

Haapanen, L., L. Kääntä & L. Lehti (toim.) 2018. Diskurssintutkimuksen menetelmistä. On the methods in discourse studies. *AFinLA-e. Soveltavan kielitieteen tutkimuksia 2018 / n:o 11*. 20–44.

Jarmo Harri Jantunen
Jyväskylän yliopisto

Korpusavusteinen diskurssintutkimus (CADS): analyysiesimerkki homouden ja heterouden digitaalisista diskursseista

The article aims at combining two methodological approaches, namely corpus linguistics and discourse studies. Both of these approaches are present in corpus-assisted discourse studies (CADS) which makes use of both quantitative and qualitative methodologies. First, an attempt is made to discuss the differences and similarities of these two traditions based on Leech's (2000) listing. However, it turns out that this listing does not include all necessary viewpoints, and those are further discussed. In the empirical section of the paper, the Suomi24 corpus is analysed in order to provide an example of CADS. The analysis was carried out through collocation and discourse prosody analyses. The results reveal recurring discourses related to homosexuality and heterosexuality on the Suomi24 discussion forum. They support earlier observations on media and CADS studies on discourses on homosexuality and also reveal new discourse patterns, as well as shed new light on discourses on heterosexuality.

Keywords: corpus-assisted discourse studies, queer linguistics, collocation, discourse prosody, homosexuality, heterosexuality

Asiasanat: korpusavusteinen diskurssintutkimus, pervokielitiede, kollokaatio, diskurssi-prosodia, homoseksuaalisuus, heteroseksuaalisuus



VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

1 Johdanto

Artikkelin teemana on korpusavusteinen diskurssintutkimus (*Corpus-assisted Discourse Studies, CADS*), jossa yhdistyvät korpustutkimuksen ja diskurssintutkimuksen tavoitteet ja menetelmät. Artikkelissa luodaan aluksi katsaus korpus- ja diskurssintutkimukseen vertailemalla tutkimusaloja erityisesti aineistojen laadun ja tutkimusmenetelmien näkökulmasta. Tämä käsittely luo perustan artikkelin pääteemalle, korpusavusteiselle diskurssintutkimukselle. Katsaus lähtee liikkeelle Leechin (2000) luettelemista menetelmien peruseroista. Näitä eroja käsitellään artikkelissa kriittisesti. Tämän lisäksi nostetaan esiin eroja, jotka ovat yhtä tärkeitä mutta jotka on jätetty aiemmassa tutkimuskirjallisuudessa (esim. Leech 2000; Partington ym. 2013; Partington & Marchi 2015) käsittelemättä eksplisiittisesti ja kootusti. Erojen lisäksi tarkastellaan myös niitä yhtäläisyyksiä, jotka ylipäättään mahdollistavat menetelmien yhdistämisen. Artikkelin empiirisessä osassa esitellään Suomi24-korpuksen avulla korpusavusteisen lähestymistavan tarjoamia keinoja tutkia diskursseja; käsittelytapa on tilastoperustainen, fraseologinen ja kielenainesten yhteisvalintaa korostava, mikä onkin tyypillistä juuri CADS:lle (ks. Partington & Marchi 2015). Empiirisen osan tavoitteena on esitellä nimenomaan niitä menetelmiä, jotka eivät ole olleet tyypillisiä niin sanotussa perinteisessä diskurssintutkimuksessa. Pervokielitieteen (*Queer Linguistics*, vrt. pervotutkimus) alaan kuuluvassa analyysissä paljastuu samalla, millaisia diskursseja Suomen suurimmalla keskustelufoorumilla tuotetaan homoista ja heteroista.¹

2 Korpustutkimus ja diskurssintutkimus: menetelmien eroja ja rajapintoja

Korpusavusteinen diskurssintutkimus yhdistää nimensä mukaisesti kaksi tutkimusperinnettä: lähinnä kvantitatiiviseksi mielletyn korpustutkimuksen (KT) etupäässä laadulliseen diskurssintutkimukseen² (DT). Molemmilla tutkimusaloilla on pitkät perinteet, mutta vasta 2000-luku on nähnyt tutkimusalojen ja -menetelmien systemaattisen yhdistämisen. (Ks. esim. Biber & Jones 2005: 151–152; Ädel & Reppen 2008: 1–2.) Tutkimusalojen on aiemmin jopa katsottu eroavan aineistoiltaan ja me-

1 Kiitän professori Esa Lehtistä ja kahta nimetöntä arvioijaa tätä käsikirjoitusta koskeneista rohkaisista ja arvokkaista kommentista.

2 Diskurssintutkimuksella tarkoitetaan tässä laajasti erilaisia tutkimussuuntauksia, joskin tilanteeseen kielenkäyttöön (interaktionaalinen suuntaus) ja merkityksen rakentumiseen yhteisöllisenä toimintana (konstruktivistinen suuntaus) keskittyvät lähtökohdat painottuvat käsittelyssä ja vastaavasti diskurssien rakenteisiin (tekstuaalinen suuntaus) keskittyvästä tutkimuksesta kumpuavat taustaoletukset jäävät vähemmälle (suuntauksista ks. Luukka 2000: 144–156).

netelmiltään niin monella tavalla, ettei niiden yhdistämistä ole katsottu hedelmälliseksi (ks. mm. McEnery ym. 2006: 111). Tärkeänä periaatteellisena erona voidaan pitää erilaista suhtautumista kielen funktionaaliseen luonteeseen ja kontekstiin: Diskurssintutkimuksessa aineiston analyysissä ja tulosten tulkinnessa on tärkeää ottaa huomioon kielenkäytön tavoitteet ja tehtävät sekä yhteiskunnallinen ja sosiokulttuurinen konteksti, sillä tutkimusalan lähtökohtien mukaan merkitykset syntyvät vuorovaikutuksessa ja muodostavat yhteisen, jaetun repertuaarin (Luukka 2000: 137, 140–142; Pietikäinen & Mäntynen 2009: 24, 26–27; Partington ym. 2013: 2). Korpus tutkimuksessa on vuorostaan mahdollista keskittyä toistuviin rakenteisiin, frekvensseihin ja tilastolliseen merkitsevyyteen ilman, että aineistoa käsiteltäisiin kontekstiinsa sidoksissa olevana koherenttina kommunikaationa (esim. Tognini-Bonelli 2004: 18; Flowerdew 2008: 115–116; Partington ym. 2013: 8).

Muitakin selviä tavoitteisiin, menetelmiin ja aineistoihin liittyviä eroja on olemassa. Käsittelen eroja aluksi tutkimuskirjallisuuden avulla, minkä jälkeen pohdin niitä vielä omiin havaintoihini pohjautuen. Esittelyssä kiinnitän huomiota tutkimustraditioiden ydinkohtiin, eikä tämä valinta kata luonnollisestikaan kaikkia menetelmiin liittyviä nyansseja.

Leech (2000: 678–680) on luetellut tutkimusalojen eroiksi seuraavat viisi piirrettä³ (kursivoitu), mutta erot eivät välttämättä ole niin jyrkkiä kuin hän esittää (omat kommenttini kursiivitta). Näistä eroista ensimmäinen liittyy edellä käsiteltyyn kontekstin tärkeyteen.

- 1) *DT:ssa fokusoidaan tekstien sisältöihin ja vuorovaikutuksen keinoihin ja analysoidaan sisällön suhdetta ympäröivään yhteiskuntaan, KT on ollut pitkälti itse kielen rakenteisiin ja sanastoon keskittyvää.* Kuitenkin etenkin tekstuaalispainotteisessa, kielentutkimukseen keskittyvässä DT:ssa (ks. Luukka 2000) tarkastellaan nimenomaan kieltä, sanastoa ja rakenteita (eli mikrotasoa) silloinkin, kun tavoitteena on saavuttaa tietoa yhteiskunnasta ja kulttuurista (makrotasosta). Silloin pyritään löytämään tavat, joilla tietyt tavoitteet halutaan saavuttaa ja joilla pyritään vaikuttamaan toisiin ihmisiin, sekä tarkastelemaan valintojen seurauksia (mm. Pietikäinen & Mäntynen 2009: 13, 17; Partington ym. 2013: 2–5). Toisaalta vaikka KT:ssa saatetaan keskittyä kielenpiirteiden kuvaukseen ilman yhteiskunnallista ulottuvuutta, tällainen KT:n määrittely on kuitenkin varsin kapea, koska KT:ta voidaan pitää pikemminkin menetelmänä (ks. McEnery ym. 2008: 7; myös Cheng 2012: 6; kuitenkin mm. Tognini-Bonelli 2001), jonka avulla on mahdollista tarkastella hyvin erilaisia tutkimuskysymyksiä (ks. esim. Cheng 2012: 6–8; Partington ym. 2013: 5–9).

3 Leech (2000) keskittyy luokittelussaan puhuttuihin aineistoihin, mutta piirteet ovat yleistettävissä myös muunlaisiin aineistoihin.

- 2) *DT:ssa yksittäiset, kokonaiset tekstit ovat analyysin kannalta merkittäviä, kun taas KT:ssa tärkeää on edustava otos eivätkä tekstit ole aineistossa aina kokonaisina.* DT:ssa tekstikäsitys on hyvin laaja, sillä aineisto voi koostua tutkijan keräämistä valokuvista, haastatteluista, videotallenteista, päiväkirjamerkinnöistä ja monista muista materiaaleista (Pietikäinen & Mäntynen 2009: 157–158, 160), joihin oletus tekstin kokonaisuudesta ei välttämättä sovi. Lisäksi yhdessä tutkimuksessa useat, erityyppiset tekstit ja tekstifragmentit voivat muodostaa aineistokokonaisuuden. KT:ssa aineiston edustavuus on kyllä erityisen tärkeää, mutta tallennuskapasiteetin kasvettua tekstit sisällytetään yhä useampaan korpukseen kokonaisina.
- 3) *DT:ssa aineiston kerääjä, käsittelijä ja analysoija ovat usein yksi ja sama henkilö, mutta KT:ssa analysoija käyttää usein valmiita, toisten aiemmin keräämiä aineistoja.* Ei kuitenkaan ole tavatonta, että korpustutkimuksessakin tutkija aloittaa työnsä aineiston kokoamisella, jos valmiita sopivia materiaaleja ei ole käytössä; tämä koskee nykyisinkin erityisesti pieniä kielialueita.
- 4) *DT:ssa aineistot ovat perinteisesti vain analysoijan käytössä ja saatavuus on rajoitettu (ellei aineisto ole esim. laajemman tutkimusryhmän käytössä), mutta KT:ssa aineistot ovat tyypillisesti (käyttölupien rajoissa) laajemmalle tutkijajoukolle tarkoitettuja.*
- 5) *Jos käytössä on sähköisiä aineistoja, DT on silloinkin tyypillisesti laadullista tutkimusta, KT:lle taas ovat luonteenomaisia kvantitatiivisen tiedon saavuttamiseen tähtäävät analyysit ja työkalut.* On kuitenkin huomattava, että myös DT:ssa toisteisuudella on merkityksensä ja tutkimuksessa haetaan säännönmukaisuuksia, joihin tulkinnat pohjautuvat (ks. esim. Pietikäinen & Mäntynen 2009: 37, 167). Vaihtelun tarkastelussa (ks. mts. 37) kiinnitetään huomiota laadun lisäksi myös määrälliseen informaatioon esimerkiksi silloin, kun halutaan selvittää tekstien moniäänisyyttä. KT:ssa puolestaan määrällistä tietoa selitetään usein laadullisin keinoin ja analyysi nojaa pohjimmiltaan laadullisiin kielellisiin kategorioihin, kuten sanaluokkiin ja lauseenjäseniin (ks. Heikkinen ym. 2012: 330). Ylipäätäänkin aloja voi vain harvoin luokitella puhtaasti laadullisiksi tai määrällisiksi.

Leechin (2000) luettelemien erojen lisäksi on vielä tarpeen nostaa esiin seuraavat, etenkin aineistojen luonteeseen ja menetelmiin liittyvät seikat:

- 6) *DT:ssa aineistot ovat menetelmien vuoksi usein pieniä, jopa yhdestä tai muutamasta kymmenestä tekstistä koostuvia, ja aineistoja voidaan rajata tutkimuksen edetessä yhä pienemmiksi. KT:ssa tekstien määrä kasvaa tuhansiin, jolloin saneina laskettuna aineistoa voi olla miljoonia tai miljardeja saneita.* Aineiston koon määrittely tapahtuu DT:ssa tavallisesti

kyllääntymisen avulla (Pietikäinen & Mäntynen 2009: 160), KT:ssa puolestaan aineiston määrä liittyy pikemminkin tilastollisten menetelmien ja korpustyökalujen soveltuvuuteen ja etenkin aineiston edustavuuteen (esim. Biber 1993).

- 7) DT:ssa analysoija tuntee aineistonsa läpikotaisin tutkimusmenetelmään kuuluvan aineiston keräämisen, analyysiprosessin (mm. aineiston pilkkominen, purkaminen, uudelleen kokoaminen; Pietikäinen & Mäntynen 2009: 166) ja myös sen pienuuden vuoksi. KT:ssa näin on harvoin, koska aineistot ovat laajoja ja korpuksia luetaan fragmentaarisesti kokonaisten tekstien sijaan (Tognini-Bonelli 2004: 18). KT:ssa tutkijalle ovat tärkeitä pikemminkin aineiston meta- eli kuvailutiedot (koonti- ja luokitteluperiaatteet [ks. esim. Tyrkkö & Taavitsainen 2012], tekstintuottajatiedot, tekstilajitiedot, tuottamisaika jne.), annotointitavat, erilaiset listat (sana-, avainsana- ja klusterilistat) ja konkordanssit (eli vertikaalisesti luettavat listaukset sanojen esiintymistä lähikonteksteineen ja metatietoineen, ks. myöhemmin kuvaa 1).
- 8) DT:ssa tutkimusaineisto valitaan tarkkaan edustamaan tarkasteltavana olevaa tekstilajia, -tyyppiä tai rekisteriä, jolloin aineisto kuvaa juuri näissä tavattavia merkityksiä ja diskursseja ja soveltuu yleensä vain yhteen tai yhdentyyppiseen tutkimukseen. KT:ssa ovat yhä tavallisempia niin sanotut yleiskorpuksot, jotka kuvaavat kieltä tai kielivarianttia mahdollisimman laajasti. Tosin erilaisia tekstilajikorpuksia tai kielivarianttikorpuksia kootaan edelleen laajasti.
- 9) DT:ssa tutkimus etenee materiaalin keräämisen jälkeen aineiston perusteelliseen läpikäymiseen ja siihen tutustumiseen sekä kirjoittamisen avulla tehtävään prosessointiin. Tutkimus- ja raportointiprosessi on monipolvinen, spiraalimainen, jaksottainen ja pakoilee selvärajaisuutta ja suoraviivaisuutta; kirjoittaminen ja raportointi ovat analyysiprosessin oleellisia osia, eikä analyysia, pohdintaa ja kirjoittamista voi erottaa toisistaan. (Pietikäinen & Mäntynen 2009: 142–143, 166–167.) KT:ssa tutkimus voi alkaa esimerkiksi erilaisten sanalistojen ja konkordanssien tarjoamista frekvenssitiedoista ja tilastollisista analyyseistä ja edetä huomattavasti suoraviivaisemmin analyysistä kohti raportointia. Kirjoittaminen ei ole menetelmällisesti yhtä kiinteästi osa analyysia vaan liittyy pikemminkin tutkimuksen raportointiin.
- 10) DT:ssa aineistoa analysoidaan tyypillisesti manuaalisesti, joskin saatavilla on laadullisen aineiston hallintaan ja käsittelyyn sopivia ohjelmia. Sitä vastoin KT perustuu tavallisesti käyttöliittymien, korpustyökalujen (kuten sanalistageneraattorien ja konkordanssiohjelmien) ja tilastotyökalujen käyttöön sekä automaattiseen analyysiin.

- 11) DT:ssa analyysi on lähtökohdiltaan teksti- ja aineistovetoista, jolloin tutkija syventyy tekstien sisällön esiin nostamiin ilmiöihin ja merkityksiin. KT:ssa tutkija voi lähestyä aineistoaan aineistovetoisesti tai -pohjaisesti, joista ensimmäinen tapa muistuttaa DT:n lähestymistapaa. Jälkimmäisessä puolestaan keskitytään pikemminkin ennalta valittuihin kielenpiirteisiin ja merkityksiin (käsitteistä tarkemmin Jantunen 2009; ks. myös Lee 2008: 88–92).
- 12) DT:ssa tavoitteena on harvoin pyrkiä tyhjentävään analyysiin ja tulosten yleistettävyyteen (Biber & Jones 2005: 152; Pietikäinen & Mäntynen 2009: 157–158). KT:ssa sen sijaan tämä on keskeinen tavoite, ja siihen pyritään valitsemalla aineisto ulkoisten valintakriteereiden avulla (Biber & Jones 2005: 152; McEnery ym. 2008: 13–15). KT:ssa tavoitteena on jo aineiston koostamisvaiheessa varmistua siitä, että aineistoa voidaan pitää edustavana ja saatuja tuloksia voitaisiin yleistää (ks. Biber 1993; McEnery ym. 2008: 13–15; Jantunen 2009). DT:ssa luotettavuus syntyy valintojen avoimesta perustelemisesta ja läpinäkyvyydestä (Pietikäinen & Mäntynen 2009: 168), KT:ssa laajoista aineistoista, frekvensseistä, tilastollisista menetelmistä ja toistettavuudesta.
- 13) DT:ssa tutkimusaineistoja koskeva laajamittainen teoreettinen tutkimus, menetelmälliset analyysit ja metatietojen kerääminen ovat harvinaisia, KT:ssa taas niin sanottu metatutkimus (aineistojen kerääminen, annotointi ja tallentaminen, ks. Partington ym. 2013: 6) muodostavat ison osan alan tutkimusta.

Monet edellä mainituista eroista liittyvät tutkimussuuntausten perustavoitteiden ja siten aineistojen kokojen ja menetelmien eroihin: Diskurssintutkimuksen pienet aineistot on mahdollista tuntea perin pohjin, ne on helppo kerätä tutkimusta varten itsenäisesti, niihin sopii laadullinen tutkimus paremmin kuin määrällinen, niiden tarkoitus ei ole antaa edustavaa kuvaa aineiston ulkopuolisesta maailmasta, eikä niiden avulla pyritä tyypillisesti tyhjentävään kuvaukseen. Korpustutkimuksessa tilanne on usein juuri päinvastoin, kun pyrkimyksenä on aineistojen edustavuus ja tulosten laaja yleistettävyyys. Seuraavaksi siirryn tarkastelemaan CADS-tutkimusta, jossa tässä luvussa käsitellyt menetelmät yhdistetään.

3 Korpusavusteinen diskurssintutkimus (CADS) ja fraseologinen kielikäsitely

CADS-tutkimus juontaa juurensa 1990-luvulla alkaneesta korpusavusteisesta, usein poliittisiin teksteihin liittyvästä tutkimuksesta (mm. Krishnamurthy 1996; Stubbs 1996), mitä ennen isojen aineistojen käyttö ja pyrkimys tulosten yleistettävyyteen

oli vähäistä DT:ssa (Biber ym. 1998: 106). Korpustutkimuksessa CADS on saavuttanut jonkinasteista jalansijaa, mutta diskurssintutkimuksessa laajojen tekstiaineistojen ja määrällisten menetelmien käyttö on ollut toistaiseksi melko harvinaista (alan kehityksestä mm. McEnery ym. 2006: 111–113; Lee 2008: 86–87; Partington ym. 2013: 10–11). Faircloughin (2003: 6) mukaan määrällisellä analyysillä voidaan kuitenkin tukea laadullista DT:ta ja näitä lähestymistapoja kannattaa käyttää rinnan. CADS siis yhdistää kaksi lähtökohdiltaan hyvin erilaista lähestymistapaa: laajoihin tekstimateriaaleihin perustuvat määrälliset ja tilastolliset menetelmät otetaan tutkimuksessa lähtökohdaksi ja yhdistetään laadulliseen analyysiin, ja näin vastataan kritiikkiin, jota on esitetty DT:n aineistojen pienuudesta, edustavuudesta ja tulosten yleistettävyydestä (esim. Stubbs 1997). Ensimmäinen laaja suomalainen CADS-tutkimus on Lehdon (2018) analyysi japaninsuomalaisten kielipuheesta.

Aineistojen ja menetelmien näkökulmasta CADS voi tarkoittaa jo valmiiksi kerättyjen aineistojen käyttöä tai uusien kokoamista, erilaisten sanalistojen generoimista, konkordanssien (ks. kuvaa 1) määrällistä ja laadullista analyysia sekä aineistojen vertailua esimerkiksi avainsana-analyysin avulla.

eivät halua lapsilleen opettaa tätä ateistista	homo	oppia ihmiselämästä, joutuisivat pakon eteen jos
ja kristittyjen asiat etusijalle mm. ateististen	homojen	asioihin nähden, hän ei ole toiminut oikein paimer
Näihin menee vain liberaalit, ateistit,	homot	ja lesbot, ja sitten ne liberaalit papit ja piispat ja lu
indut, butdhalaiset, komfutsealaiset, ateistit,	homot	, lesbot sekä kommarit joutuvat POLVISTUMAAN
Ja Juutalaiset, ateistit,	homot	ja lesbot ja ympärileikkaavat suomi hu*ranaiset ja
Onko liberaalit tai ateisti	homoista	puhumattakkaan tunnustanut syntejään täällä?
n elelee helvetissä jota kansoittavat ateistit,	homot	, abortintekijät ja muut hänen suuresti pelkäämän
Kirkko on kehittynyt riittävästi, kun ateistit,	homot	ja uuspakanat liittyvät joukolla kirkkoon.
Tietenkin muitakin ryhmät kuten ateistit,	homot	, kommarit, vihreät ym. saavat tuntea nahoissaan i
uterilaisten palstallakin mellastavat ateistit	homot	.
Mukaanlukien ateistit,	homot	ja kommunistit taistelivat maansa ja kotinsa puol
Vihapuheitta puhuu ateistit,	homot	ja liberaalit.
Siinä ateistit,	homot	kun varkaatkin (ja kaikki muu synti) asettuvat nii
kaansa paskamyrskyn, mm. tunnettu ateisti,	homojen	ja mamujen hyysäri pappi Teemu Laajasalo on kut
Tyhjää ateisti	homon	vouhotusta.
Tämä ateisti	homo	pyrkii kaikessa kirjoittelussaan herjaamaan KRS,

KUVA 1. Kuvakaappaus Korp-käyttöliittymän (Kielipankki) tuottamasta *homo*-sanon konkordanssista (järjestetty vasemman kontekstin mukaan).

Edellisessä luvussa kävi ilmi, että CADS:ssa yhdistyvillä menetelmillä on monia periaatteellisia eroja, mutta niitä kuitenkin yhdistävät nojautuminen toistuvuuteen ja säännönmukaisuuteen sekä todellisten kieliaineistojen käyttö tutkimuksessa. CADS-tutkimus perustuukin kielikäsitteeseen, jonka mukaan kieli on kertautuvaa, yhteisesti jaettua ja kontekstuaalista – käsitys on samanlainen kuin muussakin DT:ssa. Kielenkäytön nähdään perustuvan siis toistuviin ja säännönmukaisiin rakenteisiin: kommunikointiamme leimaa pikemminkin yhteinen, jaettu repertuaari kuin uniikkien ilmausten jatkuva tuottaminen. Tutkimuksessa on keskeistä näkemys, jonka mukaan kieltä ei tuoteta yksittäisiä kielellisiä elementtejä muistista poimien, vaan

niin sanotusti fraseologisesti eli muistiin varastoituneiden ja kielenkäyttäjien jakamien monisanaisten yksiköiden sekä leksikaalis-kieliopillisten ja leksikaalis-semanttisten assosiaatioiden avulla (ks. tarkemmin esim. Hoey 2005; Jantunen & Brunni 2012; Pace-Sigge & Patterson 2017).

Fraseologinen kielikäsite kuvastuu muun muassa Sinclairin (1991: 110) idiomiperiaatteessa (*idiom principal*) ja Hoeyn (2005) leksikaalisen primingin (*lexical priming*) teoriassa: molemmat korostavat sitä, että tuotamme kommunikoidessamme jatkuvasti ilmauksia, jotka koostuvat toistensa kanssa säännöllisesti yhdessä esiintyvistä sanoista ja rakenteista, kuten kollokaatioista. Leksikaalisen primingin perusta on psykolingvistinen: kielenkäyttömme perustuu sekä tiedostamattomiin että tiedostettuihin valintoihin, ja etenkin tiedostamattomat valinnat tuottavat taaajaan toistuvia rakenteita, jotka takaavat viestinnän nopean sujumisen (Hoey 2007: 8; Pace-Sigge & Patterson 2017: XII–XIV). Hoeyn sanoin kielenkäyttömme rakentuu seuraavasti:

[W]hen we use the word, we will characteristically use it with one of its typical collocations, in its usual grammatical function, in the same semantic context, in the domain we have come to associate it with, as part of the same genre, in a familiar social context, with a similar pragmatics and to similar textual ends. (Hoey 2007: 8.)

CADS-tutkimuksessa fraseologinen lähestymistapa näkyy tyypillisesti siten, että analyysissa keskitytään usein leksikaalisiin myötäesiintymärakenteisiin eli kollokaatioihin (esim. Baker ym. 2008), klustereihin (*n*-grammeihin) eli toisteisiin samalaisina esiintyviin monisanaisiin yksiköihin (esim. Partington & Morley 2004; Lehto 2018) sekä semanttisiin preferensseihin ja diskurssiprosodioihin (esim. Stubbs 2001: 65–66, 105–108; Lehto 2018). Tämänkin artikkelin esimerkkitutkimus jakaa fraseologisen kielikäsitteen, ja siinä hyödynnetään yllä mainittuja keskeisiä käsitteitä.

Tutkimuksessa tällaisen toisteisuuden ja säännönmukaisuuden havaitsemiseksi tarvitaan suhteellisen isoja aineistoja sekä myös menetelmiä, joilla määräperustaisen säännönmukaisuuden pystyy havaitsemaan. Joidenkin kielenpiirteiden, kuten toistuvien kollokaatioiden tai tekstijaksojen (esim. klustereiden ja muiden monisanaisten yksiköiden), saavuttaminen pienestä aineistosta on hankalaa, koska tekstiä ei ole tarpeeksi toisteisuuden havaitsemiseksi. Perinteisesti DT:ssa kielenpiirteiden eksakti toistuvuus ei olekaan ollut tutkimuksen fokuksessa juuri menetelmien ja aineistojen pienuuden vuoksi, vaan aineistoista on pyritty löytämään erilaisia, toisiaan täydentäviä kielenilmiöitä, jotka tuottavat tiettyä diskurssia. CADS:ssa päätaivoitteena voi sen sijaan pitää sellaisten merkitysten löytämistä, jotka jäävät havaitsematta pieniin aineistoihin perustuvassa manuaalisessa, ”paljain silmin” tehtävässä tutkimuksessa (Partington ym. 2013: 11).

4 CADS käytännössä: digitaalisia diskursseja seksuaalivähemmistöistä ja -enemmistöistä

CADS-tutkimusta havainnollistavassa esimerkkitutkimuksessani selvitän, millaisia diskursseja eli yhteiskunnassa vallalla olevia ja yhteisiä sosiaalisia merkityksiä liittää verkkokeskusteluissa seksuaalivähemmistöistä ja -enemmistöistä puhumiseen. Tässä tutkimuksessa tarkastelen vähemmistödiskursseja *homo*-sanan ja enemmistödiskursseja *hetero*-sanan avulla. Keskityn artikkelissa nimenomaan tilastollisiin ja määrällisiin lähtökohtiin, koska ne ovat perinteisessä diskurssintutkimuksessa harvinaisempia.

4.1 Taustaa ja lähtökohtia

Seksuaalivähemmistöjä koskevassa keskustelussa nostetaan yleensä esiin muun muassa sairaudet ja luonnottomuus, tasavertaisuus, uskonto ja politiikka tai keskustelussa käsitellään tiettyjä paikkoja ja henkilöitä tai henkilöryhmiä. Erityisesti väkivalta ja vihapuhe ovat voimakkaasti läsnä digitaalisessa vähemmistödiskurssissa. (Esim. Baker 2004b; Baker 2005; Bachmann 2011; Jantunen 2018.) Perhe- ja yksiavioisuusdiskurssit (Baker 2005) sekä (oman) seksuaalisuuden pohdinta, tunteista puhuminen ja sukupuolisuusdiskurssit (Jantunen 2018) ovat puolestaan heterouteen liittyviä diskursseja digitaalisissa keskusteluissa. Suomi24:ää koskeva aiempi tutkimukseni (Jantunen 2018) on tehty myös CADS-tutkimuksena, mutta siinä menetelmänä on tilastollinen avainsana-analyysi toisin kuin käsillä olevassa tutkimuksessa, joka lähtee liikkeelle kollokaatioanalyysistä (avainsana-analyysistä CADS:ssa ks. myös Baker 2004a; Karjalainen ym. 2017).

Tarkastelen diskursseja Suomi24-keskustelufoorumista kootusta, Kielipankin lemmatisoimattomasta Suomi24-korpusaineistosta (Aller Media Oy 2014; Lagus ym. 2016). Aineistossa on verkkokeskusteluja vuosilta 2001–2016, ja sen koko on noin 2,6 miljardia sanetta (Meta-Share 2016). Tutkimusaineiston kokoamisen perustaksi otettiin harvinaisempi *hetero*-sana kappaleen mittaisine konteksteineen (n=89 000). Tämän jälkeen korpuksesta haettiin *homo*-sanat konteksteineen ja valittiin systemaattisen otannan avulla suunnilleen samansuuruinen otos. Tämän jälkeen aineistoista poistettiin mahdollisimman hyvin toisteiset ketjut ja samanlaisina toistuvat tekstijaksot, jotka ovat tyypillisiä verkkoaineistoille muun muassa viestiketjuissa toistettujen aiempien viestien vuoksi. Aineistojen ko'oisiksi tuli näin 2,2 (*hetero*) ja 2,1 miljoonaa sanetta (*homo*); *hetero*-sanojen lukumäärä näissä aineistoissa on 72 000 ja *homon* 94 000 sanaa. (Aineistosta tarkemmin Jantunen 2018.)

Tarkastelen ilmauksia aluksi keskittymällä lähikontekstin kollokaatteihin. Kollokaatiot ovat toistuvia, syntagmaattisessa suhteessa olevia myötäesiintymiä (mm. Sinclair 1991: 115–121; Jantunen 2004: 15–16), jotka rakentavat tekstin sisäistä ko-

heesiota ja voivat olla diskurssispesifejä (mm. Fairclough 2003: 37, 130–131). Tässä tutkimuksessa kollokaatioanalyysi perustuu tilastolliseen analyysiin, joka on tehty seuraavan kolmen rajauksen avulla⁴:

- 1) Tarkasteluväli on 4L–4R eli neljä sanaa hakusanan molemmin puolin. Tarkastelualueita on pidetty optimaalisena (mm. Jones & Sinclair 1974), ja sen sisään mahtuu myös suomen keskipituinen, 5–7 sanan mittainen lause (ks. Hakulinen ym. 1980: 106; Heikkinen ym. 2001).
- 2) Tilastollinen testaus on tehty MI- eli *Mutual Information* -testillä, joka korostaa hakusanan ja ympäristön sanojen yhteisesiintymän todennäköisyyttä ja joka yleensä nostaa merkitseviksi sisältösanat kieliopillisten sanojen sijaan (ks. Barnbrook 1996: 95–99).
- 3) Kollokaatiksi laskettavan sanan vähimmäisfrekvenssi on 50, jolloin vähennee mahdollisuus, että kollokaatti on peräisin yhden tekstintuottajan kirjoituksista tai verkkokeskustelulle tyypillisistä toistetuista jaksoista.

Tarkasteluun olen ottanut mukaan 200 merkitsevintä kollokaattia, joista käsittelen ensin 20:tä yksityiskohtaisesti. Keskittymällä pieneen määrään tilastollisia kollokaatteja päästään kiinni yksityiskohtaiseen analyysiin; myöhemmin 200 merkitsevälle kollokaatille tehdyn merkitysryhmittelyn avulla pystytään havainnoimaan laajemmin diskurssiprosodioita.

Diskurssiprosodiolla tarkoitetaan kielellisten ilmausten ominaisuutta esiintyä niille tyypillisissä diskursiivisissa konteksteissa, ja se voi kertoa kielenkäytön pragmaattisista funktioista tai asenteista. Se myös luo diskurssikoheesiota kielenkäyttöön ja teksteihin. (Stubbs 2001: 65–66.) Samasta ilmiöstä käytetään myös termejä semanttinen prosodia (mm. Sinclair 1996: 87–88) ja evaluatiivinen prosodia (mm. Partington ym. 2013: 58–61). Näistä ensimmäinen korostaa sanan inherenttiä kontekstisidonnaisuutta ja merkityksen muodostumista sanaa laajemmasta kokonaisuudesta; jälkimmäinen painottaa arvottamista ja ilmausten valinnan takana olevaa evaluatiivista harmoniaa diskurssin tuottamisen perustana (ks. myös Jantunen 2004: 25–26; Hunston 2007).

Diskurssiprosodian havaitseminen ilman laajoja tekstiaineistoja ja korpusmenetelmiä voi olla vaikeaa huolimatta siitä, että fraseologiset piirteet opitaan hallitsemaan normaalissa kielenkehityksessä ja niitä toistetaan kielenkäytössä. Tässä tutkimuksessa diskurssiprosodiaa lähestytään tarkastelemalla kollokaatteja ja etsimällä niitä yhdistäviä kontekstuaalisia merkityksiä.

4 Tarkemmin kollokaatioiden määrittelystä ja laskennasta, ks. Barnbrook (1996: 95–99) ja Jantunen (2004: 15–21, 76–80).

4.2 Kollokaatit diskurssien vihjeinä

Tarkasteltavana olevien *homo*- ja *hetero*-sanojen 20 tilastollisesti merkitsevintä kollokaattia on lueteltu merkitsevyysjärjestyksessä taulukossa 1. Aluksi keskitytään siis merkitsevimpiin kollokaatteihin, sillä ne kuvaavat sitä, millaisia diskursseja homoista ja heteroista tuotetaan Suomi24-keskusteluissa ja näyttävät myös poikkeavat aineistoittain.

TAULUKKO 1. *Homo*- ja *hetero*-sanojen merkitsevimmät kollokaatit MI-testin merkitsevyysarvoineen ja frekvensseineen.

homo		hetero	
<i>voltti</i>	7.63699 / 125	<i>terveisin</i>	9.92136 / 64
<i>sapiens</i>	7.59064 / 93	<i>pesunkestävä</i>	9.47349 / 51
<i>ime</i>	7.46316 / 100	<i>homosta</i>	9.46777 / 700
<i>synnytä</i>	7.35681 / 59	<i>umpi</i>	9.41938 / 112
<i>hirteen</i>	7.31255 / 56	<i>ristillä</i>	9.35919 / 147
<i>synnistään</i>	7.28707 / 61	<i>valkohoinen</i>	9.30290 / 58
<i>synnyttään</i>	7.26457 / 126	<i>lesbosta</i>	9.30053 / 57
<i>syö</i>	7.25789 / 143	<i>valkoinen</i>	9.28386 / 127
<i>tunnistaa</i>	7.23912 / 199	<i>täys</i>	9.27746 / 195
<i>heterosta</i>	7.23608 / 127	<i>lihaa</i>	9.24561 / 81
<i>synny</i>	7.23451 / 158	<i>ootko</i>	9.18188 / 95
<i>hommaa</i>	7.21191 / 411	<i>golgatan</i>	9.14375 / 112
<i>haista</i>	7.19103 / 188	<i>käännyttää</i>	9.13413 / 104
<i>lesbopareja</i>	7.16626 / 55	<i>muuttua</i>	9.10227 / 470
<i>lespoja</i>	7.12661 / 61	<i>tylsä</i>	9.08877 / 50
<i>lesboihin</i>	7.11757 / 67	<i>lisääntyvät</i>	9.04716 / 63
<i>olematta</i>	7.08765 / 50	<i>olenkin</i>	9.02178 / 355
<i>lespot</i>	7.07270 / 100	<i>normaalit</i>	9.01646 / 107
<i>lesboparien</i>	7.04812 / 75	<i>normaali</i>	8.99797 / 570
<i>kaappi</i>	7.04812 / 75	<i>ps</i>	8.99241 / 152

4.2.1 Homouteen liittyvät merkitsevimmät kollokaatit

Homo-sanan merkitsevin kollokaatti on hieman yllättäen *voltti*. Tämä selittyy aineistossa tavallisella idiomilla *heitä homo voltti*, jolla ei läheskään aina viitata suuntautumiseen vaan pikemminkin pyritään osoittamaan toisen tyhmyys tai nolaamaan toinen, kuten esimerkissä (1). Toki herja voi kohdistua myös homoihin, kuten esimerkissä (2), joka on osa erään kaupungin vähemmistöilmapiiriä käsittelevää keskustelua.

(1) katotaan keväällä kun me stadin kuninkaat juhliatan mestaruutta! **heitä homo voltti!**

(2) **Heitä homo voltti**, ja ota heimlich ote kaverisi kanssa.

Herjat, loukkaava kielenkäyttö ja suoranainen vihapuhe ovat tavallisia verkkokeskusteluissa, ja ne ovat osa internetin tavallista verkkovihaa (ks. esim. Pöyhtäri ym.

2013: 19, 49; Oikeusministeriö 2016: 81, 94, 129). *Homon* merkitsevimpien kollokaattien joukossa verkkovihasta kertovat myös kollokaatit *ime*, *hirteen* ja *haista* (3–5), joita kaikkia käytetään ilmauksissa, jotka voidaan luokitella haukkumiksi, vihapuheeksi tai väkivaltaisiin tekoihin kehottamiseksi. Myös näiden joukossa on tapauksia, joissa ei käsitellä homoutta sinänsä, vaan *homo*-nimitystä käytetään tuottamassa diskursssia jostakin yleensä epämieluisasta tai huonona pidetystä (3).

- (3) Kun jotkut vaan polttaa puoleen väliin ja tuhlaa tupakkaa. No jos oot niin saatanan idiootti **niin vitun homo onnea vaan ja ime kusipää munaas!**
- (4) jokaisella ihmisellä on oikeus olla sellainen kuin itse haluaa. **ja kakat! homot hirteen sillä selvä!**
- (5) Mikä sinä olet muita tuomitsemaan ainoa joka tuomitsee on jumala! **Joten haista homo vittu!**

Sapiens-kollokaatti selittyy luonnollisesti ihmistä tarkoittavasta *homo sapiens* -sanaaliitosta. *Syö*-kollokaatille ei löydy yhtä ainoaa selitystä, vaan sanaa käytetään monissa merkityksissä: 'syödä' (*Homot syö tonnikalaa .. yök*), 'viedä, käyttää' (*Joten heteroiden on syytä käydä antamassa äänensä, homot muuten syö sun äänes*); lisäksi keskustelupalstalla toistuu usein spämmi *syö paskaa homo*, mikä osaltaan lisää juuri kyseisen verbin yllättävääkin merkitsevyyttä. *Hommaa* taas selittyy taajaan käytetyllä ilmauksella *olla homojen hommaa*, jolla arvotetaan jokin tekeminen huonoksi tai toisarvoiseksi: *Sisustus on homojen hommaa!*, *Kyllä rekka hommat on homojen hommaa*. Näissäkään ei kyse ole suuntautumisesta sinänsä, vaan kuten edellä, *homo*-nimitystä käytetään jostakin huonona pidetystä asiasta tai ihmisestä. Myös tässä nimityksen käytöllä tuotetaan negatiivisuuden diskurssia, ja ilmaisemalla jokin asia homojen työkse tai puuhasteluksi arvotetaan tekeminen ei-toivottavaksi tai vähemmän arvotetuksi. Vastaava ilmaus on *tyttöjen hommaa*, ja verkosta löytyy myös useita tapauksia, joissa jokin on *homojen ja tyttöjen hommaa*. *Olla*-verbin *ma*-infinitiivi *olematta* liittyy keskusteluun, jossa pohditaan sitä, voiko jotakin tehdä tai olla jonkinlainen *olematta homo* (*Miehet voivat olla samanhenkisiä olematta homoja*).

Oman ryhmänsä kollokaattilistassa muodostavat *syntyä*-verbin muodot *synnytä*, *synny* ja *synnyttään*. Verbin avulla tuotetaan diskurssia siitä, onko homoseksuaalisuus synnynnäistä, kasvatuksen ja ympäristötekijöiden tuottamaa vai opittu ominaisuus. Tämä diskurssi liittyy vahvasti pitkään Suomessakin jatkuneeseen keskusteluun, jota on käyty esimerkiksi uskonnon, lääketieteen ja psykologian näkökulmasta mediassa (ks. esim. Juvonen 1997). Verbin kieltomuodot *synnyttä* ja *synny* kertovat selvästi useiden keskustelijoiden pitävän synnynnäisyyttä mahdottomuutena (6–7). Toisaalta keskusteluissa tuotetaan vastakkaistakin diskurssia (8). Keskustelupalstalla ovat siis läsnä samat näkökulmat kuin median ja tieteenkin tuottamissa diskursseissa.

- (6) homos on synti ja **homoksi ei synnytä**;
- (7) **Kukaan ei homoksi synny**, eikä mitään homogeeniä ole edes olemassa.
- (8) **Homoksi synnyttään**, se ei ole tahdon valinta.

Esimerkissä (9) *synti*-sanalla tuodaan keskusteluun uskonto-teema, joka on yksi voimakkaimmista homoseksuaalisuuteen liittyvistä diskursseista niin julkisessa keskustelussa (ks. esim. Bachmann 2011) kuin verkkokeskusteluissakin (Jantunen 2018). Tätä diskurssia tuotetaan taulukon 1 kollokaatilla *synnistään*, mikä näkyy selvästi esimerkiksi (9), jossa *synti* toistuu useasti. Katkelmassa on runsaasti myös muita uskonto-diskurssia rakentavia ilmauksia (kursivoitu), kuten *tehdä parannus*, *Raamattu* ja *Jumala*.

- (9) Jos homo tahtoo *tehdä parannuksen synnistään*, sillä *Raamatun* mukaan homous on **synti**, ja hän *nöyryty Jumalan edessä, tunnustaa syntinsä* ja *pyytää kokosydamisesti Jeesusta antamaan anteeksi ja tunnustaa Jeesuksen Herrakseen ja vapahtajakseen*, niin homo vapautuu **synnistään**. Ilman *Jeesusta* se ei onnistu! Homous ja *ateismi* aiheutuvat *demoneista*.

Voiko homon tunnistaa? Tunnistamisesta puhutaan laajasti verkossa, ja se on ollut viime aikoina esillä myös monien lehtien (mm. HS 2017) julkaiseman tutkimuksen (Kosinski & Wang 2018) vuoksi. Tutkimuksessa selvitettiin, voiko tekoäly tunnistaa kasvopiirteiden perusteella seksuaalisen suuntautumisen. Keskustelupalstalla taas tunnistamisen mahdollisuus voi askarruttaa yhtä hyvin itse vähemmistöön kuuluvaa, joka haluaa tutustua toiseen (10), tai puolisoa, joka pohtii kumppaninsa suuntautumista (11). Lisäksi tunnistamisesta käydään palstalla myös yleistä keskustelua (12). Epäily niin sanotun *homotutka-* eli *gaydar*-ilmiön⁵ olemassaolosta liittyy myös tunnistamiseen: keskustelupalstalla on varsin tavallista väittää, että homo tunnistaa toisen (13).

- (10) itse olen homo ja haluaisin tietää, **onko mitään mistä toisen voisi tunnistaa**. onko homon katseessa jotain.
- (11) Olen alkanut epäillä, että onko mieheni sittenkään (täysin) hetero. **Kertokaa mistä tunnistaa homon?**

5 Tunnistaminen ei ole helppoa, vaikka se on tärkeää heteronormatiivisissa yhteisöissä ja tilanteissa, joissa valtaosa on ei-homoseksuaaleja (homotutkasta [*gaydar*] ks. esim. Rieger ym. 2010). Homohysterian (pelko tulla tulkituksi homoseksuaaliksi) vähenemisen myötä selvästi feminiinisinä (ja siten myös homomaisina) ja maskuliinisina pidettyjen käyttäytymistapojen rajat ovat sekoittuneet, eikä homon erottaminen heterosta senkään vuoksi ole yksioikoista (ks. Anderson 2011).

- (12) **Ei taida tunnistaa gaykundia oikein mistään noin ulkoisesti.** Toki on monia ultimatehomoja, joista näkyy vaatetuksen jne perusteella välittömästi, että kyseessä on h-mies.
- (13) Varmasti on helppoa **tunnistaa. Kyllä homo aina homon tunnistaa.**

Kaapissa elämisen, olemisen ja kaapista tulemisen idiomeilla ja *kaappi*-vertausku- valla viitataan puolestaan homoseksuaalisuuden aktiiviseen peittelemiseen ja yh- teiskunnan ennakkoluuloihin reagointiin (Juvonen 2015: 18–19; ks. myös Sedgwick 1990). Aineistossa käsitellään runsaasti – joko itsen (14) tai toisten (15) – kaapissa oloa ja sieltä tuloa erilaisista näkökulmista. Esillä on usein pohdinta siitä, miten omasta seksuaalisuudesta voisi kertoa läheisille (14). Sanalistalle noussut perusmuo- toinen *kaappi*-kollokaatti on peräisin usein väärin kirjoitetusta yhdyssanasta *kaappi homo* (16).

- (14) En tiiä, **tätä menoa olen aina kaapissa perheelleni.** Meidän kesken kun pelkkä sana homo on tabu. TV:stä vaihetaan kanavaa jos siellä on joku homoseksuaali tai siitä puhutaan.
- (15) Homoushan on muotia nykypäivänä Joka toinen mies on – jos ei julkihomo, niin ainakin **kaappi sellainen.**
- (16) Itse olin naimisissa naisellisen miehen kanssa. Todellakaan en tiedä onko hän **BI/kaappi homo.**

4.2.2 Heterouteen liittyvät merkitsevimmät kollokaatit

Mielenkiintoisimpia *hetero*-sanana 20 merkitsevimmän kollokaatin listassa ovat kaksi sanaryhmää, joista syntyy omanlaisensa kuva heteroudesta (ks. taulukkoa 1): Ensinnäkin *umpi*, *pesunkestävä* ja *täys* kertovat yhtäältä joko kirjoittajan halusta represen- toida sukupuolienemmistöön kuulumistaan ja sitä, että sukupuoli-identiteetti on tai oli sataprosenttisesti heteron (17–18), tai toisaalta halusta käsitellä toisten henkilöi- den oletettua ja sataprosenttista heteroseksuaalisuutta (19). Aina keskustelu ei liity seksuaaliseen suuntautumiseen sinänsä, vaan jokin asia (kuten kodin siistinä pitäminen) liitetään homouteen, minkä yhteydessä halutaan ottaa esiin myös heteronäkö- kulma (20).

- (17) sukupuolellasi ei ole väliä tai sillä oletko hetero, lesbo... Itse olen **pesunkestävä hetero** ja tämä ei ole mikään treffi-ilmoitus :D
- (18) Seksuaalisuuden kanssa olin kriisissä. Pitkään jouduin näytellä **täys heteroa.**
- (19) Suurin osa ihmisistä on bi-seksuaaleja. Ihan **täys 100% heteroita** ei muistaakseni jonkun tutkimuksen mukaan ollut kuin pari prosenttia.

- (20) Olkoon kodin siistinä pitäminen kuinka homomaista tahansa niin riittää että itse tiedän olevani **umpi hetero** mutta pidän paikat myös liinavaatekaapin siistinä.

Toisen ryhmän muodostavat ilmaukset, joilla representoidaan joko heteroseksuaalisuuden tavallisuutta ja nimenomaan omaa tai toisen kuulumista enemmistöön (21–22) tai ylipäättään tavallisuutta ja normaaliutta, jota korostetaan esimerkiksi ihonvärillä ja rotupiirteillä. Tähän sanaryhmään kuuluvat kollokaatit ovat *normaali(t)*, *valkoinen*, *valkoihoinen*, *tylsä* ja *lihaa*, joista viimeinen esiintyy ilmauksessa *lihaa syövä* (23). Representaatioiden taustalla on ajatus siitä, että ihonväriltään valkoinen ja lihaa syövä hetero on (tylsä) normi ja muut – esimerkiksi seksuaalivähemmistöön kuuluvat – ovat poikkeuksia. Keskustelujen teemana ei ole läheskään aina seksuaalisuus, vaan viestiketjun aiheena voi olla esimerkiksi kaupungin taloudellinen tilanne, kuten on laita esimerkissä (24). Siinäkin luodaan tavallisuuden representaatiota sanoilla *valkoihoinen* ja *hetero*.

- (21) Rohkeena tyttönä vastasin katseeseen ja hymyyn ja kysyin, et taida olla ihan **normaali hetero** :) -- nyt onnellisesti yhdessä.
- (22) olen juntti ja helvetin ylpeä vielä jopa ihan **normaali heterokin**
- (23) Kun todelliset aiheet puuttuvat, niin eurooppalainen **valkoinen työssäkäyvä lihaa syövä hetero** kehittää omantunnon kolkutuksen tyhjänpäiväisesä asiasta.
- (24) Jos olet **aikuinen, valkoihoinen ja hetero** niin lopeta tuo tyhjänpäiväinen lässytys. Oulu vetää itseään kaulakiikkun kiihtyvää vauhtia, ja irtisanotut asuntovelkaiset siinä samassa.

Käännyttää ja muuttua liittyvät keskustelupalstan ulkopuolellakin runsaasti käytyyn keskusteluun siitä, voiko vähemmistöön kuuluva muuttua heteroksi tai voiko heidät sellaiseksi käännyttää (25–27). Yksi tällainen runsaasti lehdissä ja verkossakin käyty keskustelu oli Amadeus Oilingin tapaus (mm. MeNaiset 2013), josta keskusteltiin myös Suomi24-palstalla (25). Muuttumiseen ja käännyttämiseen liittyvä keskustelu on vahvasti uskontoväritteistä, ja uskontotuloa voidaan kuvata voimana, jonka avulla onnistutaan muuttumisessa. Uskontodiskurssi näkyy siis hetero-aineistosakin, vaikka ei niin voimakkaana kuin homo-aineistossa. Käännyttämisen suunta voi keskustelussa olla päinvastainenkin (27), joskaan palstan viesteistä ei voi olla aina varma, onko viesti aito kannanotto vai pikemminkin huumoria. Sama muuttumis-käännyttämis-diskurssi motivoi kollokaattilistalle nousseet elatiivimuodot *homonista* ja *lesbosta*, kuten alla olevistakin esimerkeistä havaitsee.

- (25) Tässä kaikille koskettava stoori siitä, kuinka homo nimeltä Amadeus **muuttui heteroksi ja meni naimisiin**, kun oli ensin kohdannut Jeesusken pelastajanaan ja vapahtajanaan.

(26) Mitään faktaa ei ole siitä, että **kukaan olisi homosta missään heteroksi muuttunut**. Miksi olisikaan **muuttunut**: me aikuiset kun älyämme, että homous ei ole asia, jonka tulisikaan **muuttua**.

(27) Kyllä **bodaus käännyttää heteron kuin heteron homoksi**, turha puolustella.

Myös kyky saada jälkeläisiä on diskurssi, joka on usein esillä homoudesta keskusteltaessa. Heteroparisuhteen ensiarvoisuutta voidaan perustella (oletetulla) kyvyllä saada lapsia, ja homoparisuhteita voidaan pitää luonnottomina. Adoptio-oikeus halutaan evätä juuri homoparin biologisella kyvyttömyydellä saada itse jälkeläisiä. Aineistossa tämä näkyy merkitsevien kollokaattien luettelossa erityisesti hetero-aineistossa, jossa *lisääntyä (lisääntyvät)* kollokoi *hetero*-sanana kanssa. Näissä yhteyksissä voidaan samanaikaisesti puhua vähemmistöjen ja enemmistöjen lisääntymisestä (28) tai siitä, että vain heterot voivat saada jälkikasvua (29).

(28) **Homot ja heterot lisääntyvät** ihan tasan yhtä paljon tai vähän riippumatta siitä onko homoilla avioliitto-oikeus.

(29) Heteropojat tykkää heterotyöistä ja päinvastoin. Tämä on perusjärjestys. **Vain heterot lisääntyvät**.

Liitepartikkelilliset *olla*-verbin muodot *olenkin* ja *oletko* ovat päässeet yllättävästi MI-testin tuottamalle kollokaattilistalle. Tulos on yllättävä siksi, että testi nostaa esiin pikemminkin sisältöä kuvaavia sanoja. *Olenkin*-muoto esiintyy rakenteissa *vaikka hetero olenkin / vaikka (itse) olenkin hetero*. Näissä tapauksissa oma heteroidentiteetti halutaan nostaa esiin tarkasteltaessa homoseksuaalisuutta – usein hyväksyvästä näkökulmasta – (30), omaa seksuaalisuutta ja viehtymystä samaa sukupuolta olevaan (31) tai jotain aivan muuta asiaa (32). *Oletko* puolestaan esiintyy taajaan erilaisissa kysymysketjuissa, joissa kysellään mieltymyksiä (33). Toisaalta muoto esiintyy myös pohtivissa ja neuvovissa viesteissä, joissa aiheena on epävarmuus seksuaalisuudesta (34).

(30) Mutta ei siinä mitään, **vaikka hetero olenkin**, niin ei mulla homoja vastaan mitään ole.

(31) Katsos **vaikka hetero olenkin** niin en voi väittää että ettenkö esim unissani olisi fantasioinut miesten välisestä.

(32) kylläpä mullakin on honda **vaikka hetero olenkin**

(33) 17. mikä on sun unelmien ammatti? Tanssija
18. onko lapsia? haluaisitko? on,yksi tyttö. Lisää haluan joskus myöhemmin
19. **oletko hetero**, homo, lesbo, bi.....? **hetero**

- (34) Jos haluat vastauksen siihen **oletko täysin hetero..** niin siihen pystyt loppujen lopuksi vastaamaan vain sinä itse.

Taulukon 1 uskontoon viittaavat *risti* ja *Golgatan* nousevat kollokaattilistalle taajaan lähes samanlaisena toistetusta (spämmätystä) viestistä, joka esiintyy lähinnä uskontoteemaisissa keskusteluissa: *Heprealainen Yeshua mutttui nimenomaan kidutuksen takia Golgatan ristillä homosta heteroksi* – –. Kollokaatti *terveisin* puolestaan selittyy joidenkin kirjoittajien tavalla lopettaa viesti lopputervehdykseen (*Terveisin entinen hetero, nykyisin rakastunut ihminen*). Myös lyhenne *ps* on tavallinen viestin lopussa: oma suuntautuminen voidaan representoida loppuhuomautuksessa, jos sitä ei ole tehty viestissä aiemmin, ja näin viesti saa näkökulman, josta se pitää tulkita:

- (35) Minulla on kuulemma kauniit kasvopiirteet ja siksi ajattelin että voisi kokeilla sellaista siilikampausta. **Ps: olen hetero nainen**

4.3 Diskurssiprosodioista diskursseihin: yhteiset ja erottavat diskurssit

Edellä merkitsevimpien kollokaattien käsittelystä kävi ilmi, että homouden yhteydessä esiintyy erityisesti vihapuhetta tuottavia ilmauksia, mutta myös luontaisuuden, tunnistamisen, kaapissa olemisen ja uskontoon liittyviä diskursseja. Heterouden yhteydessä olivat vuorostaan tavallisia kollokaatit, jotka assosioituvat tavallisuuteen ja absoluuttisuuteen sekä myös suuntautumisen muuttumiseen ja kykyyn lisääntyä. Näistä vihapuhe- ja uskontodiskurssit ovat tulleet esille jo aiemmassakin tutkimuksessa (Jantunen 2018). Jo siis muutamien merkitsevien kollokaattien avulla voi päästä kiinni laajassa aineistossa esiintyviin diskursseihin.

Vaikka edellä lähtökohtana olivat yksittäiset merkitsevät kollokaatit, kävi ilmi, että joitakin kollokaatteja oli mahdollista ryhmitellä merkityskokonaisuuksiksi. Ryhmittelyn avulla päästään leksikaalisen tason ohi ja voidaan keskittyä laajempiin merkitysrakenteisiin ja sitä kautta diskurssipiirteisiin. Tarkastelenkin seuraavaksi, miten kattavampi määrä (200) kollokaatteja jakautuu hakusanoittain merkitysryhmiin ja millaisia diskursseja verkkokeskusteluissa voi päätellä olevan näiden ryhmittelyjen pohjalta. Ryhmittely ja siten myös diskurssiprosodian analyysi perustuu ajatukseen, jonka mukaan yhtä tiettyä diskurssia tuotetaan ja toisinnetaan useiden eri kielenpiirteiden avulla (tässä kollokaateilla), mikä on yleensäkin diskurssintutkimukselle tyypillinen lähtökohta.

Taulukossa 2 on lueteltu nimenomaan *homo*-sanana lähikontekstissa olevia merkitseviä kollokaatteja; mikäli merkitsevien kollokaattien joukossa on sanasta useampia muotoja, kollokaatti on merkitty KAPITEELEIN. Kollokaateista erottuvat selvästi omaksi ryhmäkseen nimittelyt (*kusipää*, *pelle*), voimasanat (*vittu*) ja väkivaltaisia tekoja merkitsevät sanat (*tappaa*, *hirteen*). Näiden runsas määrä viittaa vahvasti

siihen, että kun homoja ja homoutta käsitellään, keskusteluissa tuotetaan nimittely- ja väkivaltadiskurssia, joka on ylipäättään hyvin tyypillinen verkkovihan muoto (ks. Pöyhtäri ym. 2013:19, 49) ja myös aiemmin havaittu homouteen liittyvä diskurssi (Pascoe 2005; Jantunen 2018).

Myös ihmisoikeuskysymykset ja suvaitsevaisuus ovat voimakkaasti läsnä *homo*-sanankontekstissa. Toisin kuin väkivaltadiskurssi ihmisoikeuskysymykset eivät nousseet esiin edellä aivan merkittävempien kollokaattien tarkastelussa, mutta diskurssi paljastuu, kun aineistoa analysoidaan laajemman otoksen avulla. Oikeus- ja suvaitsevaisuusdiskurssia tuotetaan monella tapaa: yhtäältä käsitellään syrjintää ja tuomitsemista (*syrjitään, tuomita*), toisaalta sietämistä ja hyväksymistä (*sietää, hyväksyy*). Tasavertaisuuden ja oikeuksien on aiemminkin havaittu liittyvän homoseksuaalisuuden käsittelyyn (ks. Baker 2005; Bachmann 2011).

Myöskään tiettyjen ihmisten ja ihmisryhmien yhtäaikainen käsittely homouden yhteydessä ei näkynyt merkittävempien kollokaattien listalla. Baker (2005: 69–70) on iltapäivälehtiä tarkastellessaan huomannut, että homot (*gays/homosexuals*) liitetään usein muihin vähemmistöryhmiin, rikollisiin ja etnisiin ryhmiin tai muuten kirjoittajan näkökulmasta ongelmallisiin ryhmiin. Näin on laita Suomi24-keskusteluissakin: homojen rinnalle nousevat neutraalit ryhmät, kuten ihmiset (*ihmisinä, väestöstä*), mutta myös *feministit, mamut, natsit, ruotsalaiset* (myös *sveikko*), *sivarit* ja *vammaiset* sekä lisäksi yksittäiset henkilöt, kuten *Antti* ja *Mannerheim*. Heteron kontekstissa ryhmää tarkoittavia ilmauksia ei ole, joskin viittauksia yksittäisiin henkilöihin, kuten viestiketjun ensimmäiseen kirjoittajaan eli *aloittajaan*, kirjoittajaan itseän (*minäkin*) ja *Niinistöön*, kuitenkin on. (Ks. myös Jantunen 2018.)

TAULUKKO 2. Etupäässä homouteen liittyvät diskurssiprosodiat.

homo		hetero
HAISTA, <i>hirteen, hyi, ime, kivittää, kusipää, painu, pelle, runkkari</i> , TAPPAA, <i>vedä (käteen, vittu päähän)</i> , VIHATA, VITTU, <i>voltti</i>	nimittely, vihapuhe, väkivalta	-
HYVÄKSYÄ, <i>hyväksytty, hyväksyminen, ihmisoikeudet</i> , OIKEUS, <i>puolustavat, sietää, syrjinnän, syrjitään, TUOMITA, vaatia, vastustavat</i>	oikeudet ja suvaitsevaisuus	OIKEUS
<i>Antti, feministit, ihmisinä,</i> MAMU, <i>Mannerheim, me,</i> <i>natsit, presidentiksi, pressaksi,</i> <i>ruotsalaiset, SIVARI, sveikko,</i> <i>vammaisia, väestöstä, vastustajat</i>	henkilöt ja ryhmät	<i>aloittaja, minäkin, Niinistö,</i> <i>presidentti</i>

Aiemmin (ks. taulukkoa 1) käsitellyt heterouteen liittyvät aitouden ja absoluutisuuden diskurssit sekä tavallisuuden diskurssi saavat vahvistusta, kun laajem-

masta kollokaattiryppästä luodaan ryhmiä (taulukko 3). Aitous ja absoluuttista heteroutta toisinnetaan Suomi24:ssä hyvin monenkirjavasti: *ihka aito hetero, ihan oikeita heteroita, täysin puhdas hetero* jne. Homouden absoluuttisuutta representoidaan ainoastaan *täys*-kollokaatilla (*Ent hetsku ja nyt täys homo*). Myös edellä käsitellyä tavallisuutta representoidaan aineistossa hyvin monella tapaa. Aiemmin esiteltyjen lisäksi myös seuraavat ovat tyypillisiä: *ootko ite joku normi hetero vai?; kaveri on sellainen perus hetero; Eli Suomalainen, aito alkuperänen, Hetero*.

Tavallisuuden diskurssia tuottavat omalla tavallaan myös monet määrään liittyvät kollokaatit, kuten *moni (moni sataprosenttinen hetero), useimmat, valtaosa (valtaosa on vielä heteroja), enemmistö (hetero enemmistö kuitenkin olettaa)*, sillä ne implikoivat keskustelussa usein sen, että heteroiden määrä on suuri ja että merkittävä osa heistä ajattelee tai toimii tietyllä tavalla. Siten heidän suuri määränsä, ajattelunsa ja toimintansa luovat heteronormatiivisen normin.⁶ On myös muistettava – kuten edelläkin tuli ilmi – että tavallisuutta korostavat ilmaukset yhdessä *hetero*-sanalla kanssa eivät tuota tavallisuuden representaatiota pelkästään heteroista, vaan yhdessä *heteron* kanssa ilmaukset korostavat myös sitä, että esimerkiksi kirjoittaja mieltää itsensä tavalliseksi kansalaiseksi. Heteroseksuaalisuus onkin silloin yksi tavallisuutta indikoiva ominaisuus eikä sukupuoli suuntautuminen ole keskustelussa sinänsä oleellista (*Suomessa normaali keski-ikä lähenevä mies, valkoihoinen hetero, pärjää kaikissa tilanteissa mainiosti*.)

TAULUKKO 3. Etupäässä heterouteen liittyvät diskurssiprosodiat.

homo	hetero
täys	aito, oikea absoluuttisuus
	aito, oikeita, pesunkestävä, puhdas, puhtaasti, täys, täysi, täysin, umpi <i>kunnon, NORMAALI, normi, perus, SUOMALAINEN, TAVALLINEN, tylsä, valkoihoinen, valkoinen; enemmistö, moni, useimmat, valtaosa, yleisempää</i>
	tavallisuus

Tietyt diskurssiprosodiat myös yhdistävät homoutta ja heteroutta; nämä on lueteltu taulukossa 4. Käsittelen niistä esimerkkinä vain perhettä ja suhdetta merkitseviä kollokaatteja eli perhe- ja suhdediskursseja, jotka osoittautuvat analyysin perusteella tavallisiksi sekä homoista että lesboista keskusteltaessa. Muita ovat rikos- ja sairausdiskurssit, seksuaaliseen suuntautumiseen liittyvä diskurssi, jonka on jo toisessa yhteydessä (Jantunen 2018) havaittu liittyvän erityisesti heteroseksuaalisuuteen, sekä uskontodiskurssi, joka liittyy tyypillisesti homouteen (Bachmann 2011; Jantunen 2018). Näitä kaikkia yhdistää muun muassa se, että samassa keskustelussa puhutaan

6 Vaihtoehtoisesti määrään voisi myös tulkita omaksi diskurssiprosodiakseen.

sekä homoista että heteroista, mikä osittain selittää sen, että diskurssit ovat ainakin osin jaettuja. Kuten kollokaateista kuitenkin näkee, yksittäiset kollokaatit voivat olla erilaisia, ja siten diskurssin alla voi olla vähemmistö- ja enemmistöspesifejä diskursseja.

Yhteiset diskurssiprosodiat eivät ole yllättäviä, sillä esimerkiksi tasa-arvoiseen avioliittoon liittyvässä keskustelussa on luontaista keskustella sekä homoista ja heteroista että heidän oikeuksistaan ja tehdä vertailuja näiden ryhmien välillä. Niinpä kollokaatti *avioliitto* on molempien merkitsevä kollokaatti (**Homojen avioliitto ei ole heteroilta mitenkään pois** –). Myös *lapsi* esiintyy monissa diskursseissa, esimerkiksi niissä, joissa käsitellään lapsen seksuaalista suuntautumista (*Jokainen lapsi on hetero kunnes toisin todistetaan*), ja myös diskursseissa, joissa käsitellään lapsen asemaa ja kasvamista perheessä (*moni heteron lapsi elää ilman isää tai äitiä; Homojen lapsi ei voi elää seksuaalineutraalisti*).

Sekä *varattu-* että *sinkku-*kollokaatit löytyvät muun muassa seuranhakuilmoituksista (*komeaksi haukuttu 188/85, tumma mies olen. Varattu ja hetero; Oon kolmikymppinen sinkku hetero nainen* –) ja keskusteluista, joissa käsitellään toisten henkilöiden suhdestatusta ja seksuaalisuutta (*mutta hän kyllä hetero ja varattu miehen kanssa*). Nämä kollokaatit ovat tyypillisiä vain hetero-aineistossa. Homojen adoptio-oikeuden käsitteleminen eduskunnassa ja mediassa näkyy myös Suomi24:ssä (*ei tuo homojen adoptio niin härskiä ole*)⁷ samoin kuin avioliittolain yhteydessä käyty vilkas vihkimiseen liittyvä keskustelu (*– siksi homoja ei voi vihkiä avioliittoon jumalan edessä*).

Verkon seksuaalivähemmistöjä koskeva keskustelu toistaa pitkälti samoja teemoja kuin mediassakin on nähty: adoptio-oikeus, mahdollisuus avioliittoon ja oikeus vihkimiseen puhuttavat myös verkon kansalaiskeskusteluissa. Heteroihin liittyvä keskustelu on osin hyvin erilaista, mikä kummunee jo yhteiskunnassa vallitsevasta tilanteesta, joka ruokkii yhtäältä polaarista asetelmaa ja toisaalta enemmistön valta-asetelmaa sekä vähemmistöön kohdistuvaa, ulkoapäin suuntautuvaa katsomista ja sen oikeudellistumista. Silloinkin kun diskurssit osoittautuvat jaetuiksi, niissä on havaittavissa erilaisia painotuksia ja teemoja. Verkkokeskustelujen tutkimisen avulla päästään tarkastelemaan sitä, millaisia kansalaiskeskustelut ovat ja miten ne saattavat muuttua (Lagus ym. 2016: 40). On kuitenkin muistettava, ettei Suomi24-aineisto anna kuvaa kaikesta Suomessa käytävästä keskustelusta tai edes kaikesta verkkokeskustelusta. Aineistoa on joka tapauksessa pidettävä otoksena, joskin se on hyvin laaja.

7 Osa nominatiivimuotoisista adoptio-kollokaateista on ilman yhdysmerkkiä väärin kirjoitetuista yhdyssanoista (adoptio oikeus).

TAULUKKO 4. Homoutta ja heteroutta yhdistävät diskurssiprosodiat (hakusanaspesifit kollokaatit lihavoitu, yhteiset kollokaatit lihavoimatta, monia taivutusmuotoja edustavat lemmat kapiteelein).

homo		hetero
<i>adoptio, avio, avioliitot, lapsista, LISÄÄNTYÄ, vihkiminen, VIHKIÄ</i>	perhe, suhde	AVIOLIITTO, <i>kasvattaa, lapsista, lisääntyvät, naimisissa, sinkku, varattu</i>
<i>eheyttää, eheytyä, parantua, PEDOFIILI, SAIRAS, sairaus</i>	rikos, sairaus	<i>eheyttää, eheytyä, kidutuksen, käännyttää, kääntyä, PEDOFIILI, raiskaa, TERVE</i>
<i>avoimesti, bi, biseksuaaleja, HETERO, hintti, kaappiin, LESBO, LESBOPARI, lesbous, LESPO, pervo, trans, TRANSU</i>	suuntautuminen	<i>bi, BISEKSUAALI, bissejä, HOMO, LESBO, pride, trans, transu</i>
ATEISTI, <i>juutalaiset, luonut, naispappeja, ristillä, SYNTI</i>	uskonto	<i>ateisti, golgatan, KRISTITTY, ristillä, syntisiä</i>

5 Pohdintaa

Vaikka korpusavusteisuus voi tuoda diskurssintutkimukseen laajojen aineistojen ja korpusmenetelmien muodossa etuja, ei menetelmiä yhdistävä CADS ole luonnollisestikaan vailla puutteita. Martin (1999: 52), Hunston (2002: 110), Biber ja Jones (2005: 174), Paltridge (2006: 174–176) ja Flowerdew (2008: 115–116) luettelevat rajoitteita, jotka CADS:ssa on otettava huomioon ja pyrittävä myös ratkaisemaan: 1) Yksittäisen kielentuottajan tekemät ratkaisut ja käyttämät kielenpiirteet häviävät tutkimuksessa taustalle. 2) Tutkimus kohdistuu helposti vain leksikaalisiin ilmiöihin, ja muut manuaalisessa tutkimuksessa tavoitettavat diskurssipiirteet voivat jäädä havaitsematta. 3) Tutkimus kohdistuu korpustutkimuksen luonteen vuoksi atomistisesti yksittäisiin kielenpiirteisiin ja kielisysteemiin, ja teksti ja laajempi sosiaalinen konteksti saattavat jäädä huomiotta tuloksia tulkittaessa. Koska analyysi perustuu ensikädessä laajaan tekstimassaan, ei yksittäisen kirjoittajan ääni korostu tutkimuksessa. Toisaalta mikään ei kuitenkaan estä ottamasta yksittäistä tekstiä laadulliseen ja siten tarkempaan tarkasteluun, kun menetelmälle voidaan osoittaa esimerkiksi korpuksesta tai yhteiskunnallisesta merkityksestä nouseva relevanssi. Leksikaalisiin piirteisiin keskittymistä voidaan tarvittaessa vähentää myös juuri tekstien lähiluvun avulla tai vaihtoehtoisesti käyttämällä annotoituja aineistoja, jolloin päästään kiinni haluttuihin rakennepiirteisiin. Kontekstittomuus on ainakin osittain ratkaistavissa sitomalla tutkimus huolellisemmin yhteiskunnalliseen keskusteluun sekä kokoamalla ja hyödyntämällä esimerkiksi aineistojen metatiedot nykyistä tarkemmin (ks. esim. Tribble 2002).

Menetelmien yhdistämisen etuina voidaan vuorostaan Biberin ja Jonesin (2005: 174), Leen (2008: 92–94), Partingtonin ja kollegoiden (2013: 331) sekä Lehdon (2018: 85–88) mukaan pitää 1) pääsyä laajoihin, edustaviin ja sisällöllisesti varioiviin aineistoihin, 2) intuitioon perustuvan tulkinnan vähenemistä, 3) määrälliseen, toistaiseen, yleistettävään ja toistettavaan tietoon perustuvan argumentoinnin lisääntymistä sekä 4) automaattisen aineistovetoisen analyysin mahdollistumista. Käsillä olevan artikkelin esimerkkitutkimuksessa toteutuvat kolme ensin mainittua kohtaa ja aiemmassa samaa teemaa käsittelevässä artikkelissani (Jantunen 2018) myös aineistovetoisuus. Näiden piirteiden osalta CADS eroaa selvästi niin sanotusta perinteisestä diskurssintutkimuksesta. Sen sijaan yhtäläisyydet, kuten tutkimuksen kohdistaminen todelliseen kielenkäyttöön, todellisen kielen käyttö aineistona, aineistovetoisuus ja säännönmukaisuuksien etsiminen, ovat mahdollistaneet menetelmien yhdistämisen; pontimena tälle on ollut tarve hyödyntää diskurssintutkimuksessa laajempia aineistoja ja kvantitatiivisia menetelmiä.

Biber ja Jones (2005: 174) kuitenkin korostavat, että korpusavusteinen tutkimus ei voi korvata perinteistä diskurssintutkimusta vaan hedelmällisintä on yhdistää eri menetelmiä: käyttämällä tilastollis-määrällisiä ja laadullisia menetelmiä on mahdollista paljastaa sellaisia diskurssipiirteitä, jotka ovat piilossa ja vähemmän itseltään selviä (Partington ym. 2013: 11). Edellä tehdylle empiiriselle analyysille olisikin hyödyksi, että määrällisestä analyysistä edetään vielä pitemmälle kohti laadullista analyysia. Tutkimusta on mahdollista syventää laadulliseen suuntaan muun muassa valitsemalla yksittäinen kollokaatio- tai diskurssiprosodia-analyysissa esiin nousnut diskurssi tarkempaan, laadulliseen analyysiin. Mahdollista olisi myös poimia aineistosta joitakin yksittäisiä tekstejä eli viestejä ja viestiketjuja laadullisesti analysoitavaksi.

Kirjallisuus

- Aller Media Oy 2014. *Suomi 24 -korpus (2016H2)* [tekstikorpus]. Kielipankki. <http://urn.fi/urn:nbn:fi:lb-2017021506>.
- Anderson, E. 2011. The rise and fall of western homophobia. *Journal of Feminist Scholarship*, 1, 80–94. <http://www.jfsonline.org/issue1/articles/anderson/>.
- Bachmann, I. 2011. Civil partnership – “gay marriage in all but name”: a corpus-driven analysis of discourses of same-sex relationships in the UK parliament. *Corpora*, 6 (1), 77–105. <https://doi.org/10.3366/cor.2011.0005>.
- Baker, P. 2004a. Querying keywords. Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics*, 32 (4), 346–359. <https://doi.org/10.1177/0075424204269894>.
- Baker, P. 2004b. ‘Unnatural acts’: discourses of homosexuality within the house of lords debates on gay male law reform. *Journal of Sociolinguistics*, 8 (1), 88–106. <https://doi.org/10.1111/j.1467-9841.2004.00252.x>.
- Baker, P. 2005. *Public discourses of gay men*. London: Routledge.

- Baker, P., C. Gabrielatos, M. Khosravini, M. Krzyzanowski, T. McEnery & R. Wodak 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19 (3), 273–306. <https://doi.org/10.1177/0957926508088962>.
- Barnbrook, G. 1996. *Language and computers. A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- Biber, D. 1993. Representativeness in corpus design. *Literary and Linguistic Computing*, 8 (4), 243–257. <http://otipl.philol.msu.ru/media/biber930.pdf>.
- Biber, D., S. Conrad & R. Reppen 1998. *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D. & J. Jones 2005. Merging corpus linguistic and discourse analytic research goals: discourse units in biology research articles. *Corpus Linguistics and Linguistics Theory*, 1–2, 151–182.
- Cheng, W. 2012. *Exploring corpus linguistics. Language in action*. London: Routledge.
- Fairclough, N. 2003. *Analysing discourse. Textual analysis for social research*. London: Routledge.
- Flowerdew, L. 2008. Corpora and context in professional writing. Teoksessa V. K. Bhatia, J. Flowerdew & R. H. Jones (toim.) *Advances in discourse studies*. London: Routledge, 115–127.
- Hakulinen, A., F. Karlsson & M. Vilkuna 1980. *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Helsinki: Helsingin yliopisto.
- Heikkinen V., O. Lehtinen & M. Lounela 2001. Kuvia kirjoitetusta suomesta. *Kielikello* 3/2001. <http://www.kielikello.fi/index.php?mid=2&pid=11&aid=1279>.
- Heikkinen, V., M. Lounela & E. Voutilainen 2012. Aineistot ja niiden käyttö tekstilajitutkimuksessa. Teoksessa V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiililä & M. Lounela (toim.) *Genreanalyysi. Tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 325–350.
- Hoey, M. 2005. *Lexical priming. A new theory of words and language*. London: Routledge.
- Hoey, M. 2007. Lexical priming and literary creativity. Teoksessa M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert (toim.) *Text, discourse and corpora. Theory and analysis*. London: Continuum, 7–29.
- HS 2017 = Tutkijat loivat tekoälyn, joka tunnistaa seksuaalisen suuntautuneisuuden kasvoista – ihmistä tarkempi ”homotutka” herättää kritiikkiä. Helsingin Sanomat 8.9.2017. <https://www.hs.fi/tiede/art-2000005360520.html> [luettu 29.1.2018].
- Hunston, S. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. 2007. Semantic prosody revisited. *International Journal of Corpus Linguistics*, 12 (2), 249–268. <https://doi.org/10.1075/ijcl.12.2.09hun>.
- Jantunen, J. H. 2004. *Synonymia ja käännössuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin*. Joensuu: Joensuun yliopiston humanistinen tiedekunta.
- Jantunen, J. H. 2009. Ei pelkästään mielikuvituksen puutteen vuoksi. – Kieliaineistojen systemaattinen käyttö kielentutkimuksessa. *Virittäjä*, 113 (1), 101–113. <https://journal.fi/virittaja/article/view/4171>.
- Jantunen, J. H. 2018. Homot ja heterot Suomi24:ssä: analyysi digitaalisista diskursseista. *Puhe ja kieli*, 38 (1), 3–22. <https://doi.org/10.23997/pk.65488>.
- Jantunen, J. H. & S. Brunni 2012. Morfologinen priming ja fraseologia vieraan kielen oppimisessa: korpustutkimus oppijansuomesta. *Lähivörtlusi. Lähivertailuja*, 22, 71–100. <https://doi.org/10.5128/LV22.03>.
- Jones, S. & J. Sinclair 1974. English lexical collocations. A study in computational linguistics. *Cahiers de Lexicologie*, 24, 15–61.

- Juvonen, T. 1997. Normatiivisen hyvät harhat. Teoksessa J. Lehtonen, J. Nissinen & M. Socada (toim.) *Hetero-olettamuksesta moninaisuuteen: lesbot, homot, bi- ja transihmiset sosiaali- ja terveystalveluiden asiakkaina*. Helsinki: Edita, 11–17.
- Juvonen, T. 2015. *Kaapista kaapin päälle*. Tampere: Vastapaino.
- Karjalainen, A.-M., M. Luodonpää-Manni & V. Laippala 2017. Hyvinvointivaltio ja kielitietoisuus: hyvinvoinnin diskurssit neljän suurimman puolueen eduskuntavaaliohjelmissa. Teoksessa S. Lomaa, E. Luukka & N. Lilja (toim.) *Kielitietoisuus eriarvoistuvassa yhteiskunnassa*. AFinLAn vuosikirja 2017. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 114–132. <https://journal.fi/afinlavk/article/view/60832>.
- Kosinski, M. & Y. Wang 2018. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*, 114 (2), 246–257. <https://psyarxiv.com/hv28a/>.
- Krishnamurthy, R. 1996. ETHNIC, RACIAL AND TRIBAL. The language of racism? Teoksessa C. R. Caldas-Coulthard & M. Coulthard (toim.) *Texts and practices: readings in critical discourse analysis*. London: Routledge, 129–149.
- Lagus, K., M. Pantzar, M. Ruckenstein & M. Ylisiurua 2016. *Suomi24. Muodonantoa aineistolle*. Helsinki: Helsingin yliopisto. https://blogs.helsinki.fi/citizenmindscape/files/2016/05/257383_HY_VALT_suomi24_muodonantoa_aineistolle.pdf.
- Lee, D. 2008. Corpora and discourse analysis. New ways of doing old things. Teoksessa V. K. Bhatia, J. Flowerdew & R. H. Jones (toim.) *Advances in discourse studies*. London: Routledge, 86–99.
- Leech, G. 2000. Grammars of spoken English. New outcomes of corpus-oriented research. *Language Learning*, 50 (4), 675–724.
- Lehto, L.-M. 2018. *Korpusavusteinen diskurssianalyysi japaninsuomalaisten kielipuheesta*. Oulu: Oulun yliopisto. <http://jultika.oulu.fi/files/isbn9789526219097.pdf>.
- Luukka, M.-R. 2000. Näkökulma luo kohteen: diskurssintutkimuksen taustaletukset. Teoksessa K. Sajaava & A. Piirainen-Marsh (toim.) *Kieli, diskurssi & yhteisö*. Jyväskylä: Soveltavan kielentutkimuksen keskus, 133–160.
- Martin, J. R. 1999. Grace: the logogenesis of freedom. *Discourse Studies*, 1 (1), 29–56.
- McEnery, T., R. Xiao & Y. Tono 2006. *Corpus-based language studies. An advanced resource book*. London: Routledge.
- MeNaiset 2013. *Amadeus Oilinki: "Eheydyin homosta heteroksi"*. https://www.menaiset.fi/artikkeli/ajankohtaista/ihmiset/amadeus_oilinki_eheydyin_homosta_heteroksi [luettu 27.1.2018].
- Meta-Share 2016. *The Suomi24 Corpus*. <http://metashare.csc.fi/repository/browse/the-suomi-24-corpus-2015h1/4db73da85ce11e4912c005056be118ea699d93902fa49d69b0f4d1e692dd5f1/2> [luettu 8.1.2017].
- Oikeusministeriö 2016. *"Usein joutuu miettimään, miten pitäisi olla ja minne olla menemättä". Selvitys vihapuheesta ja häirinnästä ja niiden vaikutuksista eri vähemmistöryhmiin*. Helsinki: Oikeusministeriö. <http://oikeusministerio.fi/julkaisu?pubid=URN:ISBN:978-952-259-496-9> [luettu 1.3.2017].
- Pace-Sigge, M. & K. J. Patterson 2017. Introduction. Teoksessa M. Pace-Sigge & K. J. Patterson (toim.) *Lexical priming. Applications and advances*. Amsterdam: John Benjamins, XI–XXIII.
- Paltridge, B. 2006. *Discourse analysis*. London: Continuum.
- Partington, A., A. Duguid & C. Taylor 2013. *Patterns and meanings in discourse. Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam: John Benjamins.
- Partington, A. & A. Marchi 2015. Using corpora in discourse analysis. Teoksessa D. Biber & R. Reppen (toim.) *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press, 216–234.

- Partington, A. & J. Morley 2004. At the heart of ideology: word and cluster/bundle frequency in political debate. Teoksessa B. Lewandowska-Tomaszczyk (toim.) *PALC 2003: practical applications in language computers*. Frankfurt: Peter Lang, 179–192.
- Pascoe, C. J. 2005. 'Dude, you're a fag': Adolescent masculinity and the fag discourse. *Sexualities*, 8 (3), 329–346. <https://doi.org/10.1177/1363460705053337>.
- Pietikäinen, S. & A. Mäntynen 2009. *Kurssi kohti diskurssia*. Tampere: Vastapaino.
- Pöyhtäri, R., P. Haar & P. Raittila 2013. *Vihapuhe sananvapautta kaventamassa*. Tampere: TUP.
- Rieger, G., J. Linsenmeier, L. Gygax, S. Garcia & J. Bailey 2010. Dissecting "gaydar": accuracy and the role of masculinity–femininity. *Archives Sexual Behavior*, 39 (1), 124–140. <https://doi.org/10.1007/s10508-008-9405-2>.
- Sedgwick, E. K. 1990. *Epistemology of the closet*. Berkeley, CA: University of California Press.
- Sinclair, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. 1996. The search for units of meaning. *Textus*, IX, 75–106.
- Stubbs, M. 1996. *Text and corpus analysis*. Oxford: Blackwell.
- Stubbs, M. 1997. Whorf 's children: critical comments on critical discourse analysis. Teoksessa A. Ryan & A. Wray (toim.) *Evolving models of language*. Clevedon: Multilingual Matters, 100–116.
- Stubbs, M. 2001. *Words and phrases. Corpus studies of lexical semantics*. Oxford: Blackwell.
- Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Tognini-Bonelli, E. 2004. Working with corpora: issues and insights. Teoksessa C. Coffin, A. Hewings & K. O'Halloran (toim.) *Applying English grammar. Functional and corpus approaches*. London: Arnold, 11–24.
- Tribble, C. 2002. Corpora and corpus analysis: new windows on academic writing. Teoksessa J. Flowerdew (toim.) *Academic discourse. London: Longman*, 131–149.
- Tyrkkö, J. & I. Taavitsainen 2012. Laajojen tekstimassojen luokitteluperusteista. Teoksessa V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiililä & M. Lounela (toim.) *Genreanalyysi. Tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 309–319.
- Ädel, A. & R. Reppen 2008. The challenges of different settings. An overview. Teoksessa A. Ädel & R. Reppen (toim.) *Corpora and discourse: the challenges of different settings*. Amsterdam: John Benjamins, 1–6. <https://doi.org/10.1075/scl.31.01ade>.