen materiaaleissa esiin 534 tapausta. Näistä yleisin muoto oli yksikön 3. persoona, jonka esiintymät kattoivat yli 72 prosenttia sulaumakonjunktio kaikkien löydöksistä. Lisäksi lähes 87 prosenttia kaikkien esiintymistä oli yksiköllisessä muodossa. Myös *jottei* ilmeni harvoin muussa kuin lauseenalkuisessa asemassa. Niissä muutamissa tapauksissa, jossa se oli lauseen toisena sanana, edelsi sitä yleisimmin rinnastuskonjunktio *ja*, joskin yksikön 3. persoonassa löytyi hakusanan edeltä jonkin verran myös pronominia *mikä* kysymyksessä ”Mikä *jottei*?” Tämän

sulauman kohdalla selkeää yleisintä kollokaattia ei ollut, vaan useimmiten sulaumamuotoa seurasi persoonapronomini.

*Jottei*-sulaumakonjunktion yhteistuloksissa yhdeksässä prosentissa esiintymistä löytyikin persoonapronomini suoraan hakusanan jäljestä. Lisäksi tämän sulaumamuodon osalta persoonapronomini löytyi vain yksikön ja monikon 3. persoonan muodoista, joskin näiden muotojen prosentiosuuksilla ei ollut keskenään suuria eroja. Kiinnostavaa kyllä, valtaosa hakusanaa seuraavista verbeistä oli konditionaalissa, kuten esimerkiksi ”Tulin itse, *jottet* järkyttyisi liikaa”. Lyhyiden lauseiden osalta *jottei*-sulaumamuoto oli kaikkein harvinaisin, vain neljä prosenttia sulauman kaikista osumista esiintyi yksi- tai kaksisanaisissa lauseissa. Lisäksi yhden sanan virkkeinä esiintyviä muotoja ei löytynyt lainkaan, eikä myöskään kahden sanan lauseissa esiintyviä muotoja monikon 2. persoonan muodoista tai monikon 3. persoonan takavokaalisista muodoista (*jotteivat*). Vaikka muutoin *jottei*-sulauman eri muodot esiintyivät harvoin lyhyissä lauseissa, ilmeni yksikön 2. persoonan muotoa *jottet* neljänneksessä kaikista sen löydöksistä kaksisanaisissa virkkeissä. Myös *jottei*-sulaumakonjunktio saattaa ottaa etu- tai takavokaalisen tunnuksen yhdistyessään kieltoverbiin monikon 3. persoonassa. Ero näiden kahden variantin välillä oli suuri, etuvokaaliset muodot kattoivat 78 prosenttia viimeisen taivutusmuodon löydöksistä.

***Miksei***. Sulaumapartikkeli *miksei* ilmeni taivutusmuotoineen tutkitussa materiaalissa yli 11 500 kertaa. Se olikin osakorpuksen toiseksi yleisin syntagmaattinen sulauma. Kaikista muista sulaumapartikkeleista poiketen *miksei*-sulauman yleisin persoonamuoto oli yksikön 2. persoona, jonka esiintymät kattoivat lähes 40 prosenttia sulaumamuodon kaikista löydöksistä. Lisäksi yksikölliset muodot veivät lähes 87 prosentin osuuden kaikista sen esiintymistä. Valtaosassa esiintymistään *miksei* taivutusmuotoineen aloitti virkkeen. Lisäksi se harvoin esiintyi lauseessa toisena sanana, ja silloinkin yleisimmin adverbien *niin* jälkeen. Yleisimmät kollokaatit *miksei*-sulaumamuodolle olivat presensmuotoinen *olla*-verbi sekä varsinkin yksikön sekä monikon 2. persoonassa perfektimuotoinen *kertoa*-verbi. Liitepartikkeli taas löytyi neljästä sulaumamuodon esiintymästä.

Lisäksi kymmenessä prosentissa *miksei*-partikkelin tuloksista hakusanaa seurasi persoonapronomini. Etenkin monikon 3. persoonan löydöksissä niiden osuus oli merkittävä (31 %:a sen esiintymistä). Lyhyiden lauseiden osalta *miksei*-sulauma taivutusmuotoinen on epätavallinen. Kaikista sen osumista 30 prosenttia löytyi yksi- tai kaksisanaisista lauseista.

Eritoten yksiköllisissä taivutusmuodoissa lyhyet lauseet olivat yleisiä, 30–37 %:a kaikista niiden esiintymistä taivutusmuodosta riippuen. *Miksei* on sikälikin harvinainen sulaumamuoto lyhyiden lauseiden osalta, että 57 prosenttia sen esiintymistä yhden tai kahden sanan lauseissa löytyikin yksisanaisina virkkeinä. Yksikön 3. persoonan muodossa jopa 74 prosenttia sen esiintymistä lyhyissä lauseissa löytyi yksisanaisina virkkeinä.

**Muttei.** Sulaumakonjunktio *muttei* löytyi eri muotoineen korpuksista miltei 4 000 kertaa, joista lähes 48 prosenttia oli yksikön 3. persoonassa ja yli 95 prosenttia kaiken kaikkiaan yksiköllisessä muodossa. Valtaosin *muttei* taivutusmuotoineen esiintyi virkkeen keskellä, aloittamassa sivulauseen. Toisin kuin monien muiden konjunktioden kohdalla, *muttei*-sulaumaa ei koskaan edeltänyt lauseessa toinen alistus- tai rinnastuskonjunktio eikä vahvistussana. Siinä missä *ehkei*-sulaumamuodon tuloksista löytyi tapauksia, joissa sulaumamuotoa edelsi rinnastuskonjunktio *mutta*, ilmeni myös *muttei*-sulauman esiintymissä tapauksia, joissa modaalipartikkeli *ehkä* seurasi *muttei*-sulaumakonjunktia, kuten esimerkissä ”Kyllä sinä ymmärrät, *muttei* ehkä muut”. Tällaisten tapausten kohdalla olisi aina ollut mahdollista yhdistää kieltoverbi myös toiseen lauseessa esiintyvään partikkeliin. *Muttei*-sulaumamuotojen selkeästi yleisin kollokaatti oli preesensmuotoinen *olla*-verbi, joskin *voida*- ja *tietää*-verbit esiintyivät myös usein preesensissä sulaumamuodon *muttei* rinnalla. Lisäksi *muttei*-sulauman yksikön 3. persoonan tuloksista löytyi yksi liitepartikkeliin päättyvä esiintymä.

Lisäksi, muihin sulaumapartikkeleihin verrattuna *muttei* oli melko poikkeuksellinen persoonapronominien osalta. Vain yhdessä prosentissa sen esiintymistä suoraan hakusanan oikealta puolelta löytyi persoonapronomini. Siinä missä muiden kohdalla monikon 3. persoonan muodoissa ilmeni eniten persoonapronomineja hakusanan vieressä, ilmeni niitä *muttei*-sulauman tuloksissa eniten yksikön 2. ja 3. persoonan muodoissa (2 %:a tuloksista molempien osalta). Lisäksi monikon 2. persoonan sekä 3. persoonan etuvokaalisten muotojen läheisyydestä ei löytynyt lainkaan persoonapronomineja. Niin ikään lyhyiden lauseiden tuloksia tarkasteltaessa *muttei*-sulauma eri taivutusmuotoineen on poikkeuksellinen tapaus. Kuten *miksei*, myös *muttei*-sulauman kaikista osumista 30 prosenttia esiintyi yksi- tai kaksisanaisissa lauseissa. Eri taivutusmuotojen prosenttiosuudet kuitenkin vaihtelevat hyvin tasaisesti, sillä ne muodostavat 21–32 prosenttia niiden kaikista esiintymistä taivutusmuodosta riippuen. Yksinäisinä virkkeinä *muttei* sen sijaan esiintyi vain murto-osassa näistä lyhyiden

lauseiden esiintymistä, joista yhtä poikkeusta lukuun ottamatta kaikki olivat kesken katkeavia sivulauseita. Lisäksi sulaumakonjunktion tuloksissa monikon 3. persoonan etu- ja takavokaalisten varianttien ero oli merkittävin verrattuna muihin konjunktioihin, jotka voivat ottaa takavokaalisen *-eivat*-tunnuksen. Etuvokaaliset variantit kattavat 92 prosenttia *muttei*-sulaumamuodon monikon 3. persoonan esiintymistä.

**Siksei.** Sulaumapartikkeli *siksei* sai ainoastaan 17 osumaa tutkitusta materiaalista ollen selkeästi harvinaisin tutkituista sulaumamuodoista. Osumia löytyi vain kolmessa persoonassa, niistäkin valtaosa (76 %:a) oli yksikön 3. persoonassa, kun taas monikollisia muotoja löytyi vain yksi monikon 3. persoonasta. *Siksei*-sulaumamuodon harvoista esiintymistä löytyi hakusanan vierestä persoonapronomini vain kuudessa prosentissa kaikista tuloksista. Kuitenkin persoonapronomineja ilmeni ainoastaan yksikön 3. persoonan muodoissa (6 %:ssa sen esiintymistä). Lisäksi 82 prosenttia *siksei*-sanankäytön taivutusmuotojen löydöksistä esiintyi ainoastaan yksisanaisina päälauseina, yksikön 1. persoonan muodon *siksen* esiintymistä jopa kaikki olivat yksisanaisia virkkeitä – valtaosa vastauksena kysymykseen ”mikset”. Kahden sanan lauseissa *siksei* sen sijaan ei esiintynyt lainkaan.

**Vaikkei.** Sulaumakonjunktio *vaikkei* ilmeni tutkimusaineistossa vajaa 2 000 kertaa. Lähes 63 prosenttia näistä esiintymistä löytyi yksikön 3. persoonan osumista, ja kaiken kaikkiaan noin 93 prosenttia *vaikkei*-sulaumamuodon esiintymistä esiintyi yksiköllisessä muodossa. Yleisimmin *vaikkei* oli taivutusmuotoineen aloittamassa sivulauseita, mutta harvoissa tapauksissa sitä saattoi myös edeltää joko alistuskonjunktio *että* tai rinnastuskonjunktio *ja*. Yleisin kollokaatti *vaikkei*-sulaumalle oli preesensmuotoinen *olla*-verbi. Huomionarvoista on kuitenkin, että myös fokuspartikkelillinen *olla*-verbin muoto *olekaan* löytyi etenkin yksikön sekä monikon 1. persoonien yleisimpien kollokaattien joukosta. Kaiken kaikkiaan muutenkin *vaikkei*-sulaumamuotojen esiintymisissä verbit olivat monesti joko *-kAAn*-päätteisiä tai konditionaalissa. Liitepartikkeliin *vaikkei*-sulauma päättyi tässä materiaalissa vain kerran.

Lisäksi, *vaikkei*-konjunktion yhteistuloksissa persoonapronomineja edeltäviä hakuosumia löytyi kuudesta prosentista. Selvästi eniten persoonapronomineja esiintyi monikon 3. persoonan muotojen rinnalla (*vaikkeivat* 14 %:a, *vaikkeivät* 18 %:a). Lyhyiden lauseiden osalta *vaikkei*-sulauma oli melko harvinaisen, vain kymmenen prosenttia sen kaikista esiintymistä löytyi yhden tai kahden sanan lauseista. Näistäkin valtaosa (99 prosenttia) esiintyi kaksisanaisissa lauseissa. Lisäksi monikon 1. persoonan muotoa *vaikkemme* ei

löytynyt lainkaan lyhyistä lauseista. Kuitenkin siitä huolimatta, että sulaumamuoto *vaikka* oli kaiken kaikkiaan melko harvinainen lyhyissä lauseissa, esiintyi 31 prosenttia sulauman monikon 2. persoonan muodoista juurikin lyhyissä lauseissa. Vokaalisoinnun muutosten osalta *vaikkei*-sulaumamuodon tuloksissa oli selkeä ero varianttien välillä. Etuvokaaliset variantit kattoivat 89 prosenttia *vaikkei*-sulauman monikon 3. persoonan esiintymistä.

### 6.1.2 Erilleen kirjoitetut muodot

Jotta osakorpuksessa esiintyviä syntagmaattisia sulaumamuotoja voitaisiin arvioida, piti niiden verrokiksi etsiä korpuksista myös samojen kielellisten elementtien erilleen kirjoitetut muodot kaikkine taivutusmuotoineen. Kuten edeltävässä luvussa on täsmennetty, tähän vertailuun otettiin mukaan kaikki tapaukset, jotka olisivat yhtä hyvin voineet alkujaan olla ruututeksteissä myös lyhempinä, sulautuneina muotoina. Tästä tutkimuksesta pois jätetyistä tapauksista on kerrottu tarkemmin alaluvussa 5.3.2. Yhteensä tähän tutkimukseen hyväksytyjä erilleen kirjoitettuja tapauksia löytyi osakorpuksista 62 931, joka on noin 0,24 prosenttia koko korpuksen sanamäärästä. Eniten löytyi yhdistelmiä *mutta + ei*, *että + ei* sekä *jos + ei*, kun taas harvinaisimpia olivat *siksi + ei* sekä *jotta + ei*, kuten on nähtävissä taulukosta 5. Kaiken kaikkiaan osakorpuksen aineistossa erilleen kirjoitettujen muotojen osalta yksikölliset persoonamuodot olivat selvästi yleisempiä kuin monikolliset muodot. Niiden osuus kaikista esiintymistä oli vajaa 84 prosenttia. Yleisin persoonamuoto oli yksikön 3. persoona, joka kattoi noin 58 prosenttia kaikista erilleen kirjoitettujen muotojen esiintymistä.

Taulukko 5. Osakorpuksen erilleen kirjoitetut sulaumakonjunktiot ja -partikkelit

MUODOT	OK
ehkä + ei	4 600
että + ei	7 904
jos + ei	12 970
jotta + ei	953
miksi + ei	3 801
mutta + ei	29 825
siksi + ei	879
vaikka + ei	1 999
<b>yhteensä</b>	<b>62 931</b>

Erilleen kahtena sanana kirjoitetut muodot ovat tässä tutkimuksessa siis vain vertailukohde varsinaiselle tutkimuskohteelle, eli syntagmaattisille sulaumille. Tästä syystä näitä muotoja ei lukumäärien ja frekvenssien laskemista lukuun ottamatta tutkittu erityisen tarkasti. Poistettavia osumia tehdessä kuitenkin tuli esiin, että *ehkä*-partikkeli on muihin tutkittuihin sulaumakonjunktioihin ja -partikkeleihin nähden poikkeuksellinen sikäli, että se voi esiintyä lauseissa kieltoverbin *ei* jälkeen, kuten lauseessa ”Hän *ei ehkä* pääsisi tänne”. Kauimmaisimmillaan tässä aineistossa kontekstisanana kieltoverbi *ei* esiintyi asemassa L3, eli kolme sanaa hakusanansa edellä, kuten esimerkiksi lauseessa ”*Ei* se ole *ehkä* niin yksinkertaista” – joskin tällaisia tapauksia löytyi materiaalista vain muutama. Yleisimmin kieltoverbin ja *ehkä*-partikkelin välillä oli vain yksi sana, useimmiten subjekti tai predikaatti. Kaiken kaikkiaan 60 prosentissa aineistosta löytyneistä *ehkä*-partikkelin tapauksista kieltoverbi *ei* esiintyi ennen partikkelia *ehkä*. Eri persoonamuotojen prosenttiluvut vaihtelivat 58 prosentista 67 prosenttiin; pienin lukema oli yksikön 3. persoonassa ja suurin monikon 3. persoonassa.

***Erilleen kirjoitetut muodot lyhyissä virkkeissä ja lauseissa.*** Vaikka erikseen kahtena sanana kirjoitettavia muotoja ei tarkasteltu tässä tutkimuksessa erityisen yksityiskohtaisesti, tutkittiin niistä vertailun vuoksi lyhyiden virkkeiden esiintymistä, eli lauseita, joissa oli kahdesta kolmeen sanaa, siis korkeintaan yksi sana hakusanan ja kontekstisanana toimineen persoonamuotoisen kieltoverbin lisäksi. Lyhyitä virkkeitä ja lauseita tarkasteltiin, jotta niiden esiintymiä voitaisiin verrata niihin sulaumamuotoihin, jotka ilmenivät osakorpuksen lauseissa korkeintaan yhden toisen sanan seurana. Kaiken kaikkiaan 13 prosenttia kaikista erilleen kahtena sanana kirjoitettujen muotojen esiintymistä löytyi kaksi- tai kolmisanaisista lyhyistä lauseista, joskin erot eri partikkelien välillä olivat selkeät. Kaikkein eniten esiintymiä löytyi yhdistelmille *miksi + ei* ja *ehkä + ei* taivutusmuotoineen, vähiten taas yhdistelmille *siksi + ei* ja *vaikka + ei*. Lisäksi yksittäisten kielellisten elementtien sisällä oli merkittävääkin vaihtelua eri taivutusmuotojen kesken. Suurin eroavaisuus löytyi *ehkä*-partikkelin esiintymistä. Kaiken kaikkiaan 21 prosenttia *ehkä*-partikkelin löydöksistä esiintyi kaksi- tai kolmisanaisissa lauseissa, mutta yksikön 1. persoonan muodoista 85 prosenttia tuli esiin lyhyissä lauseissa, kun taas monikon 3. persoonan muodoista kuusi prosenttia esiintyi kaksi- tai kolmisanaisissa lauseissa. Taulukossa 6 on esillä partikkelikohtaiset tulokset siitä, miten erilleen kirjoitetut muodot esiintyivät miljoonaan sanaan suhteutettuna tässä aineistossa lyhyissä lauseissa: vasemmassa sarakkeessa näkyy erilleen kirjoitettujen muotojen esiintyminen lyhyissä

lauseissa, jotka sisälsivät siis korkeintaan yhden muun sanan partikkelin ja kieltoverbin lisäksi, keskimmaisessä sarakkeessa taas on esillä kahtena sanana muotojen esiintyminen yksinään pää- tai sivulauseessa. Jälkimmäiset tapaukset sisältyvät lyhyiden lauseiden tuloksiin. Näitä suhteellisia frekvenssilukuja on käsitelty tarkemmin alaluvussa 6.1.3. Taulukon 6 oikeanpuoleisesta sarakkeesta sen sijaan selviää, miten suuri osa lyhyissä lauseissa ilmenneistä erilleen kirjoitetuista muodoista esiintyi kaksisanaisissa lauseissa ja virkkeissä.

Taulukko 6. Erilleen kirjoitetut muodot lyhyissä lauseissa

	EKM:t lyhyissä lauseissa (suht. F)	EKM:t 2-san. lauseissa (suht. F)	EKM:t yksin lauseissa (%:a lyhyistä lauseista)
ehkä + ei	36,7	1,4	23 %
jos + ei	49,3	1,9	10 %
että + ei	14,7	0,6	25 %
jotta + ei	1,4	0,1	0 %
miksi + ei	48,9	1,8	71 %
mutta + ei	152,3	5,8	4 %
siksi + ei	1,0	0,0	7 %
vaikka + ei	4,8	0,2	1 %
<b>kaikki</b>	<b>309,1</b>	<b>11,7</b>	<b>19 %</b>

Koska syntagmaattisten sulaumien kohdalla lyhyistä lauseista tarkasteltiin, milloin sulaumamuodot esiintyvät yksin lauseessa tai virkkeessä, tutkittiin erilleen kirjoitetuista muodoista myös tapauksia, joissa nämä kielelliset elementit esiintyvät yksinään – eli tapauksia, joissa esiintyi vain hakusana ja kieltoverbi ja jotka olisi yhtä hyvin voinut muuttaa syntagmaattiseksi sulaumaksi. Taulukon 6 oikeanpuoleisesta sarakkeesta löytyy niiden partikkelikohtaiset prosenttiosuudet kaikista lyhyissä, kaksi- tai kolmisanasissa lauseissa ilmenneistä erilleen kirjoitetuista muodoista. Kaiken kaikkiaan tutkituista lyhyiden lauseiden esiintymistä 19 prosenttia ilmeni kaksisanaisina lauseina. Tältä osin eri partikkelien välillä oli varsin suuria eroja. Eniten kaksisanaisia virkkeitä ja lauseita löytyi konjunktion *miksi* tuloksista, joissa 71 prosenttia lyhyistä lauseista löytyi kaksisanaisista virkkeistä ja lauseista. *Miksi*-konjunktion tuloksissa korostuu etenkin yksikön 3. persoona, jonka esiintymistä 89 prosenttia ilmeni kaksisanaisissa lauseissa. Toiseksi eniten kaksisanaisia lauseita löytyi konjunktion *että* tuloksista, jonka osumista 25 prosenttia esiintyi vain hakusanan ja kieltoverbin sisältävissä lauseissa. *Että*-konjunktion taivutusmuodoista suurimman luokan muodostavat monikon 3. persoonan esiintymät, joista puolet löytyi kaksisanaisista lauseista.



*Jotta*-konjunktion tuloksista näitä kaksisanaisia lauseita ei löytynyt lainkaan ja *vaikka*-konjunktion tuloksistakin löytyi vain yksi esiintymä yksikön 2. persoonasta.

### 6.1.3 Osakorpuksen tulosten vertailu

Kaiken kaikkiaan siis erilleen kirjoitettuja löytyi osakorpuksesta reilut 4 000 tapausta vähemmän kuin samojen partikkelien sulautuneita muotoja, eli ruututekstityksissä ainakin tämän aineiston mukaan käytetään enemmän syntagmaattisia sulaumia kuin erilleen kirjoitettuja muotoja. Tämä tulos tukee myös osaltaan ISK:n (2005, 165) tietoa siitä, että partikkelien ja kieltoverbin syntagmaattiset sulaumat ovat yleiskielessä yleisempiä kuin konjunktioiden ja kieltoverbin erilleen kahtena sanana kirjoitetut muodot. Erilleen kirjoitettuja muotoja on käytetty ruututekstityksissä kuitenkin eri määrissä kuin sulaumia, esimerkiksi *mutta*-konjunktio löytyi osakorpuksen materiaaleissa yli seitsemän kertaa enemmän kirjoitettuna erilleen kieltoverbistä kuin yhteensulautuneena muotona. Kaiken kaikkiaan erilleen kirjoitettuna löytyi eniten partikkeleiden *mutta*, *että* sekä *jos* taivutusmuotoja ja harvinaisimpia olivat *siksi*- ja *jotta*-partikkeleiden taivutusmuodot, kuten näkyy taulukosta 6. Tässä kohdin on huomion arvoista se, että erilleen kirjoitetuissa muodoissa konjunktio *jos* kattaa syntagmaattisten sulaumien muotojen *ellei* ja *jollei* vastineet. Syntagmaattisia sulaumia taas löytyi eniten sulaumapartikkeleista *ettei*, *miksei* ja *muttei*, harvalukuisimpia taas olivat *siksei* sekä *jottei*. Harvinaisimmat partikkelit ovat siis sekä syntagmaattisten sulaumien että erilleen kirjoitettujen muotojen osalta samat.

Koska tässä vaiheessa tutkimusta tarkasteltiin ainoastaan osakorpuksen sisäistä vertailuasetelmaa, eli syntagmaattisten sulaumien esiintymien vertautumista samojen partikkeleiden erilleen kirjoitettujen muotojen lukumääriin, olisi vertailu näiden kahden välillä voitu suorittaa myös pelkistä esiintymismääristä (poistaen syntagmaattisista sulaumista liitepartikkeliin päättyvät muodot) – onhan ne kerätty samasta korpuksesta. Kuitenkin koska kolmen erikokoisen korpuksen keskinäinen vertailu syntagmaattisten sulaumien osalta vaati suhteellisten frekvenssilukujen laskemisen, haluttiin tutkimuksen yhdenmukaisuuden vuoksi myös osakorpuksen sisäinen vertailu suorittaa suhteellisista frekvenssiluvuista. Näitä lukuja tarkastelemalla oli siis mahdollista vertailla, montako kertaa mikäkin kielellinen elementti esiintyi aineistossaan miljoonan sanan joukkiossa. Suhteellisten frekvenssilukujen laskentakaava on aiemmin jo esitelty alaluvussa 5.3.3.

Taulukko 7. Osakorpuksen suhteellisten frekvenssien vertailu

ehkei	30,3	ehkä + ei	173,8
ellei & jollei	316,2	jos + ei	490,1
ettei	1 498,3	että + ei	298,7
jottei	20,2	jotta + ei	36,0
miksei	436,6	miksi + ei	143,6
muttei	146,8	mutta + ei	1 127,0
siksei	0,6	siksi + ei	33,2
vaikkei	75,5	vaikka + ei	75,5
<b>kaikki</b>	<b>2 524,5</b>	<b>kaikki</b>	<b>2 378,0</b>

Kuten edellä olevasta taulukosta 7 näkyy, syntagmaattisten sulaumien ja erilleen kahtena sanana kirjoitettujen muotojen suhteellisissa frekvenssiluvuissa miljoonaanunaan suhteutettuna on vain vajaa 150 esiintymän ero sulaumamuotojen eduksi. Vaikka kaikki taivutusmuodot yhteenlaskettuna eroavaisuus on koko korpuksen kokoon nähden melko pieni, ovat erot eri sulaumamuotojen ja niiden erilleen kirjoitettujen vastineiden välillä melko suuret. Oheisessa taulukossa 7 on syntagmaattisten sulaumien kohdalla laskettu sulaumien *ellei* ja *jollei* suhteelliset frekvenssiluvut yhteen, sillä molemmat muodot vastaavat erilleen kirjoitettua muotoa *jos + ei*. Pienin ero syntagmaattisen sulauman ja erilleen kirjoitetun muodon välillä on *vaikkei*-sulaumakonjunktion kohdalla (vain yksi esiintymä), kun taas *jos + ei* ja *jotta + ei* taivutusmuotoineen esiintyvät osakorpuksessa lähes kaksi kertaa enemmän kuin niiden yhteensulautuneet vastineet. *Miksei*, *ettei* ja *ehkei* sen sijaan esiintyvät taivutusmuotoineen ruututeksteissä noin 3–6 kertaa useammin kuin niiden erilleen kirjoitetut muodot, kun taas *muttei* on lähes kahdeksan kertaa harvinaisempi kuin sen erilleen kirjoitettu vastine. *Siksei*-sulaumapartikkelin kohdalla ero on merkittävin: sen erilleen kirjoitetut muodot esiintyvät ruututeksteissä lähes 52 kertaa useammin kuin sen sulautuneet muodot. Kaiken kaikkiaan siis erilleen kirjoitetut muodot ovat muiden kuin *ettei*-, *miksei*- ja *vaikkei*-sulaumien osalta yleisempiä ruututeksteissä kuin niiden sulautuneet vastineet, joskin *vaikkei*-sulauman kohdalla eroa on vain yksi ainoa esiintymä. *Ettei*- ja *miksei*-sulaumien suurehko ero niiden erilleen kahtena sanana kirjoitettuihin muotoihin nostaa yhteistuloksissa syntagmaattisten sulaumien esiintymät yleisemmäksi kuin sulautumattomat muodot.

**Lyhyiden lauseiden vertailu.** Kuten tässä tutkimuksessa on jo aiemmin mainittu, kun aineistojen KWIC-rivejä karsittiin, jäi kyseenalaiseksi, voisiko lyhyissä parin sanan virkkeissä ja lauseissa olevat partikkelien ja kieltoverbin yhteisesiintymät muuttaa syntagmaattisiksi

sulaumiksi. Tästä syystä näitä lyhyitä lauseita haluttiin tarkastella vielä erikseen. Koska ajatuksena oli, että erilleen kirjoitetut muodot voitaisiin vaihtaa lauseissa vastaaviin syntagmaattisiin sulaumiin, kelpuutettiin ns. lyhyiksi lauseiksi syntagmaattisten sulaumien osalta virkkeet ja lauseet, jotka sisälsivät korkeintaan yhden sanan sulaumamuodon lisäksi, ja erilleen kirjoitettujen muotojen osalta virkkeet ja lauseet, jotka pitivät sisällään korkeintaan yhden muun sanan tutkittujen partikkeleiden ja kieltoverbin lisäksi. Vertailun ideana oli tarkastella, löytyykö sulautuneiden ja erilleen kirjoitettujen muotojen väliltä selkeitä eroja lyhyiden lauseiden osalta, eli onko näitä muotoja käytetty tekstityksissä eri tavoin. Taulukossa 8 on esillä eri kielellisten elementtien suhteelliset frekvenssit lyhyiden lauseiden osalta, joskin tässä taulukossa *ellei-* ja *jollei-*sulaumien tulokset on summattu yhteen, sillä ne molemmat vastaisivat aukikirjoitettuina sanaparia *jos ei*.

Taulukko 8. Lyhyiden lauseiden suhteelliset frekvenssit

ehkei	4,1	ehkä + ei	36,7
ellei & jollei	34,9	jos + ei	49,3
ettei	120,4	että + ei	14,7
jottei	0,8	jotta + ei	1,4
miksei	129,2	miksi + ei	48,9
muttei	43,4	mutta + ei	152,3
siksei	0,5	siksi + ei	1,0
vaikkei	7,8	vaikka + ei	4,8
<b>kaikki</b>	<b>341,1</b>	<b>kaikki</b>	<b>309,1</b>

Kaikkien partikkelien yhteistuloksissa ruututekstien lyhyissä lauseissa esiintyi jonkin verran enemmän syntagmaattisia sulaumia kuin erilleen kahtena sanana kirjoitettuja muotoja. Kaiken kaikkiaan osakorpuksen aineistosta löytyi lyhyistä lauseista 9 026 syntagmaattista sulaumaa ja 8 181 niiden erilleen kirjoitettuja muotoja. Vaikka kokonaisuutena näiden kahden variantin ero lyhyiden lauseiden osalta ei ollut suuri, löytyi yksittäisten partikkelien tuloksista merkittäviäkin eroja. Suurin eroavaisuus ilmeni *mutta*-konjunktion esiintymissä, joissa sulaumamuotojen ja erilleen kirjoitettujen muotojen välillä oli merkittävä ero jälkimmäisen hyväksi. *Mutta*-konjunktion taivutusmuotojen kesken huomattavin ero oli yksikön 3. persoonan esiintymissä (suhteelliset frekvenssit: *muttei* 22,4/*mutta + ei* 75,2). Myös *ehkä*-partikkeliä ilmeni lyhyissä lauseissa selvästi enemmän erilleen kirjoitettujen muotojen esiintymissä; partikkelin sisällä taas merkittävin ero oli yksikön 1. persoonassa (*ehken* 0,5/*ehkä + ei* 19,6). Sulaumamuodoissa taas suhteellisesti huomattavasti yleisempiä erilleen kirjoitettuihin muotoihin verrattuna olivat *ettei*-sulaumapartikkelin taivutusmuodot, etenkin

yksikön 1. persoonan osalta (*etten* 41,0/*että* + *en* 3,7), mutta myös yksikön 3. persoonan muodoissa oli selkeä ero. Lisäksi myös *miksi*-partikkeli löytyi materiaalista selvästi useammin sulautuneena muotoja kuin erilleen kieltoverbistä kirjoitettuna. Etenkin yksikön 2. ja 3. persoonan muodoissa oli havaittavissa kunnan ero erilleen kirjoitettuihin muotoihin (*mikset* 51,9/*miksei* 51,7/*miksi* + *et* 15,0/*miksi* + *ei* 19,5).

Lyhyitä lauseita tarkasteltaessa haluttiin tutkia myös yksittäin omana virkkeenään tai lauseenaan esiintyviä tapauksia, esimerkkinä ”*Ellei*, he lähtevät muualle”. Täten vertailuna erilleen kirjoitettujen muotojen osalta tarkasteltiin lauseita, jotka sisälsivät ainoastaan hakusanan ja kontekstisanana toimineen persoonamuotoisen kieltoverbin ottamatta huomioon välimerkkejä kesken katkeavien lauseiden osalta, kuten lauseessa ”Olet nätti jo nyt, *vaikka et...*” Näitä sulaumamuotojen osalta yksisanaisia ja erilleen kirjoitettujen muotojen osalta kaksisanaisia lauseita verrattiin toisiinsa suhteellisten frekvenssilukujen avulla. Tulokset ovat partikkelikohtaisesti esillä taulukossa 9, joskin myös tässä taulukossa *ellei*- ja *jollei*-sulaumien tulokset on laskettu yhteen.

Taulukko 9. Yksinään lauseessa esiintyvät muodot

ehkei	0,2	ehkä + ei	1,4
ellei & jollei	2,6	jos + ei	1,9
ettei	2,6	että + ei	0,6
jottei	-	jotta + ei	0,1
miksei	73,0	miksi + ei	1,8
muttei	0,2	mutta + ei	5,8
siksei	0,5	siksi + ei	0,0
vaikkei	0,1	vaikka + ei	0,2
<b>kaikki</b>	<b>79,2</b>	<b>kaikki</b>	<b>11,7</b>

Kokonaisuutena osakorpuksen materiaaleissa esiintyi lauseessa yksinään selkeästi enemmän syntagmaattisia sulaumia kuin erilleen kahtena sanana kirjoitettuja muotoja, kuten on nähtävissä taulukosta 9. Kaiken kaikkiaan osakorpuksesta löytyi 2 095 tapausta, joissa syntagmaattinen sulauma esiintyi yksin lauseessa tai virkkeessä, ja 1 543 tapausta, joissa erilleen kirjoitettu partikkeli ja kieltoverbi ilmenivät lauseen ainoina sanoina. Merkittävin ero ilmeni *miksei*-sulaumapartikkelin ja sen erilleen kirjoitettujen muotojen kesken, ja sulaumamuodon sisällä etenkin yksikön 2. ja 3. persoonan muodot erosivat erilleen kirjoitetuista varianteistaan (suhteelliset frekvenssit: *mikset* 22,4/*miksei* 38,3/*miksi* + *et* 0,6/*miksi* + *ei* 0,7). *Miksei*-sulaumakonjunktion osuus syntagmaattisten sulaumien yksittäin

esiintyneistä oli sen verran huomattava, että vaikka kokonaisuutena syntagmaattiset sulaumat ilmenivät ruututeksteissä selvästi useammin yksittäin lauseessa kuin partikkeli ja kieltoverbi ilman muuta kontekstia, esiintyi osakorpuksen materiaaleissa itse asiassa enemmän erilleen kirjoitettuja muotoja kuin sulaumamuotoja partikkeleiden *ehkä*, *jotta*, *mutta* ja *vaikka* osalta. Edellä mainittujen partikkelien varianttien erot olivat kuitenkin melko pieniä – joskin *jottei*-sulaumamuotoa ei löytynyt aineistosta lainkaan yksinäisenä lauseenjäsenenä.

## 6.2 Kolmen korpuksen vertailun tulokset

### 6.2.1 Syntagmaattiset sulaumat korpuksissa

Kun osakorpusta sekä Käännössuomen korpuksen sisältämiä alkuperäissuomen sekä käännössuomen sivukorpuksia lähdettiin tarkastelemaan, niistä kaikista tutkittiin ensimmäisenä niissä esiintyvät syntagmaattiset sulaumat, joiden lukumäärät laskettiin ja kirjattiin ylös. Kuten jo aiemmin on mainittu, kolmea korpusta vertaillaessa keskityttiin vain syntagmaattisiin sulaumiin, eikä niiden kahtena erillisenä sanana kirjoitettuihin muotoihin tutkimuksen vaatiman työ määrän tähden, ja sen takia, että sulaumamuodot haluttiin pitää tutkimuksen pääkohteena myös normivertailun tähden. Vertailussa on käytetty termiä *käännetty suomi* viittaamaan sekä ruututeksteihin että muuhun käännettyyn suomeen, eli toiseen osaan Käännössuomen korpuksesta, ja termiä *käännössuomi* viittaamaan ainoastaan muuhun käännettyyn suomeen.

Taulukko 10. Syntagmaattiset sulaumat korpuksissa

	OK	APS	KS
ehkei	802	14	90
ellei	6 347	633	1417
ettei	39 923	5192	8273
jollei	2 020	64	315
jottei	534	28	114
miksei	11 558	298	516
muttei	3 886	147	486
siksei	17	-	-
vaikkei	1 998	179	332
<b>kaikki</b>	<b>67 085</b>	<b>6 554</b>	<b>11 543</b>

Sulaumien yhteistiedot näkyvät edellä taulukossa 10, tarkemmat tiedot jokaisesta syntagmaattisesta sulaumasta taivutusmuotoineen löytyy liitteestä I. Korpusten yleisimmät

syntagmaattisten sulaumien taivutusmuodot ovat kuitenkin näkyvillä jäljempänä taulukossa 11. Kaikissa kolmessa korpuksessa yleisin taivutusmuoto oli yksikön 3. persoonassa oleva *ettei*. Tämäkin tulos tukee ISK:n ilmoittamaa tietoa siitä, että *ettei* on yleiskielessä yleisin sulaumakonjunktio (2005, 165). Kaiken kaikkiaan harvinaisin syntagmaattinen sulauma taivutusmuotoineen oli sulaumapartikkeli *siksei*, jota ei löytynyt lainkaan Käännösuomen korpuksen materiaaleista.

Taulukko 11. Korpusten yleisimmät sulaumat suhteellisten frekvenssien mukaan

OK		APS		KS	
ettei	888,7	ettei	1 033,9	ettei	1 017,8
etten	280,8	etten	154,0	etten	174,4
mikset	173,3	ellei	126,6	ellei	163,7

Koska osakorpus sisältää yli 26 miljoonaa sanaa ja Käännösuomen korpuksista alkuperäissuomen sivukorpus vajaa neljä miljoonaa sanaa ja käännösuomen sivukorpus vajaa kuusi miljoonaa sanaa, ei varsinaisten esiintymien määrästä vielä voi kovin paljoa päätelmiä tehdä. Jotta luvut saataisiin vertautumaan toisiinsa, laskettiin kaikille syntagmaattisten sulaumien esiintymille suhteelliset frekvenssiluvut. Luvut on suhteutettu siihen, miten monta kertaa mikäkin kielellinen elementti esiintyi miljoonan sanan joukkiossa (laskukaava esitelty alaluvussa 5.3.3). Frekvenssejä laskettaessa esiintymismäärästä on jätetty pois päätteelliset muodot. Edempää taulukosta 12 on nähtävissä sulaumapartikkelikohtaiset frekvenssiluvut, tarkempi listaus eri taivutusmuotojen suhteellisista frekvenssiluvuista löytyy liitteestä II.

Taulukko 3. Syntagmaattisten sulaumien frekvenssit

	OK	APS	KS
ehkei	30,3	3,4	15,3
ellei	239,8	165,0	240,8
ettei	1 489,3	1 329,9	1 384,4
jollei	76,3	16,7	53,4
jottei	20,2	7,3	19,4
miksei	436,7	77,0	87,4
muttei	146,8	38,4	82,6
siksei	0,6	-	-
vaikkei	75,5	46,7	56,4
<b>kaikki muodot</b>	<b>2 524,5</b>	<b>1 684,5</b>	<b>1 939,7</b>

Kaiken kaikkiaan ruututeksteissä oli näissä materiaaleissa käytetty selkeästi enemmän syntagmaattisia sulaumia kuin alkuperäissuomessa ja käännösuomessa ja täten kaiken

kaikkiaan käännetyssä suomessa enemmän kuin supisuomessa. Kaikkien sulaumapartikkelien osalta käännetyssä suomessa esiintyi enemmän sulaumamuotoja kuin alkuperäissuomessa. Käännetyn suomen varianttien sisällä ainoastaan *ellei*-sulaumakonjunktion muotoja löytyi käännoosuomesta suhteessa enemmän kuin ruututeksteistä, joskin ero oli hienoinen. Muutoin ruututeksteissä ilmeni aina enemmän syntagmaattisia sulaumia kuin muissa tutkituissa suomen varianteissa. Kokonaisuutena alkuperäissuomen ja käännoosuomen keskinäinen ero oli pienempi kuin käännoosuomen ero ruututeksteihin. Vain kahden sulaumamuodon kokonaistuloksissa käännoosuomen esiintymien tulos oli suhteellisesti lähempänä ruututekstitysten sulaumien määrää kuin alkuperäissuomen, sulaumakonjunktiolla *jollei* sekä *ellei*, joskin jälkimmäisessä esiintymiä löytyi enemmän käännoosuomesta kuin ruututeksteistä.

Sulaumapartikkelien kokonaistilanteeseen verrattuna yksittäisten sulaumapartikkelien kohdalla on tuloksissa kuitenkin enemmän vaihtelevuutta korpuksen kesken. Yleisin syntagmaattinen sulauma kaikissa korpuksissa siis oli *ettei*-sulaumakonjunktio taivutusmuotoineen. *Ettei*-sulauman osalta kolmen korpuksen tuloksissa ei ollut merkittävää vaihtelua keskenään, ruututeksteissä sitä esiintyi hieman enemmän kuin muissa suomen varianteissa. *Miksei*- ja *muttei*-sulaumapartikkeleiden osalta tuloksissa oli osakorpuksen tulosten eduksi merkittävää eroa Käännoosuomen materiaaleihin. *Jottei*-sulauman muotojen osalta taas ruututeksteillä ja käännoosuomella oli pienoinen ero, mutta alkuperäissuomen ero ruututeksteihin ja kaikkiaan käännettyyn suomeen oli merkittävä. Vaikka *siksei*-sulaumapartikkelin muotoja löytyi ruututeksteistä melko vähän, on kolmen korpuksen vertailussa kuitenkin huomattavaa, ettei Käännoosuomen kummastakaan sivukorpuksesta löytynyt ainuttakaan esiintymää tälle sulaumamuodolle. Varsinkin *miksei* esiintyy eri taivutusmuotoineen merkittävästi enemmän ruututeksteissä kuin käännoosuomessa tai supisuomessa.

Kun eri sulaumamuotojen kaikkia taivutusmuotoja vertailtiin keskenään, oli selvää, että kaikissa kolmessa korpuksessa yksikölliset muodot olivat selkeästi yleisempiä kuin monikolliset muodot. Kaiken kaikkiaan yleisin persoonamuoto kaikissa korpuksissa oli yksikön 3. persoona, joka muutoin oli kaikissa korpuksissa jokaisen sulaumapartikkelin osalta yleisin persoona. Ainoana poikkeuksena tähän sääntöön oli se, että ruututeksteissä *miksei*-sulaumamuodon yleisin persoona oli yksikön 2. persoonamuoto. Kuten taulukosta 13 käy selville, valtaosin tuloksissa ei näy mitään suuria eroja kolmen korpuksen välillä. Kaikkien

yksiköllisten ja kaikkien monikollisten muotojen osalta käännetyn suomen tulokset olivat lähes identtiset, joskin vasta pyöristysten jälkeen (ennen pyöristystä luvut olivat osakorpuksen osalta 88,5 %/11,4 % ja käännoissuomen osalta 89,3 %/10,7 %). Huomion arvoista on kuitenkin se, että yksittäisten persoonamuotojen kohdalla alkuperäissuomen ja käännoissuomen tulokset olivat lähempänä toisiaan kuin käännetyn suomen variantit keskenään. Kaiken kaikkiaan ruututeksteissä esiintymät jakaantuivat tasaisemmin eri persoonamuotojen kesken, vaikka yksikön 3. persoona olikin myös ruututeksteissä yleisin persoonamuoto. Alkuperäissuomessa ja käännoissuomessa esiintymät painoutuivat enemmän yksikön ja monikon 3. persoonalle. Merkittävin ero yksittäisen persoonamuodon kohdalla oli yksikön 2. persoonan muodoissa, joita esiintyi ruututeksteissä merkittävästi enemmän kuin alkuperäissuomessa ja käännoissuomessa.

Taulukko 4. Persoonamuodot eri korpuksissa

	OK	APS	KS
y.1.p	18 %	11 %	12 %
y.2.p.	16 %	3 %	6 %
y.3.p.	55 %	77 %	71 %
<b>YKS.</b>	<b>89 %</b>	<b>92 %</b>	<b>89 %</b>
m.1.p	3 %	1 %	2 %
m.2.p.	4 %	0 %	1 %
m.3.p.	5 %	7 %	7 %
<b>MON.</b>	<b>11 %</b>	<b>8 %</b>	<b>11 %</b>

Edellä mainitut tulokset koskevat valtaosin yksittäisten syntagmaattisten sulaumien kokonaistuloksia. Seuraavaksi on lueteltu tärkeimmät huomiot, kun kutakin syntagmaattista sulaumaa taivutusmuotoineen on vertailtu kolmen korpuksen kesken. Kaikki vertailut on tehty suhteellisia frekvenssilukuja tarkastelemalla. Kolmen korpuksen vertailussa tarkasteltiin myös liitepartikkeliin päättyviä muotoja sekä monikon 3. persoonan vokaalisoinnun muutoksia, joita käsitellään erikseen alaluvuissa 6.2.2 ja 6.2.3. Käytän partikkelikohtaisissa yhteenvedoissa sanaa *yhteistulos* viittaamaan kaikkien eri syntagmien yhteenlaskettuja tuloksia yhden korpuksen osalta ja sanaa *kokonaistulos* viittaamaan yhdessä korpuksessa esiintyviä yhden syntagman eri taivutusmuotojen tuloksia.

*Ehkei*-sulaumapartikkelin kohdalla tuloksissa ei ollut merkittäviä eroja muiden kuin yksikön 3. persoonan osalta, mutta se vaikutti jo kokonaistuloksiinkin. Eniten sulaumia oli ruututeksteissä, vähiten alkuperäissuomessa. Ruututeksteissä esiintyi selkeästi enemmän



tuloksia kuin muissa varianteissa, etenkin ero alkuperäissuomeen oli merkittävä. Kaikissa kolmessa yksiköllisiä persoonamuotoja oli selkeästi eniten – yleisin muoto oli kaikissa korpuksissa yksikön 3. persoona. Alkuperäissuomessa se tosin oli ainoa esiintyvä persoona, ja käännössuomestakin puuttuu täysin monikon 2. persoona.

*Ellei*-sulaumakonjunktio oli syntagmaattisista sulaumamuodoista ainoa, jossa käännössuomesta löytyi suhteessa yhteismäärältään enemmän sulaumia kuin ruututeksteistä, joskin ero oli hyvin pieni. Vain yksikön 3. ja monikon 3. persoonassa käännössuomessa oli enemmän sulaumia, muissa muodoissa osakorpuksen tuloksissa oli suhteessa enemmän sulaumia, joskin yksikön 1. persoonassa alkuperäissuomessa ilmeni enemmän sulaumia kuin käännössuomessa. Yksikölliset muodot olivat myös *ellei*-sulauman kohdalla yleisempiä kuin monikolliset, tosin yksikön 3. persoonan muodot muodostivat valtaosan esiintymistä kaikissa kolmessa korpuksessa. Lisäksi kiinnostava seikka on se, että alkuperäissuomessa yksikön 2. persoonassa esiintyi merkittävästi vähemmän esiintymiä kuin kahdessa muussa variantissa (*ellet*-muodon suhteelliset frekvenssit: OK 37,6/APS 5,2/KS 16,3).

*Ettei*-sulaumakonjunktio eri taivutusmuotoineen oli kaikissa kolmessa korpuksessa ylivoimaisesti yleisin syntagmaattinen sulauma. Kaiken kaikkiaan *ettei*-sulauman muotoja esiintyi eniten osakorpuksen materiaaleissa ja vähiten alkuperäissuomen korpuksessa, lisäksi yksiköllisiä muotoja oli selvästi enemmän kuin monikollisia kaikissa kolmessa korpuksessa. Vaikka sulaumamuodon kokonaistuloksissa eri korpusten erot toisiinsa olivat pieniä (suhteelliset frekvenssiluvut: OK 1 498,3/APS 1 329,9/KS 1 384,4), oli yksittäisten persoonamuotojen kohdalla kuitenkin suuria eroja. Huomattavin eroavaisuus oli se, että yksikön 3. persoonan muotoja löytyikin suhteessa eniten alkuperäissuomesta ja vähiten ruututeksteistä. Lisäksi, vaikka yksikön 3. persoonan muodot olivat kaikissa kolmessa korpuksessa yleisin taivutusmuoto, nousi tästä aineistosta esiin *ettei*-sulauman kohdalla yksikön 1. persoonan esiintymät. Ne olivat yllättävän suuria kaikkien kolmen korpuksen osalta. Yleinen tendenssi muiden sulaumien kohdalla oli, että yksikön 3. persoona erottui selkeästi muista taivutusmuodoista, ja vaikka yksikölliset muodot olivatkin monikollisia selvästi yleisempiä, ei yleensä ottaen yksikön 1. ja 2. persoonan suhteellisissa frekvenssiluvuissa ollut merkittäviä eroavaisuuksia. Lisäksi yksikön 2. persoonan muoto *ettet* nousi tästä aineistosta esiin sillä, että sitä esiintyi osakorpuksessa selkeästi enemmän kuin kahdessa muussa korpuksessa. Ruututeksteissä *ettet*-muodon suhteellinen frekvenssiluku oli

165,2, käännössuomessa 53,4 ja alkuperäissuomessa 29,2. Huomioitavaa on myös se, että monikon 3. persoonan osalta oli käännössuomessa suhteellisesti jonkin verran enemmän esiintymiä kuin käännöskorpuksessa ja selvästi enemmän kuin ruututeksteissä.

**Jollei**-sulaumakonjunktion tulokset noudattivat melko hyvin kaikkien syntagmaattisten sulaumien yhteistuloksia: eniten syntagmaattisia sulaumia esiintyi osakorpuksessa, vähiten alkuperäissuomessa; yksiköllisiä muotoja löytyi enemmän kuin monikollisia ja yleisin taivutusmuoto oli yksikön 3. persoona. Pientä vaihteluakin kuitenkin myös esiintyi, sillä alkuperäissuomesta ei löytynyt lainkaan monikon 2. persoonan muotoja. Lisäksi yksikön 2. persoonan muodoissa käännetyin suomen tulokset erosivat huomattavasti alkuperäissuomen tuloksista, kun taas ruututekstien ja käännössuomen ero toisiinsa oli pieni (muodon suhteelliset frekvenssiluvut: OK 36,9/APS 13,8/KS 35,4). Lisäksi monikon 3. persoonan etuvokaalista muotoa *jolleivät* esiintyi eniten käännössuomessa, joskin ero ruututeksteihin oli melko pieni. Monikon 3. persoonan vokaalisoinnun muutoksista on kerrottu tarkemmin alaluvussa 6.2.3.

**Jottei**-sulaumakonjunktiota esiintyi kokonaisuutena vähiten alkuperäissuomessa ja eniten ruututeksteissä, joskin osakorpuksen ja käännössuomen sivukorpuksen ero oli pieni. Myös *jottei*-sulauman kohdalla yksiköllisiä muotoja oli enemmän kuin monikollisia. Käännössuomesta puuttuivat vallan monikon 1. ja 2. persoona; alkuperäissuomesta taas puuttui sekä yksikön 1. ja 2. persoona että myös monikon 1. ja 2. persoona. Sen lisäksi ainoastaan osakorpuksen materiaaleista löytyi *jottei*-sulauman monikon 3. persoonan takavokaalinen variantti *jotteivat*. Kuitenkin yksikön 1. persoonan muotoa ja monikon 3. persoonan etuvokaalista varianttia *jotteivät* esiintyi käännössuomessa enemmän kuin ruututeksteissä.

**Miksei**-sulaumapartikkeli oli tässä aineistossa useammassakin mielessä poikkeuksellinen. Toisin kuin valtaosassa syntagmaattisten sulaumien tuloksista, *miksei*-partikkelia esiintyi osakorpuksessa merkittävästi enemmän kuin kahdessa muussa korpuksessa, ja osakorpuksen osalta kaikkien taivutusmuotojen suhteellinen frekvenssiluku olikin 436,6, kun taas käännössuomen vastaava luku oli 87,4 ja alkuperäissuomen 77,0. Yksiköllisiä muotoja oli kaikissa kolmessa korpuksessa enemmän kuin monikollisia, mutta poikkeuksellisesti osakorpuksen kohdalla yleisin muoto ei ollutkaan yksikön 3. persoona, vaan yksikön 2. persoona, vaikkakin hienoisella erolla. Yleisen linjan mukaan yksikön 1. persoonan muotoja

löytyi korpuksista enemmän kuin 2. persoonan. Lisäksi yksikön 2. ja 3. persoonan muotojen yhteenlaskettu suhteellinen osuus selittää valtaosin osakorpuksen kokonaistuloksen merkittävän eron kahteen muuhun korpukseen verrattuna. Näiden kahden persoonamuodon yhteenlaskettu suhteellinen frekvenssiluku oli osakorpuksessa 339,8, käännessuomessa taas 66,1 ja alkuperäissuomessa 65,3. Käännessuomen ja alkuperäissuomen korpuksissa yleisin persoona oli siis yleisen linjan mukainen yksikön 3. persoona. Lisäksi alkuperäissuomessa oli poikkeuksellisesti hieman enemmän yksikön 3. persoonan muotoja kuin käännessuomessa.

**Muttei**-sulaumakonjunktion kohdalla osakorpuksen kokonaistuloksen suhteellinen frekvenssiluku oli huomattavasti suurempi kuin Käännessuomen korpuksen materiaaleissa (OK 146,8/APS 38,4/KS 82,7). Lisäksi käännessuomen sivukorpuksessa ei ollut lainkaan muotoja *muttette* eikä *mutteivat*; alkuperäissuomesta taas puuttui jopa kolme muotoa, *muttemme*, *muttette* sekä *mutteivat*. Ei olekaan yllättävää, että tämänkin sulaumakonjunktion kohdalla yksikölliset muodot olivat monikollisia yleisempiä ja että yksikön 3. persoona oli kaikissa kolmessa korpuksessa yleisin muoto. Poikkeuksellista on kuitenkin se, että käännessuomessa esiintyy enemmän monikon 3. persoonan etuvokaalista varianttia *mutteivät* kuin osakorpuksen ruututeksteissä. Kiinnostavaa on myöskin se, että siinä missä suurimmassa osassa sulaumapartikkeleita (koskee kaikkia korpuksia) yksikön 1. ja 2. persoonamuodon välillä ei ollut suuria eroja, oli *muttei*-sulauman kohdalla poikkeuksellisesti merkittävä ero näiden kahden välillä osakorpuksen materiaaleissa. Myös käännessuomen osalta näiden kahden ero oli kohtuullisen suuri.

**Siksei**-sulaumapartikkeli oli tässä tutkimusaineistossa poikkeuksellinen sikäli, ettei sen muotoja löytynyt lainkaan Käännessuomen korpuksesta, ei niin alkuperäissuomeksi kirjoitetuissa teksteissä kuin suomeksi käännettyissä teksteissäkään. Myöskään osakorpukselta ei löytynyt kuin yksikön 1. ja 3. persoonan sekä monikon 3. persoonan muotoja, ensimmäistä ja jälkimmäistä vain yhteensä neljä esiintymää. Osakorpuksen yleisin muoto oli siis yksikön 3. persoona ja muutoinkin täten yksiköllisiä muotoja löytyy enemmän kuin monikollisia.

**Vaikkei**-sulaumakonjunktion yhteistuloksissa osakorpuksessa esiintyi eniten ja alkuperäissuomen sivukorpuksessa vähiten *vaikkei*-sulauman muotoja. Yhteistulosten yleisin persoona kaikkien kolmen korpuksen osalta on yksikön 3. persoona ja kaiken kaikkiaan yksiköllisiä muotoja esiintyy enemmän kuin monikollisia. Taivutusmuotojen erot toisiinsa olivat kuitenkin melkoiset. Ensinnäkin osakorpuksen materiaaleissa yksikön 1. ja 2.

persoonan tuloksissa oli merkittävä eroavaisuus (suhteelliset frekvenssit: *vaikken* 16,5/*vaikket* 6,5), kun taas Käännössuomen materiaaleissa ero näiden kahden välillä oli selvästi pienempi. Lisäksi yksikön 2. ja monikon 1. persoonan osalta alkuperäissuomessa ilmeni enemmän sulaumamuotoja kuin käännössuomessa, eniten niitä oli joka tapauksessa ruututeksteissä. Kuitenkin monikon 3. persoonan molempien varianttien (*vaikkeivat* ja *vaikkeivät*) esiintymää löytyi vertailussa eniten käännössuomen sivukorpuksesta.

### 6.2.2 Syntagmaattisten sulaumien päätteelliset muodot

Syntagmaattisia sulaumia tutkittaessa tarkasteltiin myös erikseen, miten usein jo valmiiksi yhdistyneisiin muotoihin on täydennetty liitepartikkeleita. Päätteellisiä muotoja tarkasteltiin siis ainoastaan sulaumia tutkittaessa, eikä lainkaan erilleen kahtena sanana kirjoitettujen muotojen kohdalla, jottei varsinaisen tutkimusmateriaalin määrä nousisi liikaa. Päätteellisiä muotoja tutkittiin niin osakorpuksesta kuin Käännössuomen korpuksenkin materiaaleista, erikseen alkuperäissuomeksi kirjoitetuista teksteistä sekä käännössuomen teksteistä. Tutkimuksen tarkoituksena oli tarkastella, poikkeako ruututekstien kieli tässä suhteessa jotenkin muista suomen varianteista.

Kaiken kaikkiaan liitepartikkeliin päättyviä sulaumia löytyi tutkimus- ja vertailumateriaalista varsin vähän (tarkka lista päätteistä näkyvillä edempänä taulukossa 14). Päätteellisiä muotoja löytyi vain sulaumapartikkelien *ellei*, *ettei*, *jollei*, *miksei*, *muttei* ja *vaikkei* taivutusmuodoista, sulaumien *ehkei*, *jottei* ja *siksei* taivutusmuodoissa ei siis esiintynyt lainkaan liitepartikkeliin muotoja. Kaikkien korpusten osalta ehdottomasti eniten päätteellisiä muotoja esiintyi sulauman *ettei* taivutusmuodoissa ja sulauman sisällä taas eniten yksikön 3. persoonassa. Koska *ettei* taivutusmuotoineen oli muutenkin selkeästi yleisin syntagmaattinen sulauma kaikissa korpuksissa, ei edellä mainittu tulos tuottanut yllätystä.

Taulukko 5. Sulaumien päätteet korpuksissa

	OK	APS	KS	
-hän	4	1	1	6
-kaan	-	-	1	1
-kään	-	1	1	2
-kä	1	-	-	1
-kö	272	94	131	497
-pä	2	5	-	7
-sä	-	1	1	2
<i>kaikki päätteet</i>	279	102	135	516

Kuten taulukosta 14 näkyy, valtaosa sulaumiin liittyneistä liitepartikkeleista koostui korpuksesta riippumatta päätteestä *-kö*, tarkemmin 96,3 prosenttia kaikista päätteellisistä muodoista. Muita päätteitä oli korpuksissa käytetty lähinnä yksittäisissä tapauksissa (esimerkkeinä *elleikään*, *mutteipä* ja *miksei hän*). Yleisin liitepartikkeliin päättynyt sulaumamuoto kaikissa materiaaleissa oli siis *etteikö*. Muista, satunnaisista löydöksistä kiinnostavin oli kerran käännessuomen sivukorpuksessa ja kerran alkuperäissuomen korpuksessa esiintynyt *miksetsä*, joka on puhekielinen kolmen sanan syntagmaattinen sulauma, jota ei kuitenkaan löytynyt lainkaan osakorpuksen tekstityksistä.

Kaikissa osakorpuksessa esiintyneistä sulaumamuodoista vain 0,4 prosenttia päättyi liitepartikkeliin, kun vastaava luku oli käännessuomen osalta 1,2 prosenttia ja alkuperäissuomen kohdalla 15,6 prosenttia. Alkujaan suoraan suomeksi kirjoitetuissa teksteissä siis esiintyi syntagmaattisten sulaumien kokonaismäärään nähden selkeästi enemmän liitepartikkeliin päättyviä muotoja kuin käännessä suomessa. Prosenttiluvut kuitenkin kertovat vain sulaumien kokonaismäärään vertautuvan tilanteen kunkin korpusten osalta, eikä varsinaisesti korpusten välisistä eroista.

Helpoin tapa vertailla esiintymien määrää korpusten välillä oli myös päätteellisten muotojen kohdalla laskea niiden esiintymistä suhteelliset frekvenssiluvut. Kuten sulaumien kokonaismäärien kanssa, päätteellisten sulaumamuotojen suhteelliset frekvenssit laskettiin suhteuttamalla esiintymien määrä korpuksessa miljoonaa sanaa kohti. Taulukossa 15 näkyy tämän tutkimuksen materiaaleissa esiintyneiden päätteellisten syntagmaattisten sulaumien suhteelliset frekvenssiluvut per korpus. Kuten taulukosta on nähtävillä, miljoonan sanan joukkioon suhteutettuna alkuperäissuomessa esiintyy eniten ja ruututeksteissä taas vähiten päätteellisiä syntagmaattisia sulaumia. Vaikka nämä lukemat eivät suuresti eroa toisistaan, näkyy niistä kuitenkin, että alkuperäissuomessa ja muussa käännessuomessa esiintyi näissä aineistoissa enemmän päätteellisiä sulaumamuotoja kuin osakorpuksen tekstityksissä.

Taulukko 6. Päätteellisten muotojen suhteelliset frekvenssiluvut

	OK	APS	KS
kaikki päätteelliset muodot	10,5	26,6	23,0

### 6.2.3 Syntagmaattisten sulaumien vokaalisoinnun muutokset

Kuten tässä tutkimuksessa on aiemmin todettu, voi syntagmaattisissa sulaumissa tapahtua vokaalisoinnun muutoksia, kun takavokaalinen konjunktio yhdistyy kieltoverbiin, ja tällöin monikon 3. persoonan tunnus voi olla joko etuvokaalinen *-eivät* tai takavokaalinen *-ivat*. *Iso Suomen kieliopin* mukaan etuvokaaliset muodot ovat takavokaalisia muotoja yleisempiä. (ISK 2005, 165.) Koska vokaalisoinnun muutoksia tapahtuu tässä tapauksessa vain sulaumien kohdalla, on tätä asiaa tutkittu ainoastaan syntagmaattisten sulaumien hakutuloksista, eikä tietenkään erilleen kahtena sanana kirjoitetuista muodoista. Tämän tutkimuksen aineistossa sekä etuvokaalisia että takavokaalisia syntagmaattisia sulaumia oli sanojen *jollei*, *jottei*, *muttei* ja *vaikkei* monikon 3. persoonan muodoissa ruututekstityksissä, Käännössuomen korpuksen materiaaleista ei sen sijaan löytynyt lainkaan osumia varianteille *jotteivat* ja *mutteivat*. Vaikka aiemmin on jo puhuttu yksittäisten sulaumien kohdalla hieman vokaalisointujen muutoksista, on tarkemmat tiedot siitä koottu tämän alaluvun alle. Lisäksi on huomautettava, että kaikki vokaalisoinnun tulokset on laskettu päätteettömien sulaumien esiintymistä.

Kun korpusten tuloksista verrattiin keskenään etu- ja takavokaalisten muotojen esiintymiä, oli tämän tutkimuksen tulos yhtenevä ISK:n (2005, 165) tietojen kanssa, eli etuvokaalisia muotoja esiintyi aineistossa enemmän kuin takavokaalisia muotoja, oli sitten kyse ruututeksteistä, muusta käännössuomesta tai alun perin suomeksi kirjoitetuista teksteistä. Osakorpuksen materiaalien osalta 82 prosenttia monikon 3. persoonan esiintymistä oli etuvokaalisia, kun taas alkuperäissuomessa niiden osuus oli 90 prosenttia ja käännössuomessa 92 prosenttia. Kiinnostavaa on, että ruututeksteistä löytyy enemmän takavokaalisia muotoja kuin muissa tutkituissa suomen varianteissa. Käännössuomen korpuksen aineistostahan siis puuttuivat täysin takavokaaliset muodot *jotteivat* ja *mutteivat*.

Taulukko 7. Vokaalisoinnun muutokset aineistossa (per 1 000 000 sanaa)

	OK	APS	KS
jolleivat	0,08	0,03	0,05
jolleivät	0,15	0,05	0,19
jotteivat	0,05	-	-
jotteivät	0,16	0,05	0,27
mutteivat	0,03	-	-
mutteivät	0,31	0,23	0,37
vaikkeivat	0,03	0,03	0,05
vaikkeivät	0,21	0,16	0,32

Kuten taulukosta 16 on nähtävissä, suhteellisten frekvenssilukujen mukaan syntagmaattisten sulaumien takavokaaliset (-*eivat*) variantit olivat sulaumapartikkeleiden *jollei*, *jottei* ja *muttei* kohdalla yleisempiä ruututeksteissä kuin muussa käännössuomessa, vähiten takavokaalisia muotoja taas oli alkuperäissuomessa. Sanan *vaikkei* kohdalla käännössuomessa taas esiintyi enemmän takavokaalisia muotoja kuin ruututeksteissä. Kaiken kaikkiaan takavokaalisia muotoja ilmeni alkuperäissuomessa vähemmän kuin käännetyin suomen eri variaatioissa.

Toisin kuin takavokaalisia muotoja, samojen sulaumamuotojen etuvokaalisia (-*eivät*) monikon 3. persoonan variantteja löytyi paremmin. Jokaista muotoa ilmeni jokaisessa korpuksessa; eniten käännössuomen sivukorpuksessa ja vähiten alkuperäissuomessa. Vaikka jokaisen sulaumamuodon kohdalla käännössuomessa oli suhteellisesti enemmän etuvokaalisia muotoja kuin ruututeksteissä, olivat käännössuomen sivukorpuksen ja osakorpuksen tulosten erot toisiinsa melko pieniä, käännetyin suomen ero alkuperäissuomeen taas oli merkittävämpi.

## 7 YHTEENVETO JA LOPPUPÄÄTELMÄT

Tämän tutkimuksen päätavoitteena oli siis tutkia partikkelin ja kieltoverbin syntagmaattisten sulaumien esiintymistä ruututeksteissä ja verrata sitä alun perin suomeksi kirjoitetuissa teksteissä ja käännösuomessa ilmeneviin samoihin sulaumamuotoihin. Tämän vertailun avulla oli pyrkimys arvioida, olivatko kääntäjät pitäneet tärkeämpänä kohdekielen konventioita ja katsojien mahdollisia odotusnormeja vaiko lähtötekstin ja kohdetekstin suhdenormia ruututekstikäntämisen rajoitteiden tähden. Lisäksi tutkimuksen toissijaisena tarkoituksena oli tutkia, miten paljon ruututeksteissä esiintyy syntagmaattisten sulaumien erilleen kirjoitettuja muotoja, jotta nähtäisiin, miten sulaumamuotoja on käytetty ruututeksteissä ja miten paljon niitä taas on jätetty käyttämättä. Erilleen kirjoitettujen muotojen osalta tutkimukseen hyväksyttiin vain tapaukset, jotka olisi yhtä hyvin voitu muuttaa syntagmaattisiksi sulaumiksi. Tutkimuksen hypoteesina oli, että syntagmaattisia sulaumia esiintyisi ruututeksteissä enemmän kuin muissa suomen varianteissa av-kääntämisen erityislaatuisen luonteen tähden. Tällöin oletuksena oli myös, että ruututekstejä käännettäessä suhdenormin vaikutus olisi odotusnormia suurempi juurikin tekstitysten rajoitteiden ja konventioiden takia. Toisin kuin edeltävässä tulosluvussa, tässä päättävässä luvussa referoidaan tuloksia tavoitteiden mukaisessa järjestyksessä. Ensimmäisenä siis keskitytään päätavoitteeseen, eli kolmen suomen variantin välisiin eroihin syntagmaattisten sulaumien osalta, sitten sulaumamuotojen ja erilleen kirjoitettujen muotojen esiintymistä tekstityksissä ja viimeisenä tarkastellaan syntagmaattisten sulaumien esiintymistä ruututeksteissä.

***Kolmen korpuksen vertailu.*** Kuten alaluvusta 6.2 kävi ilmi, ruututeksteissä ilmeni selkeästi enemmän syntagmaattisia sulaumia kuin alkuperäissuomessa ja käännösuomessa – ja täten kaiken kaikkiaan käännettyssä suomessa supisuomeen verrattuna. Lisäksi kokonaisuutena alkuperäissuomen tulosten ero käännösuomen tuloksiin oli pienempi kuin käännösuomen ero ruututeksteihin, ja vain yhden konjunktion kokonaistulosten osalta käännösuomesta löytyi suhteessa enemmän esiintymiä kuin ruututeksteistä ja tällöinkin ero oli hiuksenhieno. Voidaan siis sanoa, että kokonaisuutena ruututekstit eroavat syntagmaattisten sulaumien osalta niin käännösuomesta kuin alkuperäissuomestakin. Koska käännösuomen ja alkuperäissuomen tuloksissa ei ole merkittäviä eroja, voidaan siis todeta, että ruututekstikäntäminen eroaa tältä osin selvästi muista tutkituista suomen varianteista.



Vaikka kaiken kaikkiaan ruututeksteissä ilmeni eniten syntagmaattisia sulaumia, päättyivät osakorpuksen materiaaleissa syntagmaattiset sulaumat merkittävästi vähemmän liitepartikkeliin kuin alkuperäissuomessa. Oletettavasti jo valmiiksi tiivistetyt muodot yhdessä liitepartikkelin kanssa haittaisivat av-kääntäjien mielestä tekstitysten luettavuutta. Tirkkonen-Conditiin ja Mäkisalonen (2007) tutkimuksen valossa kuitenkin ei voida olettaa liitepartikkelien sinänsä kuuluvan ruututekstien rajoitusten tähden poisjätettävään ainekseen, sillä heidän tuloksensa oli, että ruututekstien kieli eroaa muista käännösuomen varianteista käyttämällä hyväkseen suppeaa ja ytimekästä kielenkäyttöä hyödyntäviä koheesiokeinoja. Kiinnostavaa kyllä, myös käännösuomessa esiintyi suhteellisen vähän liitepartikkeliin päättyviä muotoja. Selkeästi yleisin liitepartikkeliin päätynyt muoto oli kaikissa korpuksissa *etteikö*. Mielenkiintoista olisi sen osalta verrata, toteutuuko aiemmissä saman muodon tutkimuksista saadut tulokset myös ruututekstien osalta – esimerkiksi *etteikö*-muodolla alkavaa sivulausetta edeltää usein kielteinen päälause (kts. Larjavaara 1992 ja Maamies 1993).

Siinä missä liitepartikkeliin päättyvät syntagmaattiset sulaumat olivat tässä aineistossa ilmeisimmin paremminkin alkuperäissuomen kuin käännetyn suomen piirre, olivat taas sulaumamuotojen monikon 3. persoonan takavokaaliset variantit (*-eivat*) ennemminkin käännetyn suomen kuin alkuperäissuomen piirre. Tältä osin käännösuomi oli kuitenkin lähempänä alkuperäissuomea kuin ruututekstejä. Mielenkiintoista olisikin tutkia, miksi tekstityksissä ilmeni eniten takavokaalisia variantteja, kun tiivistyksen näkökulma ei tässä kohdin tarjoa selitystä asialle. Kun taas eri sulaumamuotojen taivutusmuotoja vertailtiin keskenään, oli selvää, että kaikissa kolmessa korpuksessa yksikölliset muodot olivat selkeästi yleisempiä kuin monikolliset muodot.

Yhtenä kiinnostavana seikkana kolmen korpuksen tulosten vertailussa nousi esiin syntagmaattisten sulaumien persoonamuotoisuus. Kaikissa kolmessa korpuksessa yksikölliset muodot olivat monikollisia yleisempiä, lisäksi kaikissa yleisin persoona oli yksikön 3. persoona yhtä poikkeusta lukuun ottamatta (*mikset* ruututeksteissä). Vaikka erot olivat suhteellisen pieniä, korostui Käännösuomen korpuksen materiaaleihin verrattuna eniten yksikön 2. persoonan yleisyys ruututeksteissä. Muutoinkin ruututeksteissä esiintymät jakaantuivat tasaisemmin eri persoonamuotojen kesken kuin Käännösuomen korpuksessa, jossa molemmissa varianteissa yksikön 3. ja monikon 3. persoona korostuivat. Lisäksi

monikon 2. persoonan muotoja löytyi ruututeksteistä suhteessa enemmän kuin alkuperäissuomesta ja käännessuomesta. Kaiken kaikkiaan yksikkömuotoisten syntagmaattisten sulaumien kohdalla erot kolmen korpuksen vertailussa olivat suurimmillaan ja valtaosin erot monikollisten muotojen suhteen olivat pienemmät. Ainoina merkittävinä poikkeuksina edellä mainittua vastaan ilmenivät *ettei*-sulaumakonjunktion monikon 2. persoonan muoto *ettette* sekä *miksei*-sulaumapartikkelin monikon 2. persoonan muoto *miksette*. Varsinkin muodon *ettette* yleisyys ruututekstityksissä hämmästytti, sillä alkujaan odotin kyseisen muodon olevan ruututeksteissä hyvinkin harvinainen sen ”erikoisen” ulkomuodon tähden. Mitä todennäköisimmin persoonamuotojen eroavaisuudet johtuvat tekstitysten dialogisuudesta. Monikon 2. persoonan muotojen osalta on muistettava, että Ylen ohjelmistossa esiintyy usein etenkin brittiläistä pukudraamaa ja vanhahtavia saksalaisia dekkareita, joissa olettaisi löytyvän teitittelyä.

Kuten tässä tutkimuksessa on jo aiemmin huomautettu, kolmen korpuksen vertailussa ei otettu lainkaan huomioon syntagmaattisten sulaumien erilleen kirjoitettuja muotoja. Kiinnostavaa olisi silti tutkia, miten Käännessuomen korpuksessa esiintyy erilleen kirjoitettuja muotoja, ja vertailla, miten näiden kahtena sanana kirjoitettujen muotojen esiintymät vertautuvat samassa materiaalissa ilmenneiden syntagmaattisten sulaumien määrään. Kuriositeettina huomautettakoon, että kun hakusanalistan toimivuutta testattiin ennen varsinaisen tutkimustyön alkamista yhden amerikkalaisen romaanin suomennoksen sisältävällä korpuksella (kts. lisää Koivikko 2008), olivat tässä koetutkimuksessa syntagmaattiset sulaumat harvinaisempia kuin erilleen kirjoitetut muodot. Toki noin pienellä otannalla yhden kääntäjän tekemät subjektiiviset ratkaisut vaikuttivat ratkaisevasti, mutta koetutkimuksen tulos oli vastoin ISK:n (2005, 165) tulosta, jonka mukaan sulaumamuodot ovat yleiskielessä yleisempiä kuin niiden kahtena sanana kirjoitetut muodot. Tästä syystä olikin tärkeää, että sekä varsinainen tutkimusmateriaali (osakorpus) että sen vertailumateriaali (Käännessuomen korpus) olivat tarpeeksi laajoja, jotta yhden kääntäjän ratkaisut eivät pääse liikaa vaikuttamaan tuloksiin.

***Osakorpuksen sisäinen vertailu.*** Kun osakorpuksen materiaaleista tutkittiin sekä syntagmaattisia sulaumia että samojen kielellisten elementtien erilleen kirjoitettuja muotoja, huomattiin, että ruututeksteistä löytyi vain hieman enemmän sulaumamuotoja kuin kahtena sanana kirjoitettuja muotoja. Huomattavaa on, että kaikki nämä sulautumattomat tapauksethan

olivat sellaisia, jotka olisi voinut hyvin pienellä vaivalla muuttaa syntagmaattisiksi sulaumiksi, sillä tuloksista poistettiin tutkimuksen puitteisiin sopimattomien löydöksen lisäksi myös tapaukset, joissa syntagmaattiseksi sulaumaksi muuttaminen vaatisi selkeitä muutoksia lauserakenteeseen. Sulaumamuotoja olisi voinut siis löytyä korpuksesta lähes puolet enemmän, muttei kuitenkaan löytynyt. Koska molemmat variantit on tutkittu samasta, paljon dialogia sisältävästä materiaalista, ei syytä eroavaisuudelle voi kaiken kaikkiaan vierittää dialogisuudelle. Se, miksei valtaosaa mahdollisista tapauksista ole alkujaan lyhennetty syntagmaattisiksi sulaumiksi, selittyy varmasti valtaosin sillä, että kääntäjät ovat halunneet pitää tekstityksensä mahdollisimman helppoina ymmärtää. Esimerkiksi muoto *ettei* on kirjallisessakin materiaalissa yleinen muoto, oli teksti sitten kirjoitettu alun perin suomeksi tai käännetty suomeksi, joten olisi melko oletettavaa, että muoto on jo tuttuutensa tähden katsojille helppolukuinen. Mielenkiintoista on, että ISK:n (2005, 165) mukaan nykykielessä sulaumat ovat kaiken kaikkiaan yleisemmin käytössä kuin kahtena sanana kirjoitetut muodot. Tämän huomautuksen valossa olin alun perin oletanut, että näiden kahden variantin ero olisi ruututeksteissä ollut suurempi.

Vaikka syntagmaattisten sulaumien ero erilleen kirjoitettuihin muotoihin kaikkien partikkelien yhteistuloksen osalta oli melko pieni, löytyi partikkelikohtaisissa tuloksissa melkoisenkin isoja eroavaisuuksia. Tuloksista nousivat eniten esiin erilleen kirjoitetuissa muodoissa lähes kahdeksan kertaa yleisemmin esiintynyt *mutta ei* taivutusmuotoineen ja ruututeksteissä taas lähes 52 kertaa useammin ilmennyt *siksei*. Vaikkei jälkimmäisen partikkelin kohdalla esiintymiä löytynyt ruututeksteistäkään montaa, on eroavaisuus kuitenkin merkittävä. Koska *siksei* taivutusmuotoineen oli harvinainen alkuperäissuomessa ja käännössuomessa, olettaisi tällöin sen ilmenevän myös ruututeksteissä ennemminkin erilleen kirjoitettuna kuin syntagmaattisena sulaumana. On silti huomautettava, että *siksei*-sulaumapartikkeli ilmeni ruututeksteissä kuitenkin suurimmaksi osaksi yksikön 3. persoonassa, ja varsinkin niin, että edeltävässä virkkeessä esiintyi kysyvä sulaumamuoto *miksei*, joten on syytä olettaa, että näiden kahden lähtötekstinä on englanninkielisessä dialogissa ollut sanailu *Why? - That's why*.

Kuten on nähtävissä taulukosta 17, valtaosin (partikkelien *ehkä, jos, jotta, mutta* ja *siksi* osalta) erilleen kirjoitettuja muotoja ilmeni ruututeksteissä enemmän kuin syntagmaattisia sulaumia silloin, kun sulaumamuodot olivat alkuperäissuomen ja käännössuomen tuloksissa

harvinaisempia kuin ruututeksteissä, joskin *ellei-* ja *jollei-*sulaumapartikkelien yhteenlaskettujen tuloksien ja *jottei-*sulaumakonjunktion tuloksien eroa käännetyn suomen sisällä oli vähäinen. *Että-*partikkeliä taas esiintyi kaikissa korpuksissa lähes samoissa määrissä, joten sekin kuuluu edellä mainittuun suuntaukseen, joskin päinvastaisessa muodossa, sillä sen muotoja löytyi materiaaleista enemmän yhteensulautuneina kuin erilleen kirjoitettuina. Poikkeuksen tähän sääntöön siis tuovat partikkelit *miksi* ja *vaikka*. Jälkimmäisen partikkelin kohdalla sulautuneita muotoja ja erilleen kirjoitettuja muotoja oli lähes yhtä paljon, mutta Käännösuomen korpuksessa samoja sulaumamuotoja esiintyi selvästi vähemmän kuin ruututeksteissä. Tällöin voisi olettaa näiden muotojen olevan hankalammin luettavia ja että siten ruututeksteissä *vaikka-*partikkeli ilmenisi selkeästi enemmän erilleen kirjoitettuina muotoina kuin syntagmaattisina muotoina. Edeltävä päätelmä voisi päteä myös *miksi-*partikkelin kohdalla, joskin vielä selkeämmin, eli Käännösuomen korpuksen ja ruututekstien välisen merkittävän eron tähden voisi kuvitella, että *miksei-*sulaumapartikkelin muodot olisivat tekstityksissä melko harvinaisia – olisivathan ne muun suomen varianttien tulosten valossa outoja ja hankalammin luettavia muotoja. Kuitenkin *miksei-*sulaumapartikkeli oli taivutusmuotoineen merkittävästi yleisempi aineiston ruututeksteissä kuin niiden erilleen kirjoitetut muodot. *Miksei-*sulaumapartikkelin muotojen erikoinen yleisyys ruututekstityksissä selittyy pitkälti sillä, että niillä on haettu puhekielisyyden illuusiota tekstityksiin. Kiinnostavaa olisikin tietää tarkemmin, miten yleisiä nämä *miksei-*sulaumapartikkelin muodot ovat yleispuhekielessä, eli selittyykö näiden muotojen tuttuuden tuntuisuus sillä. Koska alkuperäissuomesta ja käännösuomesta ei tutkittu lainkaan erilleen kirjoitettuja muotoja, ei edempänä olevaa taulukkoa 17 liitetty osaksi tuloksia esittelevää edeltävää lukua.

Taulukko 87. Eri materiaalien vertailu

	OK	APS	KS		OK
ehkei	30,3	3,4	15,3	ehkä + ei	173,8
ellei & jollei	316,2	181,7	294,1	jos + ei	490,1
ettei	1 498,3	1 329,9	1 384,4	että + ei	298,7
jottei	20,2	7,3	19,4	jotta + ei	36,0
miksei	436,6	77,0	87,4	miksi + ei	143,6
muttei	146,8	38,4	82,6	mutta + ei	1 127,0
siksei	0,6	-	-	siksi + ei	33,2
vaikkei	75,5	46,7	56,4	vaikka + ei	75,5
<b>kaikki</b>	<b>2 524,5</b>	<b>1 684,5</b>	<b>1 939,7</b>	<b>kaikki</b>	<b>2 378,0</b>

Kun syntagmaattisia sulaumia ja samojen muotojen erilleen kirjoitettuja varianteja tutkittiin, tarkasteltiin osakorpuksen materiaaleista erikseen lyhyissä lauseissa ja etenkin yksin virkkeissä esiintyviä partikkelimuotoja. Lyhyissä lauseissa esiintyi enemmän syntagmaattisia sulaumia kuin erilleen kirjoitettuja muotoja ja samoin yksinään lauseessa esiintyvissä tapauksissa, jälkimmäisen vertailun kohdalla ero oli suuri. Yksittäin esiintyneistä syntagmaattisista sulaumista merkittävästi yleisin oli kysyvä sulaumapartikkeli *miksei* taivutusmuotoineen. Tätä pitäisin itse yhtenä dialogin ominaisuutena, vaikkei tällaista vertailua tehtykään kolmen korpuksen materiaalien kesken. Toki on huomautettava, että mikäli näitä *miksei*-sulaumapartikkelin esiintymiä ei otettaisi huomioon yksittäin lauseessa esiintyviä muotoja tarkasteltaessa, löytyisi erilleen kirjoitetuista muodoista suhteessa enemmän esiintymiä. Sama tulos pätee myös lyhyissä lauseissa esiintyneitä muotoja tarkasteltaessa. Valtaosin siis lyhyissä lauseissa ja yksinään lauseissa ilmeni erilleen kirjoitettuja muotoja kuin sulaumamuotoja, mutta kaikkien tutkittujen partikkelien yhteistuloksissa syntagmaattiset sulaumat ovat hieman yleisempiä.

Vaikka lyhyissä lauseissa ilmenneitä sulaumamuotoja ei tämän tutkimuksen puitteissa tutkittukaan Käännösuomen korpuksen tuloksista, tarkastin tätä *miksei*-sulaumapartikkelia koskevan tuloksen myös alkuperäissuomen ja käännösuomen tuloksista tukeakseni aiempaa väitettäni. Tekstitysten osalta *miksei*-sulaumapartikkelin suhteellinen frekvenssiluku oli 73,0 sen esiintyessä pää- tai sivulauseessa ainoana lauseenjäsenenä, kun käännösuomessa vastaava luku oli 8,5 ja alkuperäissuomessa 6,0. Tämä jo yksinään todistaa, että Käännösuomen korpuksen materiaalin lyhyissä lauseissa esiintyvät syntagmaattiset sulaumat olisivat olleet hedelmällinen vertailukohde osakorpuksen tuloksille, vaikka ne eivät ikävä kyllä mahtuneet tämän tutkimuksen puitteisiin. Tämä vertailu antaa kuitenkin syytä olettaa, että yksin lauseissa ja virkkeissä esiintyvät syntagmaattiset sulaumat olisivat ainakin *miksei*-sulaumapartikkelin osalta paremminkin ruututekstien ja dialogin piirre kuin muun kirjoitetun suomen. Koska ruututekstien lyhyissä lauseissa kuitenkin ilmeni enemmän syntagmaattisia sulaumia kuin niiden erilleen kirjoitettuja muotoja ja kaiken kaikkiaan ruututeksteissä esiintyi enemmän sulaumamuotoja kuin erilleen kirjoitettuja muotoja, on siis olettaa, että nämä lyhyissä lauseissa ilmenevät syntagmaattiset sulaumat olisivat yksi ruututekstikäntämisen erityispiirteistä, joskin väite kaipaa enemmän tutkimusta taakseen.

Kaiken kaikkiaan siis eroavaisuudet syntagmaattisten sulaumien ja samojen kielellisten elementtien erilleen kirjoitettujen muotojen välillä selittyvät puhekielimäisyyden tavoittelulla ja ruututekstikäntämisen rajoituksilla. Vaikka tässä tutkimuksessa oli alun perin tarkoituksena tutkia, vaikuttiko aikarajoite merkittävästi kääntäjien ratkaisuun valita tai olla valitsematta sulaumamuotoa, ei tämä ajallinen aspekti lopulta sen vaatiman työmäärän tähden päässyt osaksi tätä tutkielmaa. Mielenkiintoista olisi kuitenkin laajentaa tutkimusta kattamaan tämä aspekti ja tarkistamaan esimerkiksi tietyllä otannalla, miten aikarajoite on näiden mikrorakenteiden kohdalla vaikuttanut. Toisaalta olisi myös mahdollista kysyä ainakin nykyisiltä Ylen kääntäjiltä, miten tietoisia he itse ovat valinnasta käyttää tai olla käyttämättä syntagmaattisia sulaumia, eli ovatko ne miten pitkälti kääntäjäkohtaisia.

*Syntagmaattiset sulaumat ruututeksteissä.* Kuten tulosluvusta tuli ilmi, syntagmaattiset sulaumat muodostivat vain pikkuruisen osan osakorpuksen sanastosta. Kuitenkin tällä pienellä tekstityksen osasella voi olla suuri merkitys siihen, miten ohjelmassa sanottu asia mahtuu yhteen repliikkiin. Kuten yleiskielessä (ISK 2005, 165), myös ruututeksteissä *ettei* oli sulaumakonjunktioista yleisin ja juurikin syntagmaattisena sulaumana, eikä erilleen kirjoitettuna muotona. Syntagmaattisten sulaumien yleisin kollokaatti, eli läheisin sana, oli Ylen ruututeksteissä preesensmuotoinen *olla*-verbi. Persoonapronomini seurasi ruututeksteissä melko harvoin syntagmaattista sulaumaa. Koska persoonapronominien ilmenemistä muissa suomen varianteissa ei tässä tutkimuksessa tarkasteltu vertailun vuoksi, ei voida suoraan olettaa, että persoonapronominin puuttuminen olisi ruututekstitysten erityispiirre, etenkin kun pikaisella otannalla *ehkei*-sulaumapartikkelin seurassa esiintyy persoonapronomini tekstityksissä 10 prosentissa kaikista esiintymistä, kun alkuperäissuomessa vastaava prosenttiluku oli 10 ja käännessuomessa 12. Tietenkin näin pienen otannan avulla ei vielä voida tehdä kaikkia partikkelin ja kieltoverbin syntagmaattisia sulaumia koskevia päätelmiä, mutta asia saattaisi olla tutkimisen arvoisen.

Kuten tässä luvussa ja edeltävässä tulosluvussa on jo tullut esiin, eri syntagmaattisten sulaumien käytössä oli ruututeksteissä selkeitä eroja. Valtaosin ne ilmenivät lauseenalkuisessa asemassa ja niiden edellä saattoi olla jokin rinnastuskonjunktio, partikkeli tai jokin muu pikkusana, mutta eroavaisuuksiakin oli, esimerkiksi *ehkei* yleensä aloitti virkkeen, kun taas *muttei* ei koskaan sijainnut lauseessa toisen sanan perässä, vaan aloitti aina lauseen, valtaosin sivulauseen. Muutoinkin *muttei*-sulaumapartikkeli oli useammassakin merkityksessä

erityislaatuinen syntagmaattinen sulauma, jonka esiintymien yleisyys todennäköisimmin selittyy puhekielimäisyyden tavoittelulla ja ruututestikääntämisen rajoituksilla. *Vaikei* erottui muista sulaumapartikkeleista sillä, että huomattava osa sitä seuranneista verbeistä ilmeni konditionaalissa tai päättyi *-kAAn*-liitepartikkeliin. Eri partikkelien väliset erot selittyvät pitkälti niiden funktiolla lauseessa, esimerkiksi kysyvä *miksei* saattoi esiintyä usein yksinäisenä lauseenjäsenenä virkkeessä, kun taas monien muiden partikkelien osalta lauseen piti jäädä kesken (esimerkiksi ”*Ellei...*”, jotta ne saattoivat esiintyä yksin ja tällöinkin ne useimmiten ilmenivät sivulauseessa, eivätkä omana virkkeenään.

**Loppuyhteenveto.** Riitta Jääskeläinen (2008, 94) on huomauttanut, että valtaosa ruututekstejä käsittelevästä tutkimuksesta on keskittynyt vain käyttämään ruututekstejä materiaalina, eikä varsinaisena tutkimuksen kohteena. Samoin ruututestikorpukselle alun perin asetettuihin odotuksiin kuului oletus, että korpusta käytetään muun muassa sen tutkimiseen, poikkeavatko ruututekstit muista suomeksi käännettyistä käännösteksteistä ja suoraan suomeksi kirjoitetuista teksteistä jotenkin (Mäkisalo 2008.) Lisäksi Yves Gambier (2008, 94) on muistuttanut, että olisi tärkeää tarkastella ”vallitsevan kirjallisen normin ja kirjoitetun tekstin konventioiden suhdetta tekstitystekstien kirjoitettuun kieleen”. Omalta osaltaan tämä tutkimus on vastannut kaikkiin kolmeen vaatimukseen. Kaiken kaikkiaan siis yleisimmin oli tässä tutkimuksessa niin, että alkuperäissuomi ja käännössuomi olivat tuloksissa kuin tuloksissa lähempänä toisiaan sekä alkuperäissuomi ja ruututekstitykset olivat kauimmaisina toisistaan. Tämä todistaa, että ainakin syntagmaattisten sulaumien osalta ruututekstit todellakin eroavat muista suomen varianteista suosimalla tiiviimpää kieltä, joskin ei liitepartikkeliin päättyvien muotojen osalta. Kokonaisuutena tutkielman hypoteesi piti siis paikkaansa ja täten voisi todeta, että syntagmaattisten sulaumien osalta ruututekstien kieli eroaa muista suomen varianteista. Sulaumamuotoja voidaan siis pitää niin sanotun tekstittäjän suomen yhtenä erityispiirteinä.

Osaltaan tämän tutkimuksen lopputulos tukee myös Sonja Tirkkonen-Conditin ja Jukka Mäkisaloon tutkimusta (2007, käsitelty tarkemmin alaluvussa 4.2), jonka lopputuloksen mukaan ruututekstien kieli eroaa muista suomen varianteista käyttämällä hyväkseen suppeaa ja ytimekästä kielenkäyttöä hyödyntäviä koheesiokeinoja. Samalla voi todeta, että Andrew Chestermanin suhdenormi vaikuttaa ruututestikääntämisessä enemmän kuin lukijoiden odotusnormi kaikkien suomen varianttien kielellisestä yhteneväisyydestä. Tietenkään nämä

kaksi normia eivät poissulje toisiaan, tämän tutkimuksen puitteissa suhdennormi kuitenkin menee tarvittaessa ensisijalle. Eniten taustalla varmasti vaikuttavat ruututekstikäntämisen ainutlaatuiset rajoitteet ja konventiot. Kuitenkin on syytä muistaa, että osakorpuksen materiaaleista valtaosa on elokuvadialogia, eli puhetta. Täten voi myös olettaa, että ainakin osittain kääntäjien ratkaisuun valita syntagmaattinen sulauma on vaikuttanut yleispuhekieli, jonka avulla on pyritty saamaan tekstitykseen puhekielimäisyyttä. Tietävästi tutkimusta syntagmaattisten sulaumien esiintymisestä kirjoitetussa kielessä verrattuna yleispuhekieleen ei ole tehty, mutta intuitiivisena ratkaisuna tällä kokemuksella, mitä itselläni on ruututekstikäntämisestä, käyttäisin itse enemmän syntagmaattista sulaumaa kuin sen erilleen kirjoitettua muotoa puhemaisuutta tavoitellessani.

Jälkikäteen, tutkimuksen suoritettuani olisin ehkäpä keskittynyt tutkimuksessa enemmän kolmen korpuksen vertailuun pienemmissäkin asioissa ja jättänyt osakorpuksen sisäisen tutkimuksen vähemmälle. Siltikin olen edelleen sitä mieltä, että ilman selvitystä siitä, miten paljon partikkelit esiintyvät aineistossa kieltoverbin rinnalla sulautumatta niihin, olisi tutkimus ollut vajavainen. Monia asioita olisi voinut tutkia tarkemminkin, joskin se olisi tuottanut liikaa työtä tähän tutkimukseen. Kiinnostavaa olisi kaiken kaikkiaan tutkia, vaikuttavatko syntagmaattiset sulaumat jotenkin ruututekstien vastaanottajiin, eli katsojiin. Vastaanottotutkimuksen (kts. lisää esimerkiksi Tuominen 2012) avulla voisi selvittää, huomaavatko katsojat näitä muotoja tietoisesti, etenkin niiden sulaumamuotojen osalta, jotka ovat muissa suomen varianteissa harvinaisempia, ja vaikuttavatko ne jotenkin katsojien katselukokemukseen ja tekstitysten ymmärtämiseen. Toinen keino olisi tutkia silmänliiketutkimuksen avulla (kts. lisää Lång, Mäkisalo, Gowases & Piettinen 2013), tarttuvatko syntagmaattiset sulaumat ja etenkin niiden harvinaisemmat muodot niin sanotusti silmään, eli ovatko ne katsojien alitajuntaisen reaktion mukaan vaikeammin luettavia kuin samojen kielellisten elementtien erilleen kahtena sanana kirjoitetut muodot – kuuluisivatko siis tällaiset harvinaisemmat muodot silmänliikekameralla tutkittuihin tekstitysten tiedostamattomiin häiriötekijöihin. Toki on myös huomautettava, että tämän tutkimuksen materiaalina oli ainoastaan television tekstityksiä. Kiinnostavaa olisi tietää, olisivatko tulokset samankaltaisia, jos samaa tutkimuskysymystä tarkasteltaisiin esimerkiksi elokuvateatterikäännöksistä tai DVD-suomennöksistä, joissa merkkimäärän rajoite on tv-tekstityksiä pienempi.



# LÄHTEET

## Tutkimusaineisto

KSK - Käännössuomen korpus

YLE-ruututekstikorpus

## Kirjallisuuslähteet

- Aaltonen, Sirkku 2004. Kun Antto Puuronen Suomeen muutti: kulttuurisidonnainen käännöstutkimus työvälineenä. Teoksessa Oittinen, Riitta & Pirjo Mäkinen (toim.). 388–406.
- Agost, Rosa 2011. Audiovisual Translation: A Complex and Unstable Field of Research at the Service of All. *International Journal of Translation* Vol. 23, No. 2, Jul-Dec 2011. New Delhi: Bahri Publications.
- Anderman, Gunilla & Margaret Rogers (toim.) 2008a. *Incorporating Corpora - The Linguist and the Translator*. Clevedon: Multilingual Matters.
- Anderman, Gunilla & Margaret Rogers 2008b. The Linguist and the Translator. Teoksessa Anderman & Rogers (toim.), *Incorporating Corpora - The Linguist and the Translator*. Clevedon: Multilingual Matters. 5–17.
- Austermühl, Frank 2001. *Electronic Tools for Translators*. Manchester: St. Jerome Publishing. 124–132.
- Bowker, Lynn & Jennifer Pearson 2002. *Working with Specialized Language: A practical guide to using corpora*. London: Routledge.
- BTI 2011. *AV-kääntämisen ohjeet II*. [yrityksen sisäinen ohje av-kääntäjille] Lähetetty 25.10.2011. Vastaanottaja: Heli-Maaria Koivikko. [Luettu 25.10.2011.]
- Chesterman, Andrew 1997. *Memes of Translation - The Spread of Ideas in Translation Theory*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Chesterman, Andrew 2007. Kääntämisen normit. Teoksessa Riikonen H.K., Urpo Kovala, Pekka Kujamäki & Outi Paloposki (toim.) 2007. *Suomennoskirjallisuuden historia 2*. Jyväskylä: Gummerus.
- Díaz Cintas, Jorge & Aline Remael 2006. *Audiovisual Translation: Subtitling*. Manchester: St. Jerome Publishing.
- de Linde, Zoé & Neil Kay 1999. *The Semiotics of Subtitling*. Manchester: St. Jerome.
- Eskola, Sari 2002. *Syntetisoivat rakenteet käännössuomessa: suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmällä*. Joensuun yliopiston humanistisia julkaisuja 30. Joensuun yliopisto.
- Gambier, Yves 2008. Audiovisuaalisen kääntämisen tutkimuksen suuntaviivoja. Teoksessa Oittinen, Riitta & Tiina Tuominen (toim.). 73–115.
- Hermans, Theo 1996. Norms and the Determination of Translation: A Theoretical Framework. Teoksessa Álvarez, Román & M. Carmen-África Vidal (toim.), *Translation, Power, Subversion*. Clevedon: Multilingual Matters Ltd. 25–51.

- Hermans, Theo 1997. Translation as institution. Teoksessa Snell-Hornby, Mary & Zuzana Jettmarová & Klaus Kaindl (toim.), *Translation as intercultural communication: selected papers from the EST Congress, Prague 1995*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 3–20.
- Hyvärinen, Riitta 2007. Suomentajana Ylessä - ei hullumpi ammatti. *Kielikello* 2/2007. Kotimaisten kielten keskus. Saatavilla: <http://www.kielikello.fi/index.php?mid=2&pid=11&aid=1807&ref=2015>. [Luettu 4.10.2017]
- Ikola, Osmo 1966. Eräistä murteellisista kieltolauseiden konjunktioista. *Virittäjä* 1/1966. Helsinki: Helsingin Liikekirjapaino Oy. 9–15.
- ISK 2005 = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho 2004. *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seuran Toimituksia 950. Helsinki: SKS.
- Ivarsson, Jan 1992. *Subtitling for the Media: a handbook of an art*. Stockholm: Transedit.
- Ivarsson, Jan & Mary Carroll 1998a. *Code of Good Subtitling Practice*. ESIST. Saatavilla: <https://www.esist.org/code-of-good-subtitling-practice/> [Luettu 14.3.2016]
- Ivarsson, Jan & Mary Carroll 1998b. *Subtitling*. Simrishamn: TransEdit.
- Jantunen, Jarmo 2004. *Synonymia ja käännessuomi - Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin*. Joensuun yliopisto, väitöskirja.
- Jantunen, Jarmo & Sari Eskola 2002. Käännessuomi kielivarianttina: syntaktisia ja leksikaalisia erityispiirteitä. *Virittäjä* 2/2002. 184–206.
- Jääskeläinen, Riitta 2008. Av-kääntämisen tutkimus ja tutkimustarpeet Suomessa. Teoksessa Oittinen, Riitta & Tiina Tuominen (toim.). 116–130.
- Katajamäki, Satu 2009. *Vastaanottavan yleisön käsitykset ruututekstien konventioista*. Tampereen yliopisto, pro gradu -tutkielma.
- Kemppanen, Hannu 2007. Avainsanat - korpustyökalu tutkijan ja kääntäjän käyttöön. *Kääntäjä - Översättaren* 3/07. 12.
- Koivikko, Heli-Maaria 2008. *The Finnish Translation Equivalents of the English 3rd Person Singulars: a Case Study on the State of Fear by Michael Crichton*. Joensuun yliopisto, julkaisematon kandidaatin tutkielma.
- Kujamäki, Pekka & Riitta Jääskeläinen 2001. Korpukset kääntäjän apuna. Teoksessa Oittinen, Riitta & Pirjo Mäkinen (toim.). 239–252.
- Larjavaara, Matti 1992. Etteikö kelpaa? *Virittäjä* 1/1992. 119–122.
- Lison, Pierre & Jörg Tiedemann 2016. *OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles*. 10. Language Resources and Evaluation - konferenssiartikkeli. Saatavilla: <http://stp.lingfil.uu.se/~joerg/paper/opensubs2016.pdf>. [Luettu 5.4.2017.]
- Lång, Juha, Jukka Mäkisalo, Tersia Gowases & Sami Piettinen 2013. Using Eye Tracking to Study the Effect of Badly Synchronized Subtitles on the Gaze Paths of Television Viewers. *New Voices in Translation Studies 10*. Seoul International Association for Translation & Intercultural Studies; Centre for Translation and

Textual Studies, Dublin City University. 72-86. Saatavilla:  
<https://www.iatis.org/images/stories/publications/new-voices/Issue10-2013/articles/article-long-2013b.pdf>. [Luettu 24.3.2017.]

- Maamies, Sari 1993. Etteikö. *Kielikello 1/1993*. 16–20.
- Martikainen, Anu-Irina 2013. *Lexicon of Finnish Subtitles. A Study on the Subtitles of Different Television Programme Types*. Itä-Suomen yliopisto, pro gradu -tutkielma.
- Mauranen, Anna 2002. Käännöskielen tutkimusta suurilla aineistoilla. *Kääntäjä - Översättaren* 04/2002. 4–5.
- Mauranen, Anna 2004. Corpora, Universals and Interference. Teoksessa Mauranen, Anna & Pekka Kujamäki (toim.), *Translation Universals*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 65–82.
- MOT Kielitoimiston sanakirja 8.5e Professional* 2017. Kotimaisten kielten tutkimuskeskus ja Kielikone Oy. Saatavilla: <https://mot-kielikone-fi.helios.uta.fi/mot/uta/netmot.exe>. [Luettu 19.3.2017.]
- Mäkinen, Pirjo 2004. Ikinuori lähdeteksti, ikääntyvä kohdeteksti? Teoksessa Oittinen, Riitta & Pirjo Mäkinen (toim.). 407–425.
- Mäkisalo, Jukka 2006. Ruututekstien korpus tutkimuksen työkaluna. *Kääntäjä - Översättaren* 10/06. 18.
- Mäkisalo, Jukka 2008. *Ruututekstit ja YLE-korpus tutkijan työkaluna*. Joensuun yliopiston opetusmateriaalit, Savonlinna. Saatavilla: <http://cc.joensuu.fi/~makisalo/PL/KtS&M/RTxtKnt2.pdf>. [Luettu 10.10.2010.]
- Mäkisalo, Jukka 2017. YLE-korpus. [Henkilökohtainen sähköpostiviesti]. Vastaanottaja: Heli-Maaria Luoma. Lähetetty 18.9.2017. [Luettu 19.9.2017.]
- Mäyränpää, Sari 2010. *Kun av-kääntäjä kohtasi tavallisen katsojan; Gilmoren tyttöjen tekstitysten vastaanottoa*. Tampereen yliopisto, pro gradu -tutkielma.
- Oittinen, Riitta & Pirjo Mäkinen (toim.) 2004. *Alussa oli käännös*. Tampere: Tampere University Press.
- Oittinen, Riitta & Tiina Tuominen (toim.), *Olellaisen äärellä - Johdatus audiovisuaaliseen kääntämiseen*. 2. painos. Tampere: Tampere University Press.
- Olohan, Maeve 2004. *Introducing Corpora in Translation Studies*. New York: Routledge.
- Oxford Dictionaries* 2017. Oxford University Press. Saatavilla: <https://en.oxforddictionaries.com>. [Luettu 19.3.2017.]
- Pedersen, Jan 2011. *Subtitling Norms for Television: An Exploration Focussing on Extralinguistic Cultural References*. Amsterdam: John Benjamins.
- Pelkonen, Tiia 2008. *Line Breaks and Noun Phrases in Subtitling*. Joensuun yliopisto, pro gradu -tutkielma.
- Puurtinen, Tiina 2004. Käännösten hyväksyttävyyys. Teoksessa Oittinen, Riitta & Pirjo Mäkinen (toim.). 82–94.
- Savijärvi, Ilkka 1988. Agricolan kieltolause. Koivusalo, Esko (toim.), *Mikael Agricolan kieli*. Helsinki: Suomalaisen kirjallisuuden seura. 69–93.

- Scott, Mike 2004. *Oxford WordSmith Tools Version 4.0*. Oxford: Oxford University Press.
- Scott, Mike 2008. Developing WordSmith. *International Journal of English Studies* vol. 8 (1). 95–106. Saatavilla: <http://dialnet.unirioja.es/servlet/articulo?codigo=2729155>. [Luettu 5.4.2017.]
- Scott, Mike 2017. WordSmith Tools Manual. Saatavilla: <http://www.lexically.net/downloads/version7/HTML/index.html>. [Luettu 19.03.2017.]
- SDI Media 2011. *Finnish SDI Style Guide*. [yrityksen sisäinen ohje av-kääntäjille] Lähetetty 27.12.2012. Vastaanottaja: Heli-Maaria Luoma. [Luettu 27.12.2012.]
- Sinclair, John 2003. *Reading Concordances*. London: Pearson.
- Teitto, Heli 2010. *Human Referents in Subtitles. A Study on Personal Pronouns and Proper Nouns in Translated and Original Finnish*. Itä-Suomen yliopisto, pro gradu -tutkielma. Saatavilla: [http://epublications.uef.fi/pub/urn\\_nbn\\_fi\\_uef-20110022/urn\\_nbn\\_fi\\_uef-20110022.pdf](http://epublications.uef.fi/pub/urn_nbn_fi_uef-20110022/urn_nbn_fi_uef-20110022.pdf). [Luettu 8.9.2017]
- Tirkkonen-Condit, Sonja 2007. Kunhan katson korpuksesta: yllättäviä havaintoja ruututekstien koheesiokkeinoista. *Kääntäjä - Översättaren* 6/07. 10.
- Tirkkonen-Condit, Sonja 2017. YLE-ruututekstikorpus. Television käännöstekstien digitaalisen tutkimusaineiston koostaminen. Itä-Suomen yliopisto, YLE-korpushankkeen katselusivu. Saatavilla: [https://wiivi.uef.fi/crisyp/dis/\\_/fi/cr\\_redir\\_all/fet/fet/sea?direction=1&id=-29931](https://wiivi.uef.fi/crisyp/dis/_/fi/cr_redir_all/fet/fet/sea?direction=1&id=-29931). [Luettu 10.4.2017.]
- Tirkkonen-Condit, Sonja & Jukka Mäkisalo 2007. Cohesion in subtitles: A corpus-based study. *Across languages and Cultures* 8 (2). Budapest: Akadémiai Kiadó. 221–230.
- Toury, Gideon 2012. *Descriptive Translation Studies – and beyond: Revised edition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Tuominen, Tiina 2012. *The Art of Accidental Reading and Incidental Listening. An empirical study on the viewing of subtitled films*. Tampereen yliopisto, väitöskirja.
- Ukkola, Taina 2014. *Using Propositions as a Tool for Analysing Two Sets of Subtitles: A Case Study of the Film Star Wars Episode IV: A New Hope*. Itä-Suomen yliopisto, pro gradu -tutkielma. Saatavilla: [http://epublications.uef.fi/pub/urn\\_nbn\\_fi\\_uef-20150056/urn\\_nbn\\_fi\\_uef-20150056.pdf](http://epublications.uef.fi/pub/urn_nbn_fi_uef-20150056/urn_nbn_fi_uef-20150056.pdf). [Luettu 20.9.2017]
- Valentini, Christina 2006a. A Multimedia Database for the Training of Audiovisual Translators. *The Journal of Specialised Translation* 6:2006. 68–84. Saatavilla: [http://www.jostrans.org/issue06/art\\_valentini.php](http://www.jostrans.org/issue06/art_valentini.php). [Luettu 19.3.2017.]
- Valentini, Christina 2006b. Developing AVT corpora for a quantitative approach to language transfer in cinematic products. Kirjallinen artikkeli XVIII CETRA Summer Sessions 4.-16.9.2006 -konferenssiesitelmästä. Saatavilla: <https://www.arts.kuleuven.be/cetra/papers/files/valentini.pdf>. [Luettu 28.3.2017]
- Valentini, Christina 2008. Forlì 1 - The Forlì Corpus of Screen Translation: Exploring macrostructures. Teoksessa Chiaro, Heiss & Bucaria (toim.), *Between Text and*

*Image - Updating research in screen translation.* Amsterdam/Philadelphia: John Benjamins Publishing Company. 36–50.

Vertanen, Esko 2008. Ruututeksti tiedon ja tunteiden tulkkina. Teoksessa Oittinen, Riitta & Tiina Tuominen (toim.). 149–170.

Wilkinson, Michael 2006. The corpus analysis tool - an under-exploited translation aid. *Kääntäjä - Översättaren* 7/06. 8–9.

Wilkinson, Michael 2008. *Specialized Corpora as Translation Aids*. Joensuun yliopiston käännöstieteen (englanti) [internetopetusmateriaali erikoiskorpuksset kääntäjän apuna kurssilla]. Saatavilla: [kvl.joensuu.fi/sorvi](http://kvl.joensuu.fi/sorvi). [Luettu 10.3.2008.]

## LIITE I: SYNTAGMAATTISTEN SULAUMIEN HAKUSANALISTA

EHKEI	OK	APS	KS
ehken	63	-	6
ehket	30	-	1
ehkei	660	11	77
ehkemme	14	-	2
ehkette	9	-	-
ehkeivät	26	2	4
<b>kaikki taivutusmuodot</b>	<b>802</b>	<b>13</b>	<b>90</b>

ELLEI	OK	APS	KS
ellen	874	65	96
ellet	996	20	152
ellei	3 441	486	963
ellemme	281	14	55
ellette	485	7	24
elleivät	270	41	127
<b>kaikki taivutusmuodot</b>	<b>6 347</b>	<b>633</b>	<b>1 417</b>

ETTEI	OK	APS	KS
etten	7 487	602	1 035
ettet	4 380	112	317
ettei	23 702	4 036	6 087
etemme	1 139	56	174
ettette	946	9	46
etteivät	2 269	377	614
<b>kaikki taivutusmuodot</b>	<b>39 923</b>	<b>5 192</b>	<b>8 273</b>

JOLLEI	OK	APS	KS
jollen	314	1	29
jollet	401	6	47
jollei	977	53	209
jollemme	90	1	11
jollette	177	-	5
jolleivät	21	1	3
jolleivät	40	2	11
<b>kaikki taivutusmuodot</b>	<b>2 020</b>	<b>64</b>	<b>315</b>

JOTTEI	OK	APS	KS
jotten	48	-	11
jottet	28	-	2
jottei	386	25	85
jottemme	14	1	-
jottette	3	-	-
jotteivät	12	-	-
jotteivät	43	2	16
<b>kaikki taivutusmuodot</b>	<b>534</b>	<b>28</b>	<b>114</b>

MIKSEI	OK	APS	KS
miksen	1 036	31	39
mikset	4 586	74	145
miksei	4 410	179	246
miksemme	188	1	19
miksette	1 061	4	31
mikseivät	277	9	36
<b>kaikki taivutusmuodot</b>	<b>11 558</b>	<b>298</b>	<b>516</b>

MUTTEI	OK	APS	KS
mutten	1 647	24	135
muttet	197	2	5
muttei	1 858	112	317
muttemme	68	-	7
muttette	27	-	-
mutteivät	7	-	-
mutteivät	82	9	22
<b>kaikki taivutusmuodot</b>	<b>3 886</b>	<b>147</b>	<b>486</b>

SIKSEI	OK	APS	KS
siksen	3	-	-
sikset	-	-	-
siksei	13	-	-
siksemme	-	-	-
siksette	-	-	-
sikseivät	1	-	-
<b>kaikki taivutusmuodot</b>	<b>17</b>	<b>-</b>	<b>-</b>

VAIKKEI	OK	APS	KS
vaikken	432	17	48
vaikket	171	6	4
vaikkei	1 257	145	256
vaikkemme	44	4	2
vaikkette	32	-	-
vaikkeivat	7	1	3
vaikkeivät	55	6	19
<b>kaikki taivutusmuodot</b>	<b>1 998</b>	<b>179</b>	<b>332</b>

	OK	APS	KS
kaikki taivutusmuodot	67 085	6 554	11 543

OK = osakorpus; APS = alkuperäissuomi; KS = käännessuomi

Taulukoissa kaikki aineistosta löytyneet syntagmaattiset sulaumat sisältäen myös liitepartikkeliin päättyvät muodot

## LIITE II: ERILLEEN KIRJOITETTujen MUOTOJEN HAKUSANALISTA

<b>EHKÄ + EI</b>	<b>OK</b>	<b>2–3-san.</b>
ehkä + en	615	95
ehkä + et	357	70
ehkä + ei	2 857	296
ehkä + emme	170	13
ehkä + ette	194	46
ehkä + eivät	407	26
<b>kaikki taivutusmuodot</b>	<b>4 600</b>	<b>546</b>

<b>JOS + EI</b>	<b>OK</b>	<b>2–3-san.</b>
jos + en	1 489	177
jos + et	2 080	474
jos + ei	7 319	421
jos + emme	564	62
jos + ette	795	155
jos + eivät	723	17
<b>kaikki taivutusmuodot</b>	<b>12 970</b>	<b>1 306</b>

<b>ETTÄ + EI</b>	<b>OK</b>	<b>2–3-san.</b>
että + en	868	97
että + et	376	52
että + ei	4 872	154
että + emme	408	28
että + ette	470	51
että + eivät	910	6
<b>kaikki taivutusmuodot</b>	<b>7 904</b>	<b>388</b>

<b>JOTTA + EI</b>	<b>OK</b>	<b>2–3-san.</b>
jotta + en	53	14
jotta + et	43	5
jotta + ei	550	7
jotta + emme	44	8
jotta + ette	22	2
jotta + eivät	241	1
<b>kaikki taivutusmuodot</b>	<b>953</b>	<b>37</b>

<b>MIKSI + EI</b>	<b>OK</b>	<b>2–3-san.</b>
miksi + en	372	125
miksi + et	1 006	398
miksi + ei	1 359	515
miksi + emme	176	28
miksi + ette	687	198
miksi + eivät	201	31
<b>kaikki taivutusmuodot</b>	<b>3 801</b>	<b>1 295</b>



<b>MUTTA + EI</b>	<b>OK</b>	<b>2–3-san.</b>
mutta + en	6 948	1 329
mutta + et	1 293	279
mutta + ei	17 864	1 989
mutta + emme	1 115	136
mutta + ette	491	90
mutta + eivät	2 114	207
<b>kaikki taivutusmuodot</b>	<b>29 825</b>	<b>4 030</b>

<b>SIKSI + EI</b>	<b>OK</b>	<b>2–3-san.</b>
siksi + en	286	21
siksi + et	35	1
siksi + ei	431	2
siksi + emme	59	1
siksi + ette	16	1
siksi + eivät	52	1
<b>kaikki taivutusmuodot</b>	<b>879</b>	<b>27</b>

<b>VAIKKA + EI</b>	<b>OK</b>	<b>2–3-san.</b>
vaikka + en	345	42
vaikka + et	141	23
vaikka + ei	1 137	38
vaikka + emme	108	7
vaikka + ette	64	13
vaikka + eivät	204	4
<b>kaikki taivutusmuodot</b>	<b>1 999</b>	<b>127</b>

	<b>OK</b>	<b>2–3-san.</b>
kaikki taivutusmuodot	<b>62 931</b>	<b>7 756</b>

OK = osakorpus; 2–3-san. = kaksi- tai kolmisanaiset lauseet

Oikeanpuoleisessa sarakkeessa erilleen kahtena sanana kirjoitettujen muotojen esiintymät kaksi- tai kolmisanaisissa, ns. lyhyissä lauseissa

### LIITE III: SYNTAGMAATTISTEN SULAUMIEN FREKVENSsit

EHKEI	OK	APS	KS
ehken	2,4	-	1,0
ehket	1,1	-	0,2
ehkei	24,9	2,9	13,1
ehkemme	0,5	-	0,3
ehkette	0,3	-	-
ehkeivät	1,0	-	0,7
<b>kaikki taivutusmuodot</b>	<b>30,3</b>	<b>3,4</b>	<b>15,3</b>

ELLEI	OK	APS	KS
ellen	33,0	17,0	16,3
ellet	37,6	5,2	25,8
ellei	130,0	126,6	163,6
ellemme	10,6	3,7	9,4
ellette	18,3	1,8	4,1
elleivät	10,2	10,7	21,6
<b>kaikki taivutusmuodot</b>	<b>239,8</b>	<b>165,0</b>	<b>240,8</b>

ETTEI	OK	APS	KS
etten	280,8	154,0	174,4
ettet	165,2	29,2	53,4
ettei	888,7	1 033,9	1 017,8
etemme	42,7	14,4	28,7
ettette	35,7	2,3	7,8
etteivät	85,2	96,1	102,2
<b>kaikki taivutusmuodot</b>	<b>1 498,3</b>	<b>1 329,9</b>	<b>1 384,4</b>

JOLLEI	OK	APS	KS
jollen	11,9	0,3	4,9
jollet	15,2	1,6	8,0
jollei	36,9	13,8	35,4
jollemme	3,4	0,3	1,9
jollette	6,7	-	0,9
jolleivat	0,8	0,3	0,5
jolleivät	1,5	0,5	1,9
<b>kaikki taivutusmuodot</b>	<b>76,4</b>	<b>16,7</b>	<b>53,4</b>

JOTTEI	OK	APS	KS
jotten	1,8	-	1,9
jottet	1,1	-	0,3
jottei	14,6	6,5	14,5
jottemme	0,5	0,3	-
jottette	0,1	-	-
jotteivat	0,5	-	-
jotteivät	1,6	0,5	2,7
<b>kaikki taivutusmuodot</b>	<b>20,2</b>	<b>7,3</b>	<b>19,4</b>

MIKSEI	OK	APS	KS
miksen	39,1	8,1	6,6
mikset	173,3	19,1	24,5
miksei	166,5	46,2	41,7
miksemme	7,1	0,3	3,2
miksette	40,1	1,0	5,3
mikseivät	10,5	2,3	6,1
<b>kaikki taivutusmuodot</b>	<b>436,6</b>	<b>77,0</b>	<b>87,4</b>

MUTTEI	OK	APS	KS
mutten	62,2	6,3	23,0
muttet	7,4	0,5	0,9
muttei	70,2	29,2	53,9
muttemme	2,6	-	1,2
muttette	1,0	-	-
mutteivät	0,3	-	-
mutteivät	3,1	2,3	3,7
<b>kaikki taivutusmuodot</b>	<b>146,8</b>	<b>38,4</b>	<b>82,6</b>

SIKSEI	OK	APS	KS
siksen	0,1	-	-
sikset	-	-	-
siksei	0,5	-	-
siksemme	-	-	-
siksette	-	-	-
sikseivät	0,0	-	-
<b>kaikki taivutusmuodot</b>	<b>0,6</b>	<b>-</b>	<b>-</b>

VAIKKEI	OK	APS	KS
vaikken	16,3	4,4	8,2
vaikket	6,5	1,6	0,7
vaikkei	47,5	37,9	43,5
vaikemme	1,7	1,0	0,3
vaikkette	1,2	-	-
vaikkeivat	0,3	0,3	0,5
vaikkeivät	2,1	1,6	3,2
<b>kaikki taivutusmuodot</b>	<b>75,5</b>	<b>46,7</b>	<b>56,4</b>

	OK	APS	KS
kaikki taivutusmuodot	2 524,5	1 684,5	1 939,7

OK = osakorpus; APS = alkuperäissuomi; KS = käännessuomi

Suhteelliset frekvenssiluvut on laskettu ilman liitepartikkeliin päättyviä muotoja

## LIITE IV: ERILLEEN KIRJOITETTUIJEN MUOTOJEN SUHTEELLISET FREKVENSsit

<b>EHKÄ + EI</b>	<b>OK</b>	<b>ETTÄ + EI</b>	<b>OK</b>
ehkä + en	23,2	että + en	32,8
ehkä + et	13,5	että + et	14,2
ehkä + ei	11,2	että + ei	184,1
ehkä + emme	6,4	että + emme	15,4
ehkä + ette	7,3	että + ette	17,8
ehkä + eivät	15,4	että + eivät	34,4
<b>kaikki taivutusmuodot</b>	<b>173,8</b>	<b>kaikki taivutusmuodot</b>	<b>298,7</b>

<b>JOS + EI</b>	<b>OK</b>	<b>JOTTA + EI</b>	<b>OK</b>
jos + en	56,3	jotta + en	2,0
jos + et	78,6	jotta + et	1,6
jos + ei	276,6	jotta + ei	20,8
jos + emme	21,3	jotta + emme	1,7
jos + ette	30,0	jotta + ette	0,8
jos + eivät	27,3	jotta + eivät	9,1
<b>kaikki taivutusmuodot</b>	<b>490,1</b>	<b>kaikki taivutusmuodot</b>	<b>36,0</b>

<b>MIKSI + EI</b>	<b>OK</b>	<b>MUTTA + EI</b>	<b>OK</b>
miksi + en	14,1	mutta + en	262,5
miksi + et	38,0	mutta + et	48,9
miksi + ei	51,4	mutta + ei	675,0
miksi + emme	6,7	mutta + emme	42,1
miksi + ette	26,0	mutta + ette	18,6
miksi + eivät	7,6	mutta + eivät	79,9
<b>kaikki taivutusmuodot</b>	<b>143,6</b>	<b>kaikki taivutusmuodot</b>	<b>1 127,0</b>

<b>SIKSI + EI</b>	<b>OK</b>	<b>VAIKKA + EI</b>	<b>OK</b>
siksi + en	10,8	vaikka + en	13,0
siksi + et	1,3	vaikka + et	5,3
siksi + ei	16,3	vaikka + ei	43,0
siksi + emme	2,2	vaikka + emme	4,1
siksi + ette	0,6	vaikka + ette	2,4
siksi + eivät	2,0	vaikka + eivät	7,7
<b>kaikki taivutusmuodot</b>	<b>33,2</b>	<b>kaikki taivutusmuodot</b>	<b>75,5</b>

	<b>OK</b>	<b>2-3 san.</b>
kaikki taivutusmuodot	2 378,0	29,31

## LIITE V: SYNTAGMAATTISTEN SULAUMIEN PROSENTUAALISET ESIINTYMÄT LYHYISSÄ LAUSEISSA

lyhyet muodot (%:a  
kaikista esiintymistä)

EHKEI	OK
ehken	19 %
ehket	10 %
ehkei	14 %
ehkemme	7 %
ehkette	11 %
ehkeivät	-
<b>kaikki taivutusmuodot</b>	<b>13 %</b>

ELLEI	OK
ellen	16 %
ellet	17 %
ellei	7 %
ellemme	6 %
ellette	17 %
elleivät	7 %
<b>kaikki taivutusmuodot</b>	<b>10 %</b>

ETTEI	OK
etten	14 %
ettet	16 %
ettei	5 %
ettemme	9 %
ettette	15 %
etteivät	3 %
<b>kaikki taivutusmuodot</b>	<b>8 %</b>

JOLLEI	OK
jollen	16 %
jollet	21 %
jollei	8 %
jollemme	10 %
jollette	21 %
jolleivat	5 %
jolleivät	15 %
<b>kaikki taivutusmuodot</b>	<b>13 %</b>

1-sanaiset muodot  
(%:a 1–2-sanaisista)

EHKEI	OK
ehken	-
ehket	-
ehkei	5 %
ehkemme	-
ehkette	-
ehkeivät	-
<b>kaikki taivutusmuodot</b>	<b>5 %</b>

ELLEI	OK
ellen	-
ellet	2 %
ellei	22 %
ellemme	-
ellette	5 %
elleivät	5 %
<b>kaikki taivutusmuodot</b>	<b>9 %</b>

ETTEI	OK
etten	2 %
ettet	2 %
ettei	3 %
ettemme	-
ettette	-
etteivät	3 %
<b>kaikki taivutusmuodot</b>	<b>2 %</b>

JOLLEI	OK
jollen	4 %
jollet	1 %
jollei	5 %
jollemme	-
jollette	-
jolleivat	-
jolleivät	-
<b>kaikki taivutusmuodot</b>	<b>3 %</b>

JOTTEI	OK
jotten	10 %
jottet	25 %
jottei	2 %
jottemme	7 %
jottette	-
jotteivat	-
jotteivät	2 %
<b>kaikki taivutusmuodot</b>	<b>4 %</b>

JOTTEI	OK
jotten	-
jottet	-
jottei	-
jottemme	-
jottette	-
jotteivat	-
jotteivät	-
<b>kaikki taivutusmuodot</b>	<b>-</b>

MIKSEI	OK
miksen	37 %
mikset	30 %
miksei	31 %
miksemme	11 %
miksette	23 %
mikseivät	10 %
<b>kaikki taivutusmuodot</b>	<b>30 %</b>

MIKSEI	OK
miksen	55 %
mikset	43 %
miksei	74 %
miksemme	45 %
miksette	37 %
mikseivät	55 %
<b>kaikki taivutusmuodot</b>	<b>57 %</b>

MUTTEI	OK
mutten	28 %
muttet	30 %
muttei	32 %
muttemme	21 %
muttette	30 %
mutteivat	29 %
mutteivät	23 %
<b>kaikki taivutusmuodot</b>	<b>30 %</b>

MUTTEI	OK
mutten	1 %
muttet	-
muttei	0 %
muttemme	-
muttette	-
mutteivat	-
mutteivät	-
<b>kaikki taivutusmuodot</b>	<b>0 %</b>

SIKSEI	OK
siksen	100 %
sikset	-
siksei	85 %
siksemme	-
siksette	-
sikseivät	-
<b>kaikki taivutusmuodot</b>	<b>82 %</b>

SIKSEI	OK
siksen	100 %
sikset	-
siksei	100 %
siksemme	-
siksette	-
sikseivät	-
<b>kaikki taivutusmuodot</b>	<b>100 %</b>

VAIKKEI	OK
vaikken	16 %
vaikket	18 %
vaikkei	8 %
vaikkemme	-
vaikkette	31 %
vaikkeivat	-
vaikkeivät	5 %
<b>kaikki taivutusmuodot</b>	<b>10 %</b>

VAIKKEI	OK
vaikken	1 %
vaikket	-
vaikkei	1 %
vaikkemme	-
vaikkette	-
vaikkeivat	-
vaikkeivät	-
<b>kaikki taivutusmuodot</b>	<b>1 %</b>

## LIITE VI: ERILLEEN KIRJOITETTUIJEN MUOTOJEN PROSENTUAALISET ESIINTYMÄT LYHYISSÄ LAUSEISSA

lyhyet muodot yhteensä  
(%:a kaikista esiintymistä)

EHKÄ + EI	OK
ehkä + en	85 %
ehkä + et	19 %
ehkä + ei	10 %
ehkä + emme	8 %
ehkä + ette	24 %
ehkä + eivät	6 %
<b>kaikki taivutusmuodot</b>	<b>21 %</b>

2-sanaiset muodot  
(%:a 2–3-sanaisten esiintymistä)

EHKÄ + EI	OK
ehkä + en	8 %
ehkä + et	23 %
ehkä + ei	47 %
ehkä + emme	46 %
ehkä + ette	17 %
ehkä + eivät	58 %
<b>kaikki taivutusmuodot</b>	<b>23 %</b>

JOS + EI	OK
jos + en	12 %
jos + et	23 %
jos + ei	6 %
jos + emme	11 %
jos + ette	19 %
jos + eivät	2 %
<b>kaikki taivutusmuodot</b>	<b>10 %</b>

JOS + EI	OK
jos + en	5 %
jos + et	5 %
jos + ei	18 %
jos + emme	5 %
jos + ette	8 %
jos + eivät	35 %
<b>kaikki taivutusmuodot</b>	<b>10 %</b>

ETTÄ + EI	OK
että + en	11 %
että + et	14 %
että + ei	3 %
että + emme	7 %
että + ette	11 %
että + eivät	1 %
<b>kaikki taivutusmuodot</b>	<b>5 %</b>

ETTÄ + EI	OK
että + en	23 %
että + et	17 %
että + ei	38 %
että + emme	11 %
että + ette	4 %
että + eivät	50 %
<b>kaikki taivutusmuodot</b>	<b>25 %</b>

JOTTA + EI	OK
jotta + en	26 %
jotta + et	12 %
jotta + ei	1 %
jotta + emme	18 %
jotta + ette	9 %
jotta + eivät	0 %
<b>kaikki taivutusmuodot</b>	<b>4 %</b>

JOTTA + EI	OK
jotta + en	0 %
jotta + et	0 %
jotta + ei	0 %
jotta + emme	0 %
jotta + ette	0 %
jotta + eivät	0 %
<b>kaikki taivutusmuodot</b>	<b>0 %</b>

MIKSI + EI	OK	MIKSI + EI	OK
miksi + en	34 %	miksi + en	49 %
miksi + et	40 %	miksi + et	62 %
miksi + ei	38 %	miksi + ei	89 %
miksi + emme	16 %	miksi + emme	54 %
miksi + ette	29 %	miksi + ette	59 %
miksi + eivät	15 %	miksi + eivät	74 %
<b>kaikki taivutusmuodot</b>	<b>34 %</b>	<b>kaikki taivutusmuodot</b>	<b>71 %</b>

MUTTA + EI	OK	MUTTA + EI	OK
mutta + en	19 %	mutta + en	1 %
mutta + et	22 %	mutta + et	1 %
mutta + ei	11 %	mutta + ei	7 %
mutta + emme	12 %	mutta + emme	1 %
mutta + ette	18 %	mutta + ette	3 %
mutta + eivät	10 %	mutta + eivät	0 %
<b>kaikki taivutusmuodot</b>	<b>14 %</b>	<b>kaikki taivutusmuodot</b>	<b>4 %</b>

SIKSI + EI	OK	SIKSI + EI	OK
siksi + en	7 %	siksi + en	5 %
siksi + et	3 %	siksi + et	0 %
siksi + ei	0 %	siksi + ei	50 %
siksi + emme	2 %	siksi + emme	0 %
siksi + ette	6 %	siksi + ette	0 %
siksi + eivät	2 %	siksi + eivät	0 %
<b>kaikki taivutusmuodot</b>	<b>3 %</b>	<b>kaikki taivutusmuodot</b>	<b>7 %</b>

VAIKKA + EI	OK	VAIKKA + EI	OK
vaikka + en	12 %	vaikka + en	0 %
vaikka + et	16 %	vaikka + et	4 %
vaikka + ei	3 %	vaikka + ei	0 %
vaikka + emme	6 %	vaikka + emme	0 %
vaikka + ette	20 %	vaikka + ette	0 %
vaikka + eivät	2 %	vaikka + eivät	0 %
<b>kaikki taivutusmuodot</b>	<b>6 %</b>	<b>kaikki taivutusmuodot</b>	<b>1 %</b>



## ENGLISH SUMMARY

### Contracted particle forms in subtitles

#### Introduction

The purpose of this study was to examine how the Finnish negation verb *ei* forms syntagmatic fusions with particles. The primary aim of the study was to observe how these syntagmatic fusions of particles and the negation verb appear in Finnish subtitles and to assess how the use of these syntagmatic fusions of particles in Subtitled Finnish compares to that of written Finnish, both Original Finnish and Translated Finnish. The secondary purpose of the study was to compare how the use of these contracted forms compares to that of the non-contracted forms of the same particles alongside the negation verb in subtitles. The hypothesis of this study was that the syntagmatic fusions of particles and the negation verb occur more frequently in Finnish subtitles than they do in Translated Finnish and Original Finnish due to the constraints of audiovisual translation.

A syntagmatic fusion (or *syntagmaattinen sulauma* in Finnish) is a combination of two word stems fused together, and it inflects for person and has other word like characteristics. These fusion forms are not compound words, but still the different parts of the fusions have their own roles in sentences (ISK 2005, 163–164). These syntagmatic fusions also have word-like qualities: they are pronounced and perceived as a single word form, they can have two variants in third-person plural form (*-eivat* and *-eivät*), they can't have infixes, only suffixes and such others (ISK 2005, 166). In particular, the syntagmatic fusions of the negation verb and the Finnish conjunctions *että* 'that', *jos* 'if', *jotta* 'in order that', *mutta* 'but' and *vaikka* 'although' as well as the particles *ehkä* 'maybe', *miksi* 'why' and *siksi* 'therefore' were selected as the topic of this thesis – the latter because they behave similarly to the above mentioned conjunctions when combined with the negation verb (ISK 2005, 164). As far as is known, no-one has ever before studied these syntagmatic fusions of particles and the negation verb with regard to subtitles.

As subtitling is a special form of translation which is limited by the time and space a single subtitle line can be on screen and Finnish tends to have long words, the use of contracted

forms seems like a good way to save characters. But while working as a subtitler I have often been wondering whether these forms are actually easy to read or in fact, frequent in other variants of Finnish. After all, with the fleeting nature of subtitles the viewer only has so little time to grasp the intended meaning of a subtitle line (Vertanen 2008, 150–151).

The theoretical background for the assumption mentioned above is adapted from Andrew Chesterman's (1997) norm theory. The emphasis of the study was in his expectancy norms and relation norms (explained further in detail). The underlying assumption of this thesis was that in case the contracted particle forms occur as frequently in Subtitled Finnish as they do in Original Finnish and Translated Finnish, then the subtitles could be claimed to conform with the expectancy norm of the audience – that the subtitles use the same kind of language as is used in other variants of Finnish. In case these contracted forms occur more frequently in Subtitled Finnish than they do in the other variants, then the subtitles could be asserted to rather conform with Chesterman's relation norms. After all, the subtitler is forced to work around the different constraints and conventions of audiovisual translation. The assumption is this study was that due to the above mentioned constraints, the influence of the relation norm would be greater on the translator than that of the expectancy norm should these two clash.

The research method of this study was corpus linguistics and the primary research material used in this study came from the *Finnish Broadcasting Company corpus of subtitles*, and to be exact, from a smaller subcorpus compiled from the materials of the previously mentioned corpus. The *Corpus of Original and Translated Finnish* was used as the comparison material for the comparison between contracted particle forms in Subtitled Finnish and in written Finnish.

## Theoretical background

Transforming an audiovisual entity into Finnish subtitles is a highly specialised form of translation with unique constraints and conventions. In translation studies the concept of norm goes hand in hand with descriptive translation studies. One of the purposes of descriptive translation studies is in fact to find norms which affect translation and direct the translators' decisions (Eskola 2002, 38). According to Theo Hermans (1997, 7), "translation is... a matter of communication" and every aspect of this communication requires a set of answers to different types of problems and norms as well as conventions are the kind of answers required

in this type of interpersonal communication. He also states that "if *this* is the feature displayed in our text, then *that* must have been the norm that triggered it" (op. cit.). Gideon Toury (2012, 63) on the other hand says that norms are "the translation of general values or ideas shared by a community – as to what would count as right or wrong, adequate or inadequate – into performance ‘instructions’ appropriate for and applicable to concrete situations. These ‘instructions’ specify what is prescribed and forbidden, as well as what is tolerated and permitted in a certain behavioural dimension." As mentioned earlier, this thesis complies with the notions of Andrew Chesterman, who divides norms into two main categories: product norms and process norms. As the product norms are mainly based on the expectancies of the target language readers, Chesterman calls these types of norms as expectancy norms (Chesterman 1997, 64). As for process norms, Chesterman states that "they are subordinate to the expectancy, because they are themselves determined by the expectancy norms" (Chesterman 1997, 67). He also divides process norms into three subcategories: accountability norms, communications norms and relation norms. As the hypothesis of this thesis revolves around expectation norms and relation norms only, the other categories are not concerned here (for further information about them cf. Chesterman 1997).

As regards to expectancy norms, Chesterman (1997, 64) states the following: "Expectancy norms are established by the expectations of readers of a translation (of a given type) concerning what a translation (of this type) should be like." The readers can have all kinds of expectations on translations, e.g. about text-type, style and register, collocations, lexical choices etc (ibid.). When it comes to relation norms, Chesterman clarifies his notions by stating that "a translator should act in such a way that an appropriate relation of relevant similarity is established and maintained between the source text and the target text" (Chesterman 1997, 69). He further explains, that "it is left for the translator to decide what kind of relation is appropriate in any given case" (loc. cit.).

As was already mentioned, subtitling is a form of translation that is governed by different constraints and conventions which vary according to medium, e.g. TV, DVD, cinema (Gambier 2008, 87), and according to countries (Pedersen 2011). The most important constraints are the time and space allotted for a single subtitle line (Vertanen 2008, 151). The most important thing is that the subtitle line will remain on screen for the duration of the speech one hear from the audiotrack (op. cit., p. 152). The reading speed of an average viewer

also affects how long a line must stay on screen, and the more complex the syntax and the lexicon of the subtitle line, the longer it must stay visible (Ivarsson 1992, 38). One could also claim that the viewers of a given programme are also a constraint as the subtitler might have to e.g. re-evaluate the reading speed of the average reader in programmes directed to the elderly or children (Pedersen 2011, 20). The conventions of television subtitling in Finland include lining the subtitles to the left side of the screen, by placing a hyphen in the end of a subtitle line to indicate that the speech continues in the next subtitle and italics are used when the speaker is out of sight, in flashbacks, quotations or when using foreign language elements in the subtitles (Vertanen 2008, 154–156), to name a few.

## Research questions

As mentioned earlier, the purpose of this thesis was to study how the Finnish negation verb *ei* forms syntagmatic fusions with particles and to look into how these syntagmatic fusions of particles and the negation verb appear in Finnish subtitles. The primary aim is to study how these syntagmatic fusions of particles and the negation verb appear in Finnish subtitles and to assess how the use of these syntagmatic fusions of particles in subtitles compares to that of written Finnish, both Original Finnish and Translated Finnish. The secondary purpose of the study was to compare how the use of these contracted forms compares to that of the non-contracted forms of the same particles alongside the negation verb in Subtitled Finnish. The hypothesis of this study was that the syntagmatic fusions of particles and the negation verb occur more frequently in Finnish subtitles than they do in Translated Finnish and Original Finnish. Finns seem to have a strong opinions on the subtitles they read (cf. Tuominen 2012) and yet they seem to have little knowledge of the conventions and constraints of the audiovisual translation (Katajamäki 2009). Therefore one could assume that the Finns await that the language used in Finnish subtitles complies to the norms of other variants of Finnish. In this thesis the notion of expectancy norm is applied to these expectations viewers are likely to have – that the use of syntagmatic fusions of particles and the negation verb complies with that of Original Finnish and Translated Finnish. Yet, due to the constraints of audiovisual translation, the translator is also likely to have to use whatever means necessary to condense the message of the speech on screen to a sensible subtitle line that conveys the message to the viewers. Therefore one could assume that the Finnish subtitlers decisions conform with Chestermans' relation norm more than they do with the expectation norm (when applied to the expectations of viewers about the exact similarity of different variants of Finnish with regards

to the syntagmatic fusions of particles and the negation verb), which in fact is the hypothesis of this thesis.

## Materials and method

The primary research material used in this study came from the *Finnish Broadcasting Company corpus of subtitles* (or the FBC or *YLE-ruututekstikorpus* in Finnish), which is among the largest corpora in world and as far as is known the largest one that consists of subtitles by professional translators. The FBC corpus was compiled in the former University of Joensuu (nowadays University of Eastern Finland) in 2005 after the Finnish Broadcasting Company YLE donated all of the subtitles that were aired in the channels of the company during 1991 to 2004. The corpus comprises almost 52,500 television programmes and films containing over 99 million words. (Mäkisalo 2006, 18.) As mentioned by Sonja Tirkkonen-Condit and Jukka Mäkisalo (2007, 223), Finland tends to favour subtitles to dubbing so most foreign and imported television programmes are subtitles and hence the FBC corpus has a wide range of genres, from television series and documents to news reports and sports programmes. The corpus also has a variety of source languages from Arabic to Vietnam, although English is the most common one – therefore the FBC is a multilingual comparable corpus (Mäkisalo 2006, 18; 2008). Also, three separate versions of the corpus were compiled: one containing the Original .890 format files that can be run through a subtitling programme with the Original videos, one containing converted subtitle files which includes time codes and additional information and one containing only the subtitle line. The second version of the corpus was selected as the material, as early on before any actual research was conducted, the aim of this study was to also view time codes in order to be able to determine if time had been a factor in choosing a contracted form or a non-contracted form.

The materials of the *Corpus of Original and Translated Finnish* (or the COTF, or *Käännösuomen korpus* in Finnish) was used as the comparison material for the comparison between contracted particle forms in Subtitled Finnish and in written Finnish. The Corpus of Original and Translated Finnish was also compiled in the former University of Joensuu between 1997 and 2000 (Mauranen 2004, 72). It is a monolingual comparable corpus comprising of three separate subcorpora: one which contains texts written Originally in Finnish, one which contains text translated into Finnish from several languages and one that contains text translated into Finnish solely from English (Jantunen & Eskola 2002, 186). In

this study the latter two were combined together. Altogether the COTF contains over ten million words of which about four million is in Original Finnish and about six million in Translated Finnish (Mauranen 2002). All three subcorpora include complete, published works from the following genres: literary texts, academic texts, children's fiction, popular fiction, popular non-fiction, entertainment and biography. All of the translations in the COTF were published between 1995 and 1999 and are translations by professional translators. (Mauranen 2004, 73-74.)

As the Original FBC corpus is so large, the size of the research material needed to be reduced, so a separate subcorpus was compiled from the subtitles that were aired between 2000 and 2004. All of the files from these years were collected into a single folder thereby removing all the re-run files as they were named with the same programme number. Then every file that contained more than three subtitle lines of Finnish were cleaned of all foreign language lines and the ones that contained less were removed altogether, so that the subcorpus would only include Finnish subtitle lines. At that point, all of the files still included time codes and general information about the files and these needed to be removed manually. The resulting cleared up subcorpus contained over 26,000 words and over 14,000 programme files.

Then the materials of the subcorpus and the Corpus of Original and Translated Finnish were separately analysed with the help of *WordSmith Tools*, which is a corpus analysis tool created by Michael Scott (Scott 2017). Every particle was entered into the corpus programme with each of its forms, both in contracted forms (such as *miksen*) as well as non-contracted forms alongside the inflected negation verb as a context word (e.g. *miksi + en*). Separate concordance lists were saved, one for a each single form, and these lists were cleaned of the material not suited for this study (e.g. misspelled words, search word and context word in separate clauses, the name *Ellen* when searching the first-person singular *ellen* etc.). The resulting Key-Word-In-Context (or the KWIC) view can be organised in different ways in order to examine the results, e.g. by re-sorting the view according to R1, i.e. the word directly to the right of the search word (Wilkinson 2006). The lists were first sorted according to the search word in question in order to find the forms which ended to a clitic as those forms were saved to a separate list. Secondly the lists were sorted by the immediate context to right of the search word (R1, R2 and R3) so that the alphabetical list of words could be used to examine the collocates, i.e. "the words which occur in the neighbourhood of your search word" (Scott

2017) – the most common verbs accompanying the search words and whether personal pronouns follow the inflected particle forms. Thirdly the lists were resorted according to L1, L2 and L3 (the words left of the search word) to examine where the particles are positioned in clauses, especially *ehkä* which can be situated behind the negation verb in non-contracted forms, unlike with the other particles.

The number of occurrences for each form was then calculated as well as tabulated and calculations were made on how many times each form appeared in its corpus within a group of a million words, resulting in a relative frequency number for each form. As the subtitle subcorpus, the Corpus of Original Finnish and the Corpus of Translated Finnish were of different sizes, just the number of occurrences gives no information on how the occurrences of a certain form in subtitles corresponds to that of Translated Finnish. The first comparison to be made was between the contracted forms and non-contracted forms which occurred in the subtitle subcorpus. Then the occurrences of the contracted forms in the subcorpus were compared against that of the Corpus of Original Finnish and that of the Corpus of Translated Finnish. All of the comparisons were conducted utilising the tables of relative frequency numbers of each contracted form in every person.

## Results and discussion

When comparing the three corpora against each other, it came apparent that syntagmatic fusions of particles and the negation verb occur clearly more frequently in Subtitled Finnish than they do in Original Finnish or Translated Finnish, the less frequent these forms were in Original Finnish. All in all the difference between the results of Original Finnish and Translated Finnish was smaller than the difference between Translated Finnish and Subtitled Finnish. As there was no remarkable difference between Original Finnish and Translated Finnish, therefore one can conclude that audiovisual translation differs greatly from these two variants in terms in syntagmatic fusions of particles and the negation verb. This result is backed up by the end result of the study of Sonja Tirkkonen-Condit and Jukka Mäkisalo (2007). The conclusion of their study states as follows: "As might be expected, Subtitled Finnish exploits cohesive devices that are short and dialogical, but, unlike Translated Finnish at large, Subtitled Finnish also exploits unique cohesive devices even more frequently than Original Finnish." (Tirkkonen-Condit & Mäkisalo 2007.)

Although the end result was that subtitles contained more contracted particle forms than Original Finnish and Translated Finnish, the contracted particle forms in subtitles ended less frequently with a suffix than those in Original Finnish. The assumption is that as these forms are already contracted forms the translators have felt that adding a suffix to the end of these forms would increase the likelihood of them being more difficult to read. After all, the message of a single subtitle line must be conveyed in a very short a time.

Another thing that arose from the comparison between the three corpora was the finite forms of the syntagmatic fusions. On the whole singular forms were more multitudinous than plural forms in all three corpora. That being said, the distribution of occurrences between different persons was more wide in Subtitled Finnish and Original Finnish and Translated Finnish had clearly relatively more occurrences in third-person singular and third-person plural forms than in Subtitled Finnish, whereas Subtitled Finnish had a visible difference to the other two variants in the occurrences of second-person singular and second-person plural forms. It is highly likely that these differences are caused by the fact that most of the material in the subcorpus are film dialogue which results in subtitles including a factor called dialogicality, "the relative frequency of a linguistic item in spoken language" as Tirkkonen-Condit and Mäkisalo (2007, 224) put it. Like the two, I'm inclined to think that subtitlers prioritize such cohesive devices that are frequent in everyday spoken language (ibid.).

When comparing only the subtitle subcorpus, the emphasis of the study was on the difference between the contracted particle forms and the non-contracted forms. Note that the latter was not investigated when comparing the differences of the three corpora. The purpose of this comparison was to find out how the non-contracted forms of the syntagmatic fusion forms occurred in Subtitled Finnish, as only the forms that could have just as well have been converted to contracted forms without many changes on the sentence structure. The result of the comparison was that the contracted forms and non-contracted forms occurred in subtitles almost similarly, the difference was only a good four thousand occurrences in favour of the contracted forms.

Although all in all the difference was quite small, there were substantial differences between a single contracted form and its' non-contracted variant, e.g. the form *siksen* 'therefore I didn't' occurred nearly 52 times more often in its' contracted variant in the subtitles than it did in its' non-contracted form *siksi en*. It is highly likely, that the difference between the contracted



forms and non-contracted forms as a whole is mainly due to the fleeting nature of subtitles and the writers of subtitles have wanted to keep their subtitle lines as easy to read as possible. Also the notion of dialogicality surely has made a difference here, as e.g. the different grammatical person forms of *miksei* 'why not' appear much more frequently in subtitles than the different non-contracted forms of the same syntagmatic fusions of particles, which is likely due to an attempt to make the language of appear more colloquial.

The differences of the contracted forms and non-contracted forms were also compared when they appeared in short sentences, i.e. sentences that only included one or no other words than the syntagmatic fusion form – or one word at the most besides the particle and the negation verb when it comes to the non-contracted forms. Here the difference between the contracted and the non-contracted forms was greater than when comparing the two as a whole, especially when looking into the differences between the two variants in cases where they occurred alone without any other words in the sentence. For me, this is mainly due to dialogicality, although one cannot forget that time limit may have pushed translators to use more condensed language.

So all in all, the hypothesis of this study held up, as the syntagmatic fusions of particles and the negation verb occurred more frequently in Subtitled Finnish as they do in Original Finnish or Translated Finnish. Interestingly, although Subtitled Finnish and Translated Finnish are both obviously translated, Original Finnish and Translated Finnish were all in all closer together in the results of this study than Translated Finnish was with Subtitled Finnish. This proves that with regards to the syntagmatic fusions of particles and the negation verb Subtitled Finnish does differ from the other variants of Finnish and therefore these microstructures of Finnish can be considered as a characteristic of the so called subtitlers' Finnish.

In addition, one can state that Andrew Chesterman's relation norm takes more effect in subtitles than the expectation norm (when applied to the expectancies of the viewers about the notion of exact similarity between Subtitled Finnish and other variants of Finnish). Partly this is due to dialogicality and partly to the transient nature of subtitles and the translators need the condense the speech to a limited number of characters on screen. Of course, these two norms don't necessarily exclude each other, but within the limits of this study the relation norm seems to take priority over the applied expectation norm in Subtitled Finnish.

This study also tried to adhere to the demands laid out for the corpus studies of subtitles, as Riitta Jääskeläinen (2008, 94) has noted that most of the studies concerning subtitles have only used subtitles as the material of the study, not in fact as the actual object of observation. In addition, when the original *Finnish Broadcasting Company corpus of subtitles* was compiled, it was laid out that the material should be used to study whether Subtitled Finnish differs from Original Finnish and Translated Finnish in some way (Mäkisalo 2008). Besides, as Yves Gambier (2008, 94) has remarked, it is important to look at the relationship of the dominant literary norm and the conventions of written texts between the language used in subtitles. For one, this study has answered to all of these demands.