

**Synonyymien EHKÄ ja MAHDOLLISESTI kollokaatit
ja semanttiset preferenssit**

Maisterintutkielma

Silja Pirkola

Suomen kieli

Kielten laitos, Jyväskylän yliopisto

Tammikuu 2016

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanistinen	Laitos – Department Kielten laitos
Tekijä – Author Silja Pirkola	
Työn nimi – Title Synonyymien EHKÄ ja MAHDOLLISESTI kollokaatit ja semanttiset preferenssit	
Oppiaine – Subject Suomen kieli	Työn laji – Level Pro gradu -tutkielma
Aika – Month and year Tammikuu 2016	Sivumäärä – Number of pages 105 sivua + liitteet
Tiivistelmä – Abstract <p>Tutkielmassa tarkastellaan lähimerkityksisten modaalisanojen EHKÄ ja MAHDOLLISESTI synonymiaa. Työssä selvitetään ensinnäkin, millaisia kollokaatteja modaalisanoilla on eli millaisia sanoja tyypillisesti esiintyy sanojen kontekstissa. Toiseksi tutkitaan, millaisia semanttisia preferenssejä sanoilla on eli millaisia merkitysryhmiä sanojen kollokaatit tyypillisesti edustavat. Tavoite on näitä kontekstuaalisia leksikaalisia piirteitä tutkimalla selvittää, miten synonyymit eroavat toisistaan, ja toisaalta myös tarkastella sitä, mitä yhteisiä piirteitä sanoilla on. Tutkielman taustateorianana on siten synonyymiatutkimus. Työ pohjautuu myös kontekstuaalisen semantiikan ja fraseologisen tutkimuksen näkemyksiin, joiden mukaan kieltä tulee tutkia kontekstilähtöisesti. Tutkimusmenetelmänä työssä on korpuslingvistiikka ja aineistona sähköinen sanomalehtien vuosikertoja sisältävä korpus, Suomen kielen tekstikokoelma. Aineistoa analysoidaan tilastollisin menetelmin.</p> <p>Tutkimus paljastaa, että tärkeimpiä sanoja yhdistäviä kontekstuaalisia leksikaalisia tekijöitä sanomalehtikielessä ovat kymmenen tilastollisesti merkitsevää kollokaattia ENSI, JOKIN, JOKU, JOPA, MYÖHEMMIN, TAI, TULEVAISUUS, TULLA, VIELÄ ja VOIDA. Kummankin sanan kontekstissa esiintyy melko tasaisesti eri merkitysryhmiä, mutta tyypillisimpiä sanoille yhteisiä merkitysryhmiä ovat 'aika', 'prosessi' ja 'määrä'. Synonyymeja erottavia piirteitä puolestaan ovat EHKÄ-sanan huomattavasti suurempi adjektiivi- ja adverbikollokaattien määrä sekä mentaalisia verbejä edustavien verbikollokaattien määrä. Modaalisanon erottavat toisistaan myös MAHDOLLISESTI-sanon synonyymiaan vahvemmat semanttiset preferenssit 'yhteiskunnallisuus, institutionaalisuus' ja 'konkreettiset oliot' sekä vastaavasti EHKÄ-sanon vahvempi preferenssi 'mentaalisuus'. Tutkimustulosten perusteella näyttää siltä, että EHKÄ edustaa synonyymiaan tyypillisemmin subjektiivista modaalisuutta ja yksilön tietoa ja MAHDOLLISESTI puolestaan objektiivista modaalisuutta ja yhteisön tietoa. Kaiken kaikkiaan tutkimuksen tuloksia on mahdollista hyödyntää kielenopetuksessa sekä leksikografiassa.</p>	
Asiasanat – Keywords synonymia, fraseologia, kontekstuaalisuus, modaalisuus	
Säilytyspaikka – Depository	
Muita tietoja – Additional information	

SISÄLLYS

1 JOHDANTO	1
2 TEORIATAUSTA	3
2.1 Kontekstuaalinen semantiikka	3
2.1.1 Kontekstuaalisen semantiikan lähtökohdat	3
2.1.2 Kontekstuaalinen valinta	5
2.2 Fraseologia	7
2.2.1 Fraseologinen yksikkö	7
2.2.2 Kollokaatio	11
2.2.3 Semanttinen preferenssi	13
2.3 Synonymia	15
2.4 Modaaliset partikkelit ja adverbit	21
2.4.1 Modalisuus	21
2.4.2 Modaalisanojen EHKÄ ja MAHDOLLISESTI sanaluokka	22
3 AINEISTO JA METODIT	24
3.1 Aineisto	24
3.2 Perusteet tutkittavien lekseemien valinnalle	26
3.3 Fraseologisen yksikön tutkiminen korpuslingvistisin menetelmin	29
3.3.1 Tilastolliset testit	29
3.3.2 Kollokaatioväli	31
3.3.3 Kollokaatioanalyysin aineiston muodostaminen	33
3.3.4 Semanttisen preferenssin analyysin aineiston muodostaminen	35
3.3.5 Aineisto ja intuitio tutkimusmenetelmien perustana	39
4 KOLLOKAATIT	41
4.1 Tilastollisesti merkitsevimmät kollokaatit	41
4.1.1 T-testi	41
4.1.2 MI-testi	43
4.1.3 Eri testien antama kuva modaalisanojen kollokoinnista	45
4.2 Eri sanaluokkien kollokaatit	48
4.2.1 Nominikollokaatit	49
4.2.1.1 Substantiivikollokaatit	49
4.2.1.2 Adjektiivikollokaatit	52
4.2.1.3 Numeraalikollokaatit	55

4.2.1.4 Pronominikollokaatit	56
4.2.2 Verbikollokaatit	59
4.2.3 Adverbikollokaatit	63
4.2.4 Partikkeli- ja adpositiokollokaatit	68
5 SEMANTTINEN PREFERENSSI	73
5.1 Luokittelun perusteet	73
5.2 Semanttisen preferenssin analyysin tulokset	82
6 PÄÄTÄNTÖ	90
6.1 Modaalisanoja yhdistävät ja erottavat kollokaatit ja semanttiset preferenssit	90
6.2 Tutkimuksen arviointia	94
6.3 Tutkimustulosten hyödyntäminen	97
LÄHTEET	100
LIITTEET	

1 JOHDANTO

Synonymia on kielentutkimuksessa ja myös sen ulkopuolella laajalti tunnettu käsite. Jokaisen on kielitajunsa pohjalta helppo luetella monia synonyymipareja: ISO ja SUURI, AALTO ja LAINE, PUHDISTAA ja PUTSATA, VASTA ja VIHTA, HEVONEN ja KONI ja niin edelleen. Synonyymit määritellään usein samaa merkitseviksi sanoiksi. Kuitenkaan synonymiassa ei ole kyse siitä, että sanat olisivat täysin identtisiä keskenään. Synonyymeja voivat erottaa toisistaan esimerkiksi murre- ja sävyerot: samaa merkitsevät sanat VASTA ja VIHTA esiintyvät eri murteissa, ja synonyymisia sanoja HEVONEN ja KONI puolestaan erottaa erilainen tyylisävy. Edelleen esimerkiksi sanat ISO ja SUURI poikkeavat toisistaan siten, että sanaa ISO käytetään tyypillisesti konkreettisissa yhteyksissä ja sanaa SUURI taas abstrakteissa (Kangasniemi 1997: 42). Suomen kielessä synonymiaa on tutkittu esimerkiksi seuraavien sanaparien tai -ryhmien osalta: HYVIN, KOVIN ja OIKEIN (Jantunen 2004), KESKEINEN ja TÄRKEÄ (Jantunen 2001), AJATELLA, MIETTIÄ, POHTIA ja HARKITA (Arppe 2008), VUOKSI ja TAKIA (Suoraniemi 2011) ja ASTI ja SAAKKA (Päiviö 2007). Tutkimuksissa on löydetty erilaisia synonyymeja toisistaan erottavia piirteitä. Voidaankin siis sanoa, että synonyymit ovat itse asiassa tarkkaan ottaen lähisyntonyymeja – kaikilla synonyymeilla on oletettavasti jonkinlaisia eroja keskenään.

Tässä tutkielmassa tarkoitetaan selvittää, miten synonyymit EHKÄ ja MAHDOLLISESTI eroavat toisistaan. Sanat ovat modaalisanoina ja ilmaisevat kumpikin eräänlaista epävarmuutta. Kielellisen intuition perusteella voidaan sanoa, että EHKÄ ja MAHDOLLISESTI eivät ole täydellisiä synonyymeja, vaan niillä tuntuu olevan ainakin jonkinlainen tyyliero. MAHDOLLISESTI kuulostaa jossain määrin virallisemmalta sanalta, ja EHKÄ voi tuntua hieman puhekielisemmältä. Kovin tarkasti sanojen eroja ei voida kuitenkaan pelkän intuition avulla kuvailla, vaan erot selviävät tutkimalla. Keskityn etsimään sanojen eroja kollokaateista ja semanttisista preferensseistä, jotka ovat kontekstuaalisia eli sanojen lähikontekstissa esiintyviä piirteitä. Sekä kollokaatio että semanttinen preferenssi ovat fraseologisen tutkimuksen ydinkäsitteitä. Kollokaatiolla tarkoitetaan kahden tai useamman sanan esiintymää tekstissä toistensa läheisyydessä (Sinclair 1991a: 170). Tutkin, millaisia sanoja tyypillisesti esiintyy toisaalta EHKÄ-sanoina ja toisaalta MAHDOLLISESTI-sanoina kontekstissa. Semanttinen preferenssi puolestaan on suhde, joka vallitsee sanan sekä merkityksensä puolesta toisiinsa liittyvien sanojen joukon välillä (Stubbs 2001b: 65). Semanttista preferenssiä tutkitaan tarkastelemalla, mitä merkitysryhmiä sanan kollokaatit tyypillisesti edustavat; kyse on siis siitä, mihin merkitysryhmiin kuuluvia sanoja tyypillisesti esiintyy sanojen EHKÄ ja MAHDOLLISESTI kontekstissa.

Tutkin siis sanojen eroja, mutta samalla selvitän, millaiset kollokaatit ja semanttiset preferenssit puolestaan ovat molemmille sanoille tyypillisiä eli millaiset kontekstin leksikaaliset ja semanttiset piirteet yhdistävät synonyymeja toisiinsa. Tutkimus tarjoaa siis tietoa paitsi synonyymien eroista, myös kahden erillisen modaalisanan kontekstuaalisista piirteistä. Tutkimuskysymykset ovat:

- 1) Millaisia kollokaatteja synonyymeilla EHKÄ ja MAHDOLLISESTI on?
- 2) Millaisia semanttisia preferenssejä synonyymeilla EHKÄ ja MAHDOLLISESTI on?
- 3) Mitä kollokaatiot ja semanttiset preferenssit kertovat näiden sanojen synonyymisyydestä?

Ensimmäinen ja toinen tutkimuskysymys koskevat siis sanojen kontekstuaalisia ominaisuuksia, ja kolmantena tutkimuskysymyksenä taas tarkastelen, miten kollokaatiot ja semanttiset preferenssit erottelevat synonyymisia modaalisanoja toisistaan ja toisaalta yhdistävät niitä.

Tutkielma on korpuslingvistinen tutkimus eli perustuu laajan sähköisen kieliaineiston hyödyntämiseen. Korpuslingvistiikan keinoin voidaan tutkia kielen todellista käyttöä ja saadaan tutkittavasta ilmiöstä kvantitatiivista, todelliseen kielenkäyttöön perustuvaa tietoa. Etuna korpusmenetelmässä on myös tutkimusaineistojen laajuus sekä tietokoneohjelmilla toteutettavat automaattiset analyysitoiminnot. (Biber, Conrad & Reppen 2006 [1998]: 4.) Tutkielman aineistona on Suomen kielen tekstikokoelmasta peräisin oleva sanomalehtiaineisto, jota tutkin kvantitatiivisesti juuri kielentutkimuksen tarpeisiin kehitetyn Lemmie-käyttöliittymän avulla.

Tutkielma etenee siten, että ensin luvussa 2 kuvailen tutkielman taustalla olevia kontekstuaalisen semantiikan ja fraseologisen tutkimuksen periaatteita. Esittelen myös synonymiaan liittyvää teoriaa, ja koska tutkimuksen kohteena ovat sanat ovat nimenomaan modaalisanoja, käsittelen myös modaalisuutta ilmiönä. Luvussa 3 kuvailen tarkemmin käyttämäni aineistoa, selostan perusteet tutkittavien lekseemien valinnalle ja esittelen kollokaatioiden ja semanttisen preferenssin tutkimiseen vaadittavat tutkimusmenetelmät. Luvussa 4 analysoin sanojen kollokaatteja ja luvussa 5 semanttista preferenssiä. Viimeiseksi luvussa 6 pohdin tutkimusta kokonaisuutena ja arvioin tutkimuksen antia.

2 TEORIATAUSTA

2.1 Kontekstuaalinen semantiikka

2.1.1 Kontekstuaalisen semantiikan lähtökohdat

Semantiikka on kielitieteen alue, jossa tutkitaan ilmausten merkityksiä. Eri teoreettiset suuntaukset mallintavat merkitystä keskenään hieman eri tavoin, joten erillisiksi suuntauksikseen erottuvat esimerkiksi generatiivinen semantiikka ja kognitiivinen semantiikka. (Kangasniemi 1997: 22.) Tässä tutkimuksessa nojaan pääasiassa kontekstuaalisen semantiikan käsityksiin kielen merkityksistä ja niiden tutkimisesta. Kontekstuaalinen semantiikka tutkii kielen merkityksiä kontekstilähtöisesti, ja merkityksiä lähestytään siinä tutkimalla joko lähintä tekstikontekstia tai laajempaa sosiaalista kontekstia (Jantunen 2004: 7).

Lähestymistavan juuret ovat brittiläisessä kielentutkimuksessa (Stubbs 1996: 22), ja sen varhaisimpana edustajana voidaan pitää J. R. Firthiä (1957, 1968). Firth käsittää kielen laajasti sosiaalisen, inhimillisen toiminnan muotona ja korostaa, että kieli kuuluu aina johonkin kielenkäyttötilanteeseen, josta ilmausten merkitys riippuu. Hän käyttää käsitettä tilannekonteksti, jolla hän viittaa sosiaalisten prosessien toistuviin, tyypillisiin tapahtumiin. Tilannekonteksti on abstrakti käsite, johon kuuluvat osanottajat, heidän verbaalinen ja nonverbaalinen toimintansa, muut tilanteen olennaiset piirteet ja verbaalisen toiminnan vaikutukset. Nämä kaikki tekijät ovat yhteydessä toisiinsa, ja kun ne toistuvat kielessä taajaan tietyllä tavalla, ne muodostavat yhdessä abstraktion, tilannekontekstin. (Firth 1957: 182; 1968: 175–176, 199.) Tilannekontekstin eri piirteet eivät ole tässä tutkimuksessa olennaisia, mutta käsitteessä keskeistä on idea siitä, että tietyt asiat toistuvat kielessä. On siis olemassa tyypillisiä, kertautuvia tapoja käyttää kieltä. Firth (1957: 35; 1968: 170) myös tähdentää, että on tutkittava nimenomaan objektiivisesti havainnoitavissa olevaa kieltä. Kieltä ei tule pitää mielensisäisinä tiloina eikä tutkia introspektion menetelmin, sillä mielensisäiset tapahtumat eivät ole havainnoitavissa. Merkitykset eivät Firthin mukaan liity piilossa oleviin mentaalsiin prosesseihin, vaan etupäässä tilannekontekstiin ja sellaiseen kieleen, josta voidaan tehdä havaintoja. Täten kontekstuaalinen semantiikka merkityksentutkimuksen suuntauksena eroaa etenkin konseptuaalisesta semantiikasta, jossa merkitystä pidetään kielenkäyttäjän tajunnassa olevana käsitteenä, ideana tai mielikuvana (Kangasniemi 1997: 10; Jantunen 2004: 7).

Firthin ajatusten mukaisesti kontekstuaalisen semantiikan keskeinen ajatus kaiken kaikkiaan onkin, että kielen merkityksiä tutkittaessa oleellista on tarkastella todellista, autenttista kielenkäyttöä (Stubbs 1996: 22–23.) Kontekstuaalista semantiikkaa niin ikään edustava Sinclair huomauttaa (1991a: 36–40), että pelkkä introspektiivinen havainnointi ei tarjoa tietoa kielenkäytöstä. Esimerkiksi sanakirjaa tehtäessä ei voida käyttää lähteenä ainoastaan kielenkäyttäjien ajatuksia kielestä, sillä siten saavutettaisiin vain käsityksiä, ei faktoja. Toisaalta Sinclair toteaa, että kielenkäyttäjän intuitiota ja introspektiota kyllä myös tarvitaan tutkimuksessa ja että ne ovatkin siinä joka tapauksessa läsnä jollakin tavalla: esimerkiksi sanakirjan teossa leksikografi väistämättä käyttää omaa arviotaan käsitellessään sanakirjan pohjana olevaa aineistoa (intuition hyödyntämisestä tarkemmin ks. luku 3.3.5). Toisaalta taas kielenkuvaus, jonka tueksi ei esitetä todisteita kielestä, ei ole Sinclairin mukaan lainkaan uskottavaa. Merkityksiä ei pidä myöskään tutkia sellaisista tekstikatkelmista, jotka on erotettu kontekstistaan (Stubbs 1996: 23), vaan konteksti on otettava huomioon. Tässä tutkielmassa keskityn nimenomaan sanojen toistuvien, tyypillisten kontekstuaalisten piirteiden tutkimiseen.

Keskeistä kontekstuaalisessa semantiikassa on lisäksi ajatus kieliopin ja leksikon yhteydestä. Muoto ja merkitys ovat erottamattomat ja leksikko ja kielioppi toisistaan riippuvaiset (Stubbs 1996: 23). Sinclair (1991b: 495) esimerkiksi havaitsee englannin kielen sanasta *EYE* 'silmä', että yksikkömuotoisena se ei yleensä ole korvattavissa monikkomuodolla eikä päinvastoin. Monikkomuotoinen *eyes* esiintyy viittaamassa silmään, ihmiseen anatomiseen osaan, sekä muun muassa sellaisten adjektiivien kuin *CRITICAL* 'kriittinen' ja *DISBELIEVING* 'epäilevä' yhteydessä kuvaannollisissa merkityksissä. Yksikkömuotoa *eye* puolestaan käytetään esimerkiksi sellaisissa kuvaannollisissa fraaseissa kuin *an eye for an eye* 'silmä silmästä' ja *keep an eye on something* 'pitää silmällä'. Toisessa yhteydessä Sinclair (1991a: 57) puolestaan havaitsee, että *YIELD*-sana merkityksessä 'antaa periksi, taipua' esiintyy tyypillisesti intransitiivisena: aineiston 49 tapauksesta se esiintyy 33 kertaa intransitiivisena ja vain 16 kertaa transitiivisena. Sinclair (1991a: 103–105) huomauttaa, että kieliopeissa otetaan vain harvoin huomioon tällaisia ilmiöitä, siis esimerkiksi sitä, että joskus jokin kieliopillinen muoto on tyypillinen vain tietyille sanan merkityksille, ei kaikille. Samoin hän toteaa, että sanakirjoissa puolestaan ei aina kiinnitetä huomiota muotoon, esimerkiksi siihen, että sanan tietty merkitys voi esiintyä tyypillisesti joissakin tietyissä syntaktisissa rakenteissa. Sinclair asettaakin kyseenalaiseksi perinteisten kielioppien ja perinteisten sanakirjojen esittämän kielen kuvauksen, jossa kielioppi ja leksikko eivät vaikuta toisiinsa. Hänen mukaansa ne todellisuudessa ovat yhteydessä toisiinsa, ja kieltä kuvattaessa onkin etsittävä säännöllisiä ja tyypillisiä yhteyksiä niiden välillä. Vaikka sanan yksi keskeinen piirre siis kiistattomasti on, millaisia kieliopillisia kytkentöjä se saa kontekstissaan, tässä tutkielmassa ei kuitenkaan ole tutkielman rajallisen laajuuden vuoksi mahdolli-

suutta perehtyä systemaattisesti sanojen EHKÄ ja MAHDOLLISESTI kieliopillisiin kontekstuaalisiin piirteisiin. Analyysi keskittyy sitä vastoin kollokaatioihin ja semanttiseen preferenssiin.

2.1.2 Kontekstuaalinen valinta

Kontekstuaalisessa semantiikassa tärkeitä käsitteitä ovat kontekstuaalinen valinta sekä paradigmaattinen ja syntagmaattinen suhde. Kun tutkitaan ilmausten kontekstuaalista valintaa, tarkastellaan sitä, millä perusteella ilmaus, esimerkiksi jokin tietty sana, valitaan paradigmaattisesta suhteesta syntagmaattiseen suhteeseen. Näistä paradigmaattinen suhde kuvaa sitä, että jokin tietty ilmaus valitaan käyttöön muiden mahdollisten ilmausten joukosta – vaihtoehtoisia ilmauksia ovat muun muassa kyseisen ilmauksen synonyymit ja hyponyymit. Kun siis jossakin lauseessa puhutaan vaikkapa KOIRASTA, sana on valittu käyttöön muun muassa sellaisten ilmausten joukosta kuin HAUVA, KISSA ja ELÄIN. Sana KOIRA on siis paradigmaattisessa suhteessa sanoihin HAUVA, KISSA ja ELÄIN – kuten myös lukuisiin muihin kielen sanoihin. Syntagmaattisella suhteella puolestaan tarkoitetaan sitä, että ilmaus valitaan muiden ilmausten kanssa lineaariseen jatkumoon, kuten johonkin tiettyyn lauseeseen. Esimerkiksi lauseessa *Koira haukkuu* sanat KOIRA ja HAUKKUA ovat toistensa kanssa syntagmaattisessa suhteessa. Kontekstuaalisessa semantiikassa kiinnostus kohdistuu siis niihin sääntöihin ja preferensseihin, joiden mukaisesti ilmaus valitaan paradigmaattisesta suhteesta syntagmaattiseen suhteeseen. On siis pyrittävä selvittämään, millä tavalla sanat yhdistyvät toistensa kanssa kielenkäytössä ja miten sanat valikoituvat kulloiseenkin kontekstiinsa. (Firth 1957: 196; Cruse 1991: 100–101; Jantunen 2004: 8–9; Karlsson 2006: 219; esimerkit omia.)

Sinclair (1991a: 109–114) esittää kaksi kontekstuaalisessa semantiikassa sittemmin tunnetuksi tullutta periaatetta, jotka ohjaavat ilmaisujen kontekstuaalista valintaa: vapaan valinnan periaatteen (*open choice principle*) ja idiomaattisen valinnan periaatteen (*idiom principle*). Näistä ensin mainittu tarkoittaa sitä, että ilmaisuun, esimerkiksi lauseeseen, voidaan valita sanoja aivan vapaasti kielen kaikista ilmaisuvaroista. Esimerkiksi jokainen teksti on näin ollen tulosta lukuisista monimutkaisista valinnoista. Ainoa tekijä, joka rajoittaa ilmaisujen valintaa, on kieliopillisuus: muodostettavien lauseiden on oltava kieliopin mukaisia. Vapaan valinnan periaate ei kuitenkaan riitä selittämään ilmaisujen valintaa ja tekstien rakentumista, sillä sanat eivät esiinny teksteissä satunnaisesti. Jossain määrin ensinnäkin ympärillä oleva maailma heijastuu kieleen ja vaikuttaa siihen, etteivät valinnat ole vain satunnaisia: esimerkiksi fyysisessä maailmassa yhdessä esiintyvät asiat todennäköisesti mainitaan myös kielessä toistensa läheisyydessä. Kielenulkoisella maailmallakaan ei kuitenkaan voida selittää kaikkea kielenkäyttöä, sillä on olemassa silti lukuisia vaihtoehtoisia tapoja ilmaista

asioita. Idiomaattisen valinnan periaate selittää sen, mitä vapaan valinnan periaate tai ulkoisen maailman ominaisuudet eivät selitä. Idiomaattisen valinnan periaatteella Sinclair tarkoittaa, että kielenkäyttäjällä on käytettävissään suuri määrä osittain valmiita fraaseja, joista kukin muodostaa yksittäisen valintayksikön. Esimerkiksi englannin kielen ilmaus *of course* 'tietysti' koostuu kahdesta sanasta, mutta ilmaus kuitenkin toimii yhtenä kokonaisuutena yksikkönä ja valitaan kielenkäyttöön nimenomaan yhtenä kokonaisuutena, ei yksittäisinä lekseimeinä OF ja COURSE. Sinclair katsoo, että yleensä tekstit rakentuvat pääasiassa idiomiperiaatteen mukaan.

Idiomiperiaatteen mukaan sanat ja muut yksiköt valikoituvat kielenkäyttöön siis valmiiden, olemassaolevien rakenteiden pohjalta. Suomen kielessä idiomiperiaate ilmenee muun muassa idiomissa *katsoa läpi sormien*. Yksittäisiä sanoja ei ole valittu ilmaukseen syntagmaattiseen suhteeseen sattumalta, vaan sanat KATSOA, LÄPI ja SORMI muodostavat sellaisen syntagmaattisen suhteen, joka on vakiintunut kieleen edustamaan merkitystä 'hyväksyä kielletty toiminta' (idiomin merkitys Muikku-Werner, Jantunen & Kokko 2008 s.v. *katsoa/katsella läpi sormien*). Sanat siis valitaan kielenkäyttöön yhtenä yksikkönä. On kuitenkin syytä huomata, että idiomiperiaatteen mukaan kielenkäyttöön eivät valikoidu ainoastaan jähmeät, vakiintuneet idiomit, vaan myös esimerkiksi kollokaatiot – ja kaiken kaikkiaan ylipäätään fraseologiset yksiköt, jota käsitettä tarkastelen tarkemmin luvussa 2.2.1. Esimerkiksi sanat EHKÄ ja MAHDOLLISESTI eivät oletettavasti muodosta idiomeja, vaan pikemmin löyhempiä kollokaatiosuhteita toisten sanojen kanssa, mutta kyse on silti niidenkin osalta nimenomaan Sinclairin idiomiperiaatteesta: valmiit rakenteet ja kielen tyypilliset käyttötavat vaikuttavat siihen, miten modaalisanoja valitaan toisten sanojen kanssa syntagmaattisiin suhteisiin.

Jantunen (2004: 15) huomauttaa, että sanojen yhdistymistä toisten sanojen kanssa tulee pitää jatkumona. Kontekstuaalisen valinnan toisessa ääripäässä on sanojen vapaa valinta ja toisessa päässä puolestaan valintarajoitukset, jotka rajaavat lekseimin kontekstuaaliset ominaisuudet tiukasti. Näiden ääripäiden väliin sijoittuvat valintapreferenssit, joissa kyse ei ole siitä, mitkä kontekstuaaliset piirteet ovat mahdollisia, vaan siitä, mitkä ovat tyypillisiä. Valintarajoitukset siis määrittävät, että vain tietyt kollokaatiot ja muut kontekstuaaliset ominaisuudet ovat mahdollisia, kun taas valintapreferenssit kuvaavat väljemmin sitä, millaiset kontekstuaaliset piirteet ovat sanalle tyypillisiä.

Kaiken kaikkiaan sana ei siis ole kontekstistaan irrallinen, vaan sana ja konteksti ovat keskinäisessä vuorovaikutuksessa: konteksti ohjaa sitä, mikä sana valitaan käyttöön, ja toisaalta valittu sana toimii kontekstina toiselle yksikölle (Sinclair 1998: 8; Jantunen 2004: 29–30). Jantunen (mts. 30–32) havainnollistaa tätä kontekstuaalisen semantiikan periaatetta kokoamalla yhteen leksikaalisen yksikön kontekstuaalisen profiilin osatekijät eli ne tekijät, jotka muodostavat kontekstuaalisen merkitysyksikön. Näitä tekijöitä ovat ensinnäkin tilannekonteksti ja toisekseen koteksti, jossa keskiön muodostavat lemma ja kulloinenkin sananmuoto. Kotekstiin kuuluvia tekijöitä ovat Jantunen

mukaan myös kollokaatit, kolligaatit, semanttinen preferenssi ja semanttinen prosodia ja eräältä kannalta katsottuna myös morfologinen preferenssi. Käsittelen näitä tekijöitä tarkemmin luvussa 2.3 synonymian yhteydessä, sillä nämä tekijät usein myös erottelevat synonyymisia sanoja toisistaan.

2.2 Fraseologia

2.2.1 Fraseologinen yksikkö

Kontekstuaalisen semantiikan periaatteet todellisen kielenkäytön ja kontekstuaalisten suhteiden tutkimisen tärkeydestä toteutuvat fraseologisessa kielentutkimuksessa. Fraseologiassa tutkitaan kielen vakiintuneita rakenteita, kuten idiomeja (*mennä kuuroille korville*), rutiineja (*Hyvää päivää, Ei kestä*) ja kollokaatioita (Granger & Paquot 2008: 27–28, 36; Jantunen 2009b: 360–361). Fraseologia on vasta viime aikoina vakiinnuttanut asemansa omana kielentutkimuksen alanaan; kiinnostus fraseologiaan on kasvanut huomattavasti vasta viimeisen noin kahdenkymmenen vuoden aikana (Granger & Meunier 2008: XIX; Gries 2008: 3). Fraseologiaan kohdistuvaa kiinnostusta on lisännyt erityisesti korpusten kehittyminen. Nykyään saatavilla olevat laajat, sähköiset korpukset ovat tärkeä lähde fraseologiseen analyysiin, ja ne mahdollistavat kvantitatiivisen lähestymistavan. (Sinclair 2008: XVI; Hoffmann, Fischer-Starcke & Sand 2013: 1.) Tässäkin tutkielmassa fraseologiaa lähestytään siis kvantitatiivisin menetelmin.

Eri tutkijat ovat määritelleet fraseologian eri tavoin, ja jotkut fraseologista tutkimusta tekevät eivät ylipäätään lainkaan anna käsitteelle selkeää määritelmää. Terminologia alalla on sekavaa, ja yhdelle ja samalle ilmiölle voi olla olemassa monia eri termejä. (Gries 2008: 4; Hoffmann ym. 2013: 1.) Edes fraseologista yksikköä, jota voidaan pitää fraseologian varsinaisena tutkimuskohteena, ei määritellä aina selkeästi. Gries (mts. 4–10) kuitenkin tarttuu tähän ongelmaan ja esittää eksplisiittisesti kuusi parametria, joiden avulla fraseologinen yksikkö voidaan määritellä. Gries tarjoaa oman määritelmänsä (ks. mts. 4–6), mutta toteaa, ettei kaikkien tutkijoiden tarvitse noudattaa hänen antamaansa fraseologisen yksikön määritelmää. Tärkeää sen sijaan on ottaa määrittelyssä huomioon aina kuusi ulottuvuutta. Seuraavaksi selostan, mitkä Griesin kuusi parametria ovat ja miten otan ne huomioon tässä tutkimuksessa. Samoja fraseologisen yksikön piirteitä käsittelee myös Jantunen (2004), jonka näkemyksiä niin ikään esittelen.

Ensinnäkin fraseologisessa tutkimuksessa on Griesin mukaan määriteltävä, millaisista elementeistä fraseologinen yksikkö voi koostua. Gries pitää fraseologista yksikköä sanan (sananmuodon tai lemman) sekä jonkin muun kielellisen elementin myötäesiintymänä, ja tämä toinen kielelli-

nen elementti voi olla esimerkiksi toinen sana tai kieliopillinen rakenne. Kahden sanan myötäesiintymästä esimerkki on vaikkapa englannin kielen ilmaus *strong tea*, jossa kaksi lekseemiä esiintyy toistensa seurassa. Sanan ja kieliopillisen rakenteen yhteisesiintymä taas toteutuu esimerkiksi englannin verbissä HEM: sanaa käytetään pääasiassa passiivimuodossa. (Gries 2008: 5–6.) Ensin mainitussa tapauksessa kyse on kollokaatiosta ja jälkimmäisessä tapauksessa morfologisesta preferenssistä. Täsmällisempänä esityksenä fraseologisen yksikön elementeistä voidaan kuitenkin pitää Jantusen (2004: 31) kokoamaa kuvausta leksikaalisen yksikön kontekstuaalisesta profiilista. Jos Griesin ajatusta hieman laajennetaan Jantusen esityksen mukaiseksi, voi ”toinen kielellinen elementti” olla myös sanan kontekstissa esiintyvä kieliopillinen kategoria, jolloin kyseessä on kolligatio (kolligatation käsitteestä ks. myös esim. Firth 1968: 181; Sinclair 2004: 32). Jantunen (mp.) mainitsee lisäksi semanttisen preferenssin ja semanttisen prosodian. Sanan myötäesiintymää tiettyjen kontekstuaalisten semanttisten piirteiden kanssa voidaan siis niin ikään pitää fraseologisena yksikkönä – jolloin on kyse semanttisesta preferenssistä tai semanttisesta prosodiasta –, joskin tällöin on kysymys hieman laajemman ja abstraktimman tason yksiköstä. Vaikka kaikkia mainittuja piirteitä voidaan pitää olennaisina fraseologisen yksikön piirteinä, tässä työssä keskityn siis kollokaatioihin ja semanttisen preferenssiin.

Toisekseen tutkimuksessa on päätettävä, koostuuko fraseologinen yksikkö kahdesta vai useammasta elementistä (Gries 2008: 4–5). Sanoille voi olla tyypillistä, että ne muodostavat kolme- tai useampisanaisia leksikaalisia yksiköitä. Kenny (2001) ja Jantunen (2004) käyttävät tällaisista leksikaalisista yksiköistä nimitystä klusterit. Esimerkiksi Jantunen (mts. 103) havaitsee sanan KOVIN esiintyvän tyypillisesti klusterissa *ei olla kovin*. Tässä tutkielmassa tutkimuskysymykseni ei kohdistu siihen, millaisissa klustereissa sanat EHKÄ ja MAHDOLLISESTI esiintyvät, vaan keskityn tarkastelemaan, millaisissa kollokaatioissa eli kahden sanan myötäesiintymissä ne esiintyvät. Jos aineistosta kuitenkin selkeästi nousee tapauksia, joissa tutkittavat sanat muodostavat laajempia kuin kahden sanan kokonaisuuksia, nostan ne mainintoina esille.

Edelleen kolmantena huomioon otettavana tekijänä Gries (2008: 4–5) mainitsee sen, kuinka monta kertaa ilmaisun on esiinnyttävä, jotta se laskettaisiin fraseologiseksi yksiköksi. Kuten Jantunen (2004: 16) toteaa, kollokaatiolla viitataan yleensä sellaisiin sanojen yhteisesiintymiin, jotka ovat toistuvia. Esimerkiksi jo Firth (1968: 180) puhuu nimenomaan *tyypillisistä* kollokaatioista (*habitual collocation*). Tutkielmassani keskityn sellaisiin kollokaatioihin, jotka esiintyvät aineistossa vähintään viisi kertaa. Luku viisi on kuitenkin vain absoluuttinen vähimmäisluku; lisäksi arvioin kollokaatioiden tilastollista merkitsevyyttä tilastollisten testien avulla. Semanttisen preferenssin osalta en puolestaan ole määritellyt, kuinka monta kertaa tietyn semanttisen luokan on aineistossa esiinnyttävä, jotta laskisin myötäesiintymän fraseologiseksi yksiköksi. Semanttista preferenssiä kä-

sitellään muissakin tutkimuksissa puhumalla vain esimerkiksi ”vahvoista” ja ”voimakkaista” preferensseistä (ks. esim. Jantunen 2004: 138–140), eikä yleensä varsinaisesti aseteta tiettyä rajaa sille, mikä on varsinainen preferenssi.

Neljäntenä tekijänä on kiinnitettävä huomiota siihen, kuinka kaukana toisistaan fraseologisen yksikön elementit voivat esiintyä. Joissakin tutkimuksissa tarkastellaan fraseologisina yksiköinä ainoastaan keskenään vierekkäisiä elementtejä, mutta myös katkonaisia ilmaisuja voidaan pitää fraseologisina yksiköinä. (Gries 2008: 4–5.) Esimerkiksi Liu, Wang, Wu ja Li (2009: 487) pitävät kollokaatioina sekä yhtenäisenä esiintyvää ilmausta *by accident* että myös ilmaisua *take...advice*, jonka sanojen välissä voi esiintyä muita sanoja. Jantusen (2004: 19) mukaan kollokaatioissa elementtien välillä voi olla useitakin sanoja, kun taas klustereissa sanat ovat tyypillisesti peräkkäin. Tässä tutkimuksessa kriteerinä ei ole, että fraseologisen yksikön elementtien – siis modaalisanan ja sen kollokaatin – olisi esiinnyttävä nimenomaan vierekkäin, vaan niiden välissä voi hyvin olla muita sanoja. Mahdollisiksi kollokaateiksi lasken kuitenkin vain tietyn välimatkan päässä tutkittavasta sanasta olevat sanat (ks. luku 3.3.2).

Viidentenä tekijänä tulee määritellä, missä määrin fraseologisen yksikön elementeille sallitaan leksikaalista ja syntaktista joustavuutta. Esimerkiksi englannin kielen ilmaus *by and large* ’yleisesti ottaen’ on joustamaton, kun taas ilmausta *kick the bucket* ’potkaista tyhjää’ voidaan taivuttaa eri aikamuodoissa. On myös leksikaalisesti täysin vapaita ja joustavia rakenteita: esimerkiksi englannin ditransitiivisessa rakenteessa ehtona on ainoastaan se, että siihen kuuluu subjekti, verbi ja kaksi objektia. Siihen voidaan sijoittaa monia eri lekseemejä, eli leksikaalisen joustavuuden määrä edellä mainittuihin tapauksiin verrattuna on suuri. (Gries 2008: 5–6.) Tässä tutkielmassa keskityn tutkimaan lekseemejä EHKÄ ja MAHDOLLISESTI, joten vaatimuksena fraseologiselle yksikölle on, että se sisältää jommankumman näistä lekseemeistä. Kieliopillisesti sen sijaan kyse on hyvin joustavasta fraseologisesta yksiköstä: Otan ensinnäkin mukaan analyysiin lekseemien kaikki sananmuodot (vaihtelua ei tosin ole paljoa; ainoastaan EHKÄ-lekseemi esiintyy toisinaan muodossa *ehkäpä*). Lisäksi en rajaa fraseologista yksikköä tarkoittamaan vain sellaisia sanayhdistelmiä, joissa kollokaatti esiintyy vain tietyssä taivutusmuodossa, vaan tutkittavien lekseemien kollokaatit voivat esiintyä missä taivutusmuodoissa tahansa.

Kuudes kysymys koskee fraseologisen yksikön semanttista yhtenäisyyttä (*semantic unity*) ja semanttista epäkompositionaalisuutta (*semantic non-compositionality / semantic non-predictability*). Semanttinen yhtenäisyys on monille tutkijoille tärkein fraseologisen yksikön piirre: fraseologisen yksikön elementit toimivat yhdessä yhtenä semanttisena yksikkönä, eli yksittäisistä sanoista muodostuvalla kokonaisuudella on oma merkityksensä samalla tavoin kuin yksittäisellä morfeemilla tai sanalla. Tutkimukset kuitenkin poikkeavat toisistaan siltä osin, edellytetäänkö niissä

fraseologiselle yksikölle yhdeksi piirteeksi semanttista epäkompositionaalisuutta eli sitä, että elementtien summa saa eri tulkinnan kuin elementit, jotka sen muodostavat. (Fraser 1976: v; Gries 2008: 4–6.) Idiomi *mennä penkin alle* on esimerkki sekä semanttisesta yhtenäisyydestä että semanttisesta epäkompositionaalisuudesta: sen elementit – lekseemit MENNÄ, PENKKI ja ALLE – muodostavat keskenään yhden kokonaisuuden, semanttisen yksikön, jonka merkitys on ’sujua huonosti, epäonnistua’ (merkitys Muikku-Werner ym. 2008 s.v. *mennä penkin alle*). Idiomi on myös semanttisesti epäkompositionaalinen, sillä sen merkitys ei ole summa niiden lekseemien merkityksistä, joista idiomi muodostuu: ilmauksella ei tarkoiteta sitä, että joku tai jokin kirjaimellisesti menisi penkin alapuolelle, vaan ilmaus on vakiintunut merkitsemään juuri ’huonosti sujumista, epäonnistumista’.

Perinteisessä fraseologisessa tutkimuksessa kiinnostus on kohdistunut juuri edellä kuvatun kaltaisiin leksikaalisia yksiköihin, siis sellaisiin, joiden merkitys on erilainen kuin sen yksittäisten osien summa (Granger & Paquot 2008: 31). Tässä tutkielmassa kuitenkin noudatan uudemman fraseologisen tutkimuksen näkemyksiä, jossa fraseologian käsite on laajentunut perinteiseen tutkimukseen verrattuna. Kuten Granger ja Paquot (mp.) toteavat, enää nimittäin semanttinen epäkompositionaalisuus ei niinkään ole fraseologisen yksikön kriteerinä, vaan monet tutkijat pitävät sitä toissijaisena piirteenä, joka on joillakin, mutta ei kaikilla fraseologisilla yksiköillä. Tämän tutkielman tutkimuskohteena olevat fraseologiset yksiköt eli sanojen EHKÄ ja MAHDOLLISESTI kollokaatiot ovat kaikki oletettavasti tulkittavissa kirjaimellisesti, eli ne eivät ole semanttisesti epäkompositionaalisia. Itse asiassa tässä tutkielmassa oleellista on myös se, että en edellytä fraseologisesta yksiköltä edes edellä mainittua semanttista yhtenäisyyttä. En siis vaadi tutkimiltani rakenteilta sitä, että havaitsemani kollokaatti muodostaisi yhtenäisen semanttisen yksikön modaalisanan kanssa. Vaikka Gries edellä kuvatun mukaisesti mainitseekin semanttisen yhtenäisyyden olevan monille tutkijoille fraseologisen yksikön tärkeä kriteeri, on myös paljon tutkimuksia, joissa se nimenomaan ei ole kriteerinä. Esimerkiksi Jantunen ja Brunni (2012: 75) katsovat fraseologiaan kuuluviksi sellaiset ilmaukset kuin *voi härskiintyy*, *leipä homehtuu* ja *maito happanee*. Kielenkäyttäjät liittää leivän pilaantumiseen juuri verbin HOMEHTUA eikä esimerkiksi verbiä HÄRSKIINTYÄ. Nämä kollokaatiot eivät muodosta semanttisesti yhtenäistä yksikköä, mutta näissä esimerkeissä ja samoin modaalisanojen kollokaatioissa on silti nimenomaan kyse fraseologisista yksiköistä ja Sinclairin idiomiperiaatteen ilmentymistä. Tämäkin tutkielma linkittyy näin ollen osaksi uudempaa fraseologista tutkimusta.

2.2.2 Kollokaatio

Kollokaation käsite on alun perin peräisin Firthiltä (1957, 1968). Hän on esittänyt (1968: 179) tunnetuksi tulleen lauseen, jota sittemmin on siteerattu paljon: ”*You shall know a word by the company it keeps*”. Lause kuvaa sitä, että sanastontutkimuksessa keskeistä on selvittää, millaisessa ympäristössä mikin sana esiintyy. Kollokaatioita ovat myöhemmin käsitelleet tutkimuksissaan monet tutkijat, muun muassa Sinclair, jonka esittämän määritelmän mukaan (1991a: 170) kollokaatio on kahden tai useamman sanan esiintymä tekstissä toistensa läheisyydessä. Sinclairin määritelmä edustaa kollokaation tekstuaalista määrittelyä, mutta lisäksi kollokaatiota on pidetty tilastollisena tai assosiativisena ilmiönä (Jantunen 2001: 173). Tilastollisessa mielessä kollokaatiosta puhuvat esimerkiksi Biber ja Conrad (1999: 183). Tilastollisen määritelmän mukaan kollokaatio on noodin eli tutkittavan sanan (engl. *node*) ja kollokaatin yhdistelmä, jota esiintyy frekventimmin kuin noodin sekä kollokaatin omien frekvenssien perusteella on odotuksenmukaista. Keskeistä tilastollisessa määritelmässä on siis se, että kollokaatio toistuu tutkittavassa aineistossa riittävän usein. (CSCc; Jantunen 2001: 173.) Kun kollokaatio määritellään puolestaan assosiativiseksi ilmiöksi, korostetaan sitä, että kielen sanoilla on merkitykseen pohjautuva taipumus esiintyä yhdessä tietynlaisten toisten sanojen kanssa. Esimerkiksi sanat HEVONEN ja HIRNUA liittyvät merkityksiltään toisiinsa ja kollokoivat siten kielessä keskenään. Merkitykseltään yhteen liittyvillä sanoilla on siis taipumus esiintyä kielessä yhdessä. (Leech 1990: 17; Jantunen 2001: 173; Karlsson 2006: 232.)

Kollokaatio assosiativisena ilmiönä vaatii kuitenkin hieman tarkempaa tarkastelua. Sanat HEVONEN ja HIRNUA assosioituvat toisiinsa kielessä siksi, että hevoselle on tyypillistä hirnuminen eli että kielenulkoisessa maailmassa sanojen taustalla olevat tarkoitteet liittyvät toisiinsa. Koska hevoset hirnuvat, on luonnollista, että myös kielessä sanat HEVONEN ja HIRNUA esiintyvät usein toistensa läheisyydessä. Karlssonin mukaan (2006: 233) kollokaatioiden taustalla onkin tarkoitteiden yhteen kuuluminen. Voidaan kuitenkin huomata, etteivät kaikki kollokaatiot perustu siihen, että niiden taustalla olevat tarkoitteet liittyisivät toisiinsa. Karlssonin väite ei nimittäin voi pitää paikkaansa ainakaan sellaisten sanojen osalta, joilla ei ole lainkaan kielenulkoista tarkoitetta. Tällaisia sanoja ovat kieliopilliset sanat, esimerkiksi sanat ETTÄ ja MUTTA, kuten myös sanat EHKÄ ja MAHDOLLISESTI. Ei ole olemassa ilmiötä nimeltä ”ehkä” (sitä vastoin on kyllä esimerkiksi ilmiö nimeltä ”epävarmuus”, jota sana EHKÄ sitten kielen tasolla ilmentää), joten näin ollen ilmiö ”ehkä” ei voi kielenulkoisessa maailmassa myöskään esiintyä minkään toisen tarkoitteen yhteydessä. Jantunenkin (2009b: 358) toteaa, että usein kollokaatioiden taustalla ei lainkaan ole kielenulkoiseen maailmaan liittyvää syytä. Kollokaatio assosiativisena ilmiönä ei siis liity ainoastaan siihen, että sanojen tar-

koitteet esiintyisivät kielenulkoisessa maailmassa toistensa seurassa, vaan kyse on laajemmasta ilmiöstä.

Voidaan ajatella, että kollokaatio assosiatiivisena ilmiönä on lähellä Hoeyn (2007) käyttämää priming-käsitettä. Sillä Hoey tarkoittaa, että joka kerta, kun kielenkäyttäjälle tulee vastaan esimerkiksi jokin kielen sana, tämä tiedostamattomasti painaa mieleensä sanan kontekstin. Kun sama sana tulee uudelleen vastaan, kielenkäyttäjän tajunnassa oleva lista sanan kollokaateista täydentyy. Tämän seurauksena kielenkäyttäjä jotakin tiettyä sanaa käyttäessään tyypillisesti toistaa niitä konteksteja, joissa sanaa on aiemminkin käytetty. Aikaisemmat kielenkäyttötilanteet siis virittävät kielenkäyttäjän niin, että tämä käyttää sanaa usein sen tyypillisten kollokaatioiden kanssa. Sanaa käytetään niin ikään sille tyypillisessä kieliopillisessa tehtävässä, samassa semanttisessa kontekstissa ja niin edelleen; Hoey tarkoittaa primingin käsitteellä laajasti muidenkin eri kielen piirteiden kuin kollokaatioiden toistumista kielenkäytössä. Keskeistä on joka tapauksessa se, että kielen piirteet toistuvat sellaisina kuin niitä on aiemminkin kielessä käytetty. Hoey pitää primingia ihmisen, ei sanan ominaisuutena: ei tule niinkään sanoa, että sana esiintyy tietyn kollokaatin yhteydessä, vaan että useimmat kielenpuhujat assosioivat sanan tiettyyn kollokaattiin. (Hoey 2007: 7–8.) Niinpä voidaan ajatella, että kollokaation assosiatiivinen määrittely liittyy oikeastaan kielenpuhujan tajuntaan, kognitioon: sanojen assosioituminen toisiinsa tapahtuu alun perin kielenkäyttäjän kognitiossa.

Kaiken kaikkiaan kollokaatiota voidaan siis pitää sekä tekstuaalisena, tilastollisena että assosiatiivisena ilmiönä. Tässä tutkimuksessa tarkoitan kollokaatiolla sanojen esiintymistä teksteissä toistensa läheisyydessä eli nojautun kollokaation tekstuaaliseen määrittelyyn, mutta toisaalta erityisesti kollokaatioanalyysissä korostuu myös kollokaation tilastollinen määritelmä, sillä muodostan kollokaatioanalyysin aineiston tilastollisia menetelmiä apuna käyttäen. En kuitenkaan pidä kollokaation assosiatiivista, tekstuaalista ja tilastollista määritelmää keskenään ristiriitaisina tai vastakkaisina; myöskään Jantunen ei painota niiden eroavaisuuksia tai vastakkaisuutta, vaikka siis luettelee kollokaatiolle kolme erilaista määrittelytapaa. Eri määrittelytavat ovat sen sijaan saman ilmiön eri puolia, joita voidaan kuvata eräänlaisena ketjuna. Siitä, että kielenkäyttäjä assosioi tiettyjä sanoja toisiinsa ja käyttää niitä toistensa yhteydessä (assosiatiivinen määrittely), seuraa, että sanat esiintyvät teksteissä toistensa läheisyydessä (tekstuaalinen määrittely). Sanojen esiintyessä teksteissä lähekkäin puolestaan syntyy tilastollisesti merkitseviä kollokaatioita (tilastollinen määrittely). Tutkittaessa kollokaatioita tilastollisesti tullaan siis samalla tutkineeksi kollokaatioita myös tekstuaalisena ja assosiatiivisena tai kognitiivisena ilmiönä: tilastollisesti merkitsevät kollokaatit kertovat sanoista paitsi tilastollista tietoa, myös samalla tietoa niiden esiintymisestä teksteissä sekä assosioitumisesta kielenkäyttäjän tajunnassa. Näin ollen voidaan myös ajatella, ettei tämä tutkimus eivätkä muutkaan kontekstuaalista semantiikkaa edustavat tutkimukset lopulta ole kovin erillisiä mielen-

sisäisiä käsitteitä tarkastelevista tutkimuksista, vaikka esimerkiksi Firth (ks. luku 2.1.1) vastustaakin kielen pitämistä mielensisäisinä tiloina. Palmerkin (1981: 27) on samalla kannalla kuin Firth: hänen mukaansa merkitystä ei ole syytä pitää kielenkäyttäjän tajunnassa olevana käsitteenä, sillä ainoastaan yksilö itse voisi päästä käsiksi tällaisiin mentaalisiin käsitteisiin ja saatavilla olisi täten vain täysin subjektiivisia näkökulmia kieleen. Nähdäkseni kuitenkin siis kontekstuaalinen semantiikka ja kollokaatioiden tutkiminen kertovat samalla myös siitä, millä tavalla kollokaatiot jäsentyvät kielenkäyttäjän tajunnassa. Ei ole siis tarpeen asettaa vahvasti vastakkain kontekstuaalista semantiikkaa ja niitä semantiikan suuntauksia – esimerkiksi konseptuaalista semantiikkaa –, joissa kieltä tarkastellaan mentaalisisinä tiloina, vaan kollokaatiota voidaan pitää ilmiönä, joka kertoo niin itse kielenkäytöstä kuin myös siitä, millä tavalla sanat yhdistyvät toisiinsa kielenkäyttäjän tajunnassa.

2.2.3 Semanttinen preferenssi

Semanttisen preferenssin käsitteen on ottanut esiin ensimmäisenä Sinclair (1998, 2004). Käsitettä on määritellyt myös muun muassa Stubbs (2001b: 65), jonka mukaan semanttinen preferenssi voidaan määritellä suhteeksi, joka vallitsee lemman tai sananmuodon sekä semanttisesti toisiinsa liittyvien sanojen joukon välillä. Jantunen (2009b: 359) tiivistää, että kyseessä on ”lekseemin ja kotekstissa esiintyvän merkityspiirteen syntagmaattinen suhde”. Kun tutkitaan sanan semanttista preferenssiä, kysymys on siis siitä, mihin merkitysryhmiin kuuluvia sanoja tyypillisesti esiintyy tutkittavan sanan kotekstissa eli mitä semanttisia piirteitä sana suosii kotekstissaan. Kysymys on nimenomaan sellaisista merkityspiirteistä, jotka esiintyvät noodin kotekstissa toistuvasti (Sinclair 1998: 16; Stubbs 2001a: 449). Esimerkiksi Stubbs havaitsee, että englannin kielen sananmuoto *large* ’iso, suuri’ kollokoi tyypillisesti sellaisten sanojen kanssa, jotka ilmentävät ’määrää’ ja ’kokoa’, kuten *number(s)*, ’numero(t), määrä’, *scale* ’asteikko’, ja *part* ’osa’. Lähes 25 prosenttia *large*-muodon esiintymistä 200 miljoonan sanan korpuksessa kollokoi ’määrää’ tai ’kokoa’ ilmaisevien sanojen kanssa. (Stubbs 2001b: 65.) Partington (2004: 148) taas huomaa, että ’jonkin puuttuminen’ on yleinen piirre sanojen UTTERLY, TOTALLY, COMPLETELY ja ENTIRELY (kaikkien merkitys suunnilleen ’täysin, kokonaan’) kollokaateille. Tämä preferenssi voi ilmetä esimerkiksi sellaisista kollokaateista kuin UNKNOWN ’tuntematon’, MEANINGLESS ’merkityksetön’, USELESS ’käyttökelvoton, tarpeeton’ ja WITHOUT ’ilman’.

Semanttinen preferenssi on nykyisin tärkeä käsite korpuslingvistiikassa (Bednarek 2008: 119). Tutkimuksia on tehty etenkin englannin kielestä, mutta suomen kieltä koskeva semanttisen preferenssin tutkimus on puolestaan toistaiseksi hyvin vähäistä. Ilmiötä on kuitenkin tutkinut Jantu-

nen adverbeista HYVIN, KOVIN ja OIKEIN (2004) sekä adverbista AIVAN (2009b). Hän havaitsee esimerkiksi, että OIKEIN-sanalla on selvä preferenssi 'arvoa' merkitseviin sanoihin (esim. HYVÄ, *hyvin*) ja että AIVAN-sana kollokoi tyypillisimmin 'vertailua' ilmaisevien sanojen kanssa (Jantunen 2004: 139; 2009b: 366). Semanttista preferenssiä tutkitaan merkitykseen perustuen välittämättä siitä, mihin sanaluokkaan sanat kuuluvat (Sinclair 1998: 16). Esimerkiksi sanan AIVAN semanttinen preferenssi 'selvyys' ilmenee Jantusen (2009b: 366) mukaan adjektiivista SELVÄ, adverbista *selvästi* ja verbistä KIRKASTUA.

Semanttiseen preferenssiin liittyy läheisesti semanttisen prosodian käsite. Ilmiötä on esitellyt Sinclair (1991a: 112) ottaessaan esiin, että esimerkiksi verbi HAPPEN liittyy kielenkäytössä epämiellyttäviin asioihin. Käsitteenä semanttisen prosodian on ottanut esiin ensimmäisenä Louw (1993: 157). Semanttinen prosodia ilmaisee ikään kuin ilmauksen funktiota ja ohjaa sitä, miten ilmaus tulee tulkita. Siinä on kyse puhujan asenteista ja suhtautumisesta ja lopulta syystä, miksi tiettyä ilmausta on käytetty. Usein se määritellään ilmauksen positiiviseksi tai negatiiviseksi esiintymisympäristöksi: aihetta käsittelevissä tutkimuksissa tuloksina esitetään sanojen taipumusta 'positiiviseen' tai 'negatiiviseen' prosodiaan tai esimerkiksi sille läheiseen 'epämiellyttävyyden' tai 'miellyttävyyden' prosodiaan. (Sinclair 1998: 20; 2004: 34; Partington 2004: 150; Jantunen 2004: 25–26; 2009b: 359.) Esimerkiksi Stubbs (1995b: 247–250) havaitsee, että englannin CAUSE-verbillä ('aiheuttaa') on negatiivinen semanttinen prosodia, koska sillä on kotekstissaan muun muassa sellaisia frekventisti esiintyviä sanoja kuin PROBLEM 'ongelma', DAMAGE 'vahinko, haitta', DEATH 'kuolema' ja DISEASE 'sairaus'. Semanttisen prosodian voidaan sanoa olevan ylemmällä abstraktion tasolla kuin semanttisen preferenssin sekä koskevan ehkä pikemminkin pragmatiikkaa kuin semantiikkaa (Sinclair 2004: 34; Partington 2004: 150; Begagić 2013: 405).

Semanttisen preferenssin ja prosodian ero ei kuitenkaan ole täysin selvä (Stubbs 2001b: 66). Käsitteitä on pidetty usein eri ilmiöinä, mutta joskus niitä on käytetty viittaamaan jopa samaan ilmiöön (Begagić 2013: 404). Partingtonin mukaan (2004: 149) semanttista prosodiaa voidaan pitää semanttisen preferenssin alakategoriana: jos leksikaalisella yksiköllä on preferenssi esiintyä yhdessä sellaisten sanojen kanssa, joita voidaan kuvailla pahoiksi, epäsuosituiksi tai epämiellyttäviksi, tai hyviksi, suosituiksi ja miellyttäviksi, ilmiötä voidaan pitää semanttiseen preferenssin kuuluvana. Käsitteiden välinen rajankäynti vaatii tarkastelua, mutta tässä tutkielmassa en syvenny siihen enempää. En etsi semanttisen preferenssin analyysissa modaalisanonjen kotekstista 'positiivisuutta' tai 'negatiivisuutta' ilmentäviä piirteitä enkä määrittele tällaisia merkityspiirteitä kuuluvaksi semanttiseen preferenssiin, vaan semanttiseen prosodiaan, joka ei kuulu tutkimuskohteisiini.

2.3 Synonymia

Synonymia on yksi kielen paradigmaattisista suhteista esimerkiksi hyponymian ja antonymian ohella (Karlsson 2006: 219). Se voidaan lyhyesti määritellä kielellisten ilmausten merkityksen samuudeksi (Lyons 1996: 60). Synonymia on mahdollista käsittää niin, että se koskee leksikaalista ainesta eli sanoja ja leksikaalisia fraaseja, mutta toisaalta sen voidaan nähdä koskevan myös esimerkiksi kokonaislauseita (Murphy 2003: 144; Stanojević 2009: 193–194). Tässä työssä keskityn kuitenkin lekseemien väliseen synonymiaan, sillä tutkimukseni kohteena on kaksi lekseemiä, EHKÄ ja MAHDOLLISESTI. Synonymiasta ei ole yhtä yhtenäistä käsitteistöä, vaan eri tutkijat ovat käyttäneet hieman eri termejä ilmiöstä puhuessaan. Seuraavaksi tarkastelen, millaisia termejä ja määrittelyjä synonymiaan on liitetty ja mitkä termit mahdollisesti kuvaisivat tämän tutkimuksen kohteina olevien modaalisanojen keskinäistä suhdetta.

Synonymian yksinkertaisimpia määrittelytapoja on käsittää synonymia kognitiivisen merkityksen samuudeksi. Kognitiivisesta merkityksestä käytetään myös nimityksiä konseptuaalinen sekä denotatiivinen merkitys. Kyse on siitä, että sanoilla on samanlainen viittaussuhde eli että ne viittaavat samaan kielenulkoiseen tarkoitteeseen. Näin ollen esimerkiksi sanaparit NAINEN ja AKKA, AALTO ja LAINE, ILTA ja EHTOO sekä RYHTYÄ JA RUVETA voidaan lukea synonyymeiksi. Voidaan kuitenkin huomata, että sanat eivät ole täysin samamerkityksisiä, sillä esimerkiksi ensin mainitussa sanaparissa NAINEN on merkitykseltään neutraali ja AKKA-sanalla on konnotatiivinen, asennetta ilmaiseva sivumerkitys. (Leech 1990: 9; Kangasniemi 1997: 42; Kuiri 2012: 28–31.) Olennaista onkin, että juuri denotaatio on sama: substantiiveilla NAINEN ja AKKA viitataan samaan kielenulkoiseen tarkoitteeseen. Näin ollen voidaan itse asiassa huomata, että modaalisanat EHKÄ ja MAHDOLLISESTI eivät ole määriteltävissä synonyymeiksi samanlaisen viittaussuhteen perusteella: kieliopillisina sanoina ne eivät varsinaisesti viittaa mihinkään kielenulkoiseen tarkoitteeseen, joten siten niiden viittaussuhde ei luonnollisesti voi olla keskenään samanlainen. Kuiri (mts. 30) esittääkin, että synonymia koskee vain sisältösanoina; esimerkiksi VUOKSI ja TAKIA eivät hänen mukaansa kieliopillisina sanoina ole synonyymeja. Kuitenkin niin sanat EHKÄ ja MAHDOLLISESTI kuin myös sanat VUOKSI ja TAKIA ovat selvästi samamerkityksisiä, vaikka niillä ei olekaan denotaatiota; esimerkiksi kumpikin modaalisanoista ilmaisee eräänlaista epävarmuuden merkitystä. Muun muassa Lyons (1968: 427) ja Jantunen (2004: 55–56) toteavatkin, että synonyymisia voivat olla sellaisetkin ilmaukset, joilla ei ole viittaussuhdetta kielen ulkopuolisiin tarkoitteisiin. Voidaan siis tarkastella, miten EHKÄ ja MAHDOLLISESTI suhteutuvat muihin synonymiaan liitettyihin teorioihin.

Yhtenä synonymian tyyppinä voidaan pitää absoluuttista synonymiaa. Absoluuttisia synonyymeja ovat sellaiset leksikaaliset yksiköt, joiden merkitykset ovat täysin samat: sanat ovat yhtä

oikean tuntuista kaikissa konteksteissa. Ne ovat vaihdettavissa keskenään missä tahansa kontekstissa niin, että merkitys ei muutu lainkaan. Absoluuttisen synonymian ehdon täyttävät kuitenkin vain hyvin harvat sanat, jos mitkään. Jos löydetään yksikin konteksti, jossa sanat eivät ole vaihdettavissa keskenään, absoluuttinen synonymia ei toteudu. (Cruse 1991: 265–271; 2000: 157–158.) Esimerkiksi sanojen EHKÄ ja MAHDOLLISESTI osalta absoluuttisen synonymian mahdollisuus voidaan sulkea pois muun muassa jo sillä perusteella, että Kielitoimiston sanakirjassa (KS) on EHKÄ-sanan sisältävä lause, johon sana MAHDOLLISESTI ei ole vaihdettavissa. Voidaan sanoa *Ehkäpä menemme peremmälle*, mutta lause *Mahdollisestipa menemme peremmälle* olisi omituinen tai suorastaan kieli-tajun vastainen eikä toimisi samassa kohteliaan kehotuksen merkityksessä kuin ensimmäinen lause toimii (esimerkkilause: KS s.v. *ehkä*). Kaiken kaikkiaan absoluuttista synonymiaa onkin pidettävä synonymiaa kuvaavan janan päätepisteenä, ei kielissä todella toteutuvana, merkittävänä piirteenä (Cruse 2000: 157). Crusen lisäksi myös monet muut tutkijat suhtautuvat varauksellisesti absoluuttisen synonymian olemassaoloon: esimerkiksi Zgusta (1971: 89) on pitänyt täydellisiä synonyymeja ”melko harvinaisina” (”rather rare”) ja Murphy (2013: 283) ”erittäin harvinaisina” (”exceedingly rare”).

Cruse (1991: 265–291; 2000: 156–160) puhuu absoluuttisen synonymian lisäksi myös propositionaalisesta synonymiasta sekä lähisynonymiasta eli plesionymiasta. Nämä sijoittuvat jatkumolle siten, että absoluuttinen synonymia on kuvitteellisen janan yhdessä päässä ja sitä seuraavat propositionaalinen synonymia sekä sen jälkeen lähisynonymia, joka muuttuu asteittain ei-synonymiaksi. Propositionaalisessa synonymiassa on kyse siitä, että sanat ovat keskenään vaihdettavissa siten, ettei lauseen totuusarvo muutu, kun sana vaihdetaan synonyymiinsa. Jantunen (2004: 54) antaa Crusen propositionaalisesta synonymiasta esimerkeiksi lauseet *Ruoka on valmista* ja *Sapuska on valmista*. Kun ensimmäisen lauseen substantiivi RUOKA vaihdetaan substantiiviin SAPUSKA, lauseen totuusehto ei muutu, vaan lause on edelleen tosi samoilla ehdoilla – sillä ehdolla, että puheena oleva ravinto on valmista. Sen sijaan tämän tutkielman sanojen EHKÄ ja MAHDOLLISESTI synonymian tarkasteluun totuusehdot ja propositio eivät nähdäkseni sovellu, sillä modaalisanoilla ei itse asiassa ole lainkaan propositionaalista sisältöä. Modaalii-ilmauksilla sen sijaan ilmaistaan suhtautumista lauseen propositioon: esimerkiksi väite *Koira on mahdollisesti sairas* koostuu loogisesti kahdesta osasta, modaaliopeaattorista ’mahdollinen’ sekä sen vaikutusalaan kuuluvasta propositiosta ’koira on sairas’ (VISK § 1553). Näin ollen esimerkiksi lauseiden *Koira on mahdollisesti sairas* ja *Koira on ehkä sairas* propositionaalista sisältöä ja totuusehtoja ei oikeastaan voida vertailla, jolloin propositionaalinen synonymia ei sovellu kuvaamaan sanojen EHKÄ ja MAHDOLLISESTI synonymiaa.

Lähisynonyymit eli plesionnyymit eroavat Crusen (1991: 285–287; 2000: 159–160) mukaan propositionaalisesta synonymiasta siinä, että ne tuottavat lauseita, joilla on keskenään eri totuuseh-

dot. Lähisynonyymeille on tyypillistä, että niihin pätee tietynlainen vastakkaisuus: samassa virkkeessä on mahdollista vahvistaa toinen sana ja samalla kieltää toinen. Esimerkki suomen kielen sanojen lähisynonymiasta ja niiden vastakkaisuudesta on muun muassa Kuirin esittämä (2012: 31), sanaparia JOKI ja PURO koskeva esimerkki: virtaavan veden maastosta löytänyt henkilö voi huutaa ”*Hei, täällä on pieni joki*”, ja toinen henkilö voi tavallaan kieltää tämän lauseen vastaamalla ”*Eihän tuo ole joki, pelkkä puro vain.*” Cruse (1991: 287; 2000: 159–160) itse taas mainitsee lähisynonymiasta esimerkkeinä muun muassa sanaparit BIG ’iso’ ja HUGE ’valtava’, PRETTY ’kaunis’ ja HANDSOME ’komea’ sekä MIST ’usva’ ja FOG ’sumu’. Lähisynonyymien keskinäiset erot ovat hänen mukaansa vain vähäisiä tai taustalla olevia. Lähisynonymian eroa ei-synonymiaan nähden ei ole kuitenkaan Crusen mukaan yksinkertaista määritellä, vaan niiden raja jatkumolla on liukuva. Hän myös myöntää, että lähisynonymian periaatteita on hankalaa määritellä aivan tarkasti. Tämän tutkimuksen kohteena olevista sanoista EHKÄ ja MAHDOLLISESTI voidaan todeta, että niiden synonymia muistuttaa lähisynonymiaa. Jos lähisynonymian kriteerinä kuitenkin pidetään edellä kuvattua tietynlaista vastakohtaisuutta, niin modaalisanat EHKÄ ja MAHDOLLISESTI eivät liene luettavissa lähisynonyymeiksi. Esimerkiksi lausetta ”*Huomenna on ehkä sateinen päivä*” kommentoitaessa nimittäin tuskin olisi kielitajun mukaista sanoa ”*Ei, pikemminkin mahdollisesti*”.

Crusen lisäksi myös esimerkiksi Lyonsilla (1996) on oma synonymiaa koskeva teoriansa. Lyons (mts. 44, 63) käyttää käsitettä deskriptiivinen synonymia, joka voidaan määritellä hänen mukaansa deskriptiivisten merkitysten identtisyudeksi. Deskriptiivisellä merkityksellä hän viittaa käsitteeseen, jota on hänen mukaansa nimitetty myös kognitiiviseksi, propositionaaliseksi ja referentiaaliseksi merkitykseksi. Jo edellä todettiin, että tämä käsite ei sovellu kuvaamaan modaalisanojen välistä synonymiaa, sillä kieliopillisina sanoilla niillä ei ole viittaussuhdetta kielenulkoiseen maailmaan. Lyons (mts. 60) puhuu myös lähisynonymiasta: lähisynonyymit ovat hänen mukaansa enemmän tai vähemmän samanlaisia, mutta eivät identtisiä, merkitykseltään. Modaalisanojen voidaan katsoa täyttävän tämän Lyonsin väljän määritelmän. Kyseessä on kuitenkin käsitykseni mukaan sama ilmiö kuin Crusen lähisynonymia: termi on sama kuin Crusella, ja Lyons (mts. 60) myös esimerkiksi mainitsee esimerkkeinä lähisynonymiasta muun muassa sanat MIST ja FOG. Näin ollen modaalisanojen välistä synonymiaa ei varsinaisesti voida kuvata Lyonsin lähisynonymiakäsitteelläkään.

Lyons korostaa teoriassaan lisäksi absoluuttisen synonymian sekä epäabsoluuttisen eli osittaisen synonymian (*non-absolute synonymy, partial synonymy*) välistä eroa. Absoluuttiset synonyymit täyttävät hänen mukaansa kolme ehtoa: 1) kaikki niiden merkitykset ovat identtisiä, 2) ne ovat synonyymisia kaikissa konteksteissa ja 3) ne ovat semanttisesti yhtäpitäviä merkityksen kaikilla ulottuvuuksilla, niin deskriptiivisellä kuin ei-deskriptiivisellä tasolla. Epäabsoluuttisia synonyymeja

puolestaan ovat synonyymit, jotka eivät täytä näitä kaikkia kolmea ehtoa. (Lyons 1996: 60–65.) Vaikka Lyons näin määrittelee absoluuttisen synonymian hiukan eri sanoin kuin Cruse, voidaan todeta, että esimerkiksi yllä mainittu Kielitoimiston sanakirjan esimerkki osoittaa, etteivät modaalisanat kuulu Lyonsinkaan määrittelemään absoluuttiseen synonymiaan: ne eivät ole synonyymisia kaikissa konteksteissa, eli ehto 2 ei täyty. Täten ne Lyonsin teoriassa voidaan luokitella epäabsoluuttisiksi eli osittaisiksi synonyymeiksi. Tämä määritelmä ei kuitenkaan tarjoa kattavaa kuvausta modaalisanojen välisestä synonymiasta, sillä itse asiassa kaikki tai ainakin lähes kaikki synonyymit ovat juuri epäabsoluuttisia synonyymeja – aivan kuten Crusekin siis toteaa teoriassaan, että absoluuttisia synonyymeja kielessä tuskin on. Ei voitane esimerkiksi löytää sanaparia, joka olisi täysin synonyyminen kaikissa konteksteissa eli täyttäisi Lyonsin absoluuttisen synonymian ehdon 2.

Kuten edellä on nähty, synonymiaan liittyvä käsitteistö on kaiken kaikkiaan monenkirjvaa ja mutkikasta. Vanhatalo (2005: 10) esittää, että synonyymien läheisyyden astetta tai laatua voidaan kyllä kuvata, mutta nimenomaan ajantasaisen empiirisen tutkimuksen pohjalta, ja keinotekoisia termirakennelmia tulisi välttää. Nähdäkseni edellä esiteltyjä synonymiateorioita ei ole syytä pitää yksioikoisesti keinotekoisina, pohjautuvathan ne todellisiin esimerkkeihin ja ovat siten perusteltavissa. Voidaan kuitenkin todeta, että teoriat eivät ole täysin kuvaavia, sillä mikään niistä ei nähdäkseni onnistu erityisen kuvaavasti luonnehtimaan sanojen EHKÄ ja MAHDOLLISESTI synonymiaa. Tässä tutkimuksessa lähtökohtani sanojen EHKÄ ja MAHDOLLISESTI synonymialle onkin synonymian jatkumonluonteisuus, johon ajatukseen myös esimerkiksi Vanhatalo (mp.) nojautuu. Hän ehdottaa, että synonyymeista voitaisiin puhua esimerkiksi ”toisiaan merkitykseltään enemmän tai vähemmän muistuttavina ilmauksina”. Samoin Weinreich (1980: 288) ehdottaa, että synonymiaa voidaan kuvata jatkumona, jonka toisessa päässä on joitain harvoja täydellisiä synonyymeja, jos yhtään, ja muut synonyymiparit sijoittuvat eri kohdille synonymian jatkumoa. Samoin Crusen teoriassa itse asiassa korostuu synonymian jatkumonluonteisuus; Crusen (1991: 267) mukaan synonyymit ovat sellaisia leksikaalisia yksiköitä, joiden keskeiset semanttiset piirteet ovat samoja, mutta jotka eroavat perifeerisiltä piirteiltään.

Mitä synonyymien erot, Crusen sanoin perifeeriset piirteet, sitten voivat olla? Suomen kielen tutkijoista ainakin Arppe (2008) ja Jantunen (2004) käsittelevät synonyymien välisiä eroja, ja näiden tutkimusten perusteella voidaan luetella tekijöitä, jotka erottavat synonyymeja toisistaan. Arppe (2008: 9) listaa tällaisia piirteitä taustaksi tutkimukselleen, jossa tilastollisilla monimuuttujamenetelmillä etsii AJATELLA-, MIETTIÄ-, POHTIA- ja HARKITA- verbien keskinäiseen valintaan vaikuttavia tekijöitä. Toisekseen Jantunen (2004: 57–62) käsittelee HYVIN, KOVIN- ja OIKEIN- sanojen synonymiaa koskevan tutkimuksensa taustaksi synonymian kontekstuaalista luonnetta ja synonyymien välisiä eroja. Hän perustaa käsittelynsä erityisesti leksikaalisen yksikön kontekstuaalisen pro-

fiilin osatekijöihin (Jantusen kokoamasta mallista ks. Jantunen 2004: 31). Sekä Arppen että Jantusen tutkimuksessa korostuu synonymian kontekstuaalinen luonne: juuri sanan kontekstille tyypilliset tekijät vaikuttavat siihen, miksi tietty sana valitaan paradigmaattisesta synonymiasuhteesta lauseeseen. Seuraavaksi luettelen Arppen ja Jantusen teorioiden pohjalta synonyymeja erottavia tekijöitä.

Siihen, mikä sana synonyymijoukosta valitaan, vaikuttaa Arppen mukaan ensinnäkin kielenulkoisen konteksti, kuten rekisteri, tyyli ja tilanne. Esimerkiksi sanomalehdissä politiikkaosastoilla käytetään tyypillisesti POHTIA-verbiä MIETTÄ-verbin sijaan, joka taas on tyypillinen urheilusuilla. (Arppe 2008: 9; 2009: 2.) Myös alueellisen ja sosiaalisen vaihtelun voidaan katsoa kuuluvan kielenulkoiseen kontekstiin; Jantunen (2004: 58) mainitsee nämä tekijät samassa yhteydessä, missä puhuu tyyliin ja rekisteriin liittyvistä piirteistä. Hän toteaa, että valinta esimerkiksi synonyymien VIHTA ja VASTA välillä riippuu erityisesti puhujan murretaustasta.

Toisekseen synonyymien valintaan vaikuttaa sanan ympärillä oleva lähikonteksti eli koteksti. Tällaisista piirteistä sekä Arppe mainitsee ensinnäkin leksikaalisen kotekstin; tästä voidaan Jantunen ja monen muun tutkijan tavoin käyttää myös termiä kollokointi. Esimerkiksi sanojen TÄRKEÄ ja KESKEINEN kollokointi on erilaista: on tyypillisempää sanoa *tärkeä seikka* kuin *keskeinen seikka* ja vastaavasti esimerkiksi *keskeinen kysymys* kuin *tärkeä kysymys* (Arppe 2008: 8–9; 2009: 2; Jantunen 2001: 179–180; 2004: 58–60.) Kollokaatioiden lisäksi myös kieliopilliset kotekstuaaliset piirteet erottavat synonyymeja. Kieliopillisista tekijöistä puhuessaan Arppe (2008: 9) käyttää käsitettä syntaktinen argumenttirakenne ja Jantunen (2004: 60) termiä kolligaatio. Kumpikin heistä mainitsee esimerkkinä Biberin ym. (2006 [1998]: 95–100) havainnot BEGIN- ja START-verbien syntaktisista eroista: Biber ym. havaitsivat muun muassa, että BEGIN-verbi saa objektikseen *to*-lauseeseen selvästi useammin kuin START-verbi. Esimerkkinä suomen kielen synonyymeja erottavista kolligaatioista Jantunen (2001: 182–183; 2004: 60) mainitsee myös tekemänsä havainnon TÄRKEÄ- ja KESKEINEN-sanojen erilaisesta sanaluokkakolligaatiojakaumasta: vasemmanpuoleisten vieruskolligaattien joukossa adverbikolligaatin osuus on TÄRKEÄ-sanalla huomattavasti suurempi kuin KESKEINEN-sanalla.

Synonyymeja erottavia kotekstuaalisia tekijöitä ovat siis leksikaalinen sekä kieliopillinen koteksti, mutta lisäksi Arppe mainitsee kotekstuaalisista tekijöistä myös syntaktisten argumenttien semanttiset piirteet. Esimerkiksi MIETTÄ-verbin tekijä on usein yksilöllinen (*minä mietin*), kun taas synonyymisen POHTIA-verbin puolestaan kollektiivinen (*hallitus pohtii*). (Arppe 2002: 27–28; 2008: 9; 2009: 2.) Leksikaalisen ja kieliopillisen kotekstin lisäksi eroja synonyymien kotekstista voidaan siis etsiä myös tarkastelemalla siinä esiintyviä semanttisia piirteitä. Niin ikään Jantunen (2004: 61) ottaa esiin kotekstin semanttiset piirteet: semanttinen preferenssi ja semanttinen prosodia voivat erottaa synonyymeja toisistaan. Kuten jo luvussa 2.2.3 mainittiin, esimerkiksi OIKEIN-sanalla on

voimakas semanttinen preferenssi 'arvo'; sitä taas ei ole sen synonyymeilla KOVIN ja HYVIN (Jantunen 2004: 138–139). Vastaavasti esimerkki synonyymien erilaisesta semanttisesta prosodiasta on Tuurin (2014: 96) havainto, jonka mukaan lähimerkityksiset adverbit VUOKSI, TAKIA, TÄHDEN, ANSIOSTA ja JOHDOSTA eroavat toisistaan muun muassa siten, että adverbilla ANSIOSTA on selvä positiivinen semanttinen prosodia lähikontekstissaan, kun taas neljällä muulla negatiivinen semanttinen prosodia on yleisin. Toisin kuin Arppen toteamaa piirrettä tutkittaessa, semanttista preferenssiä ja prosodiaa tutkittaessa ei selvitetä kontekstin tietyissä syntaktisissa funktioissa toimivien sanojen merkitystä, vaan kontekstin semantiikkaa tutkitaan yksinkertaisesti luokittelemalla merkitysluokkiin kollokaatteja tietyltä kollokaatioväliltä. Kyse on joka tapauksessa sekä Arppella että Jantusella kontekstin semanttisista piirteistä.

Kielenulkoisen kontekstin sekä kontekstuaalisten piirteiden lisäksi synonyymit poikkeavat toisistaan myös sen osalta, missä morfologisissa muodoissa niillä on taipumus esiintyä. Esimerkiksi PYSTYÄ- ja VOIDA-verbeistä ensin mainittu on tyypillinen imperfekti- ja partisiippimuodoissa, kun taas jälkimmäistä käytetään etenkin passiivi- ja konditionaalimuodoissa. (Arppe 2001: 9–13; 2009: 2–3.) Jantunen nimittää tätä piirrettä paradigmaattiseksi morfologiseksi preferenssiksi. Synonyymeja erottavista tekijöistä tämä ei ole kontekstuaalinen tekijä siinä mielessä, että se koskee lekseemin sisäisiä valintoja. (Jantunen 2004: 27–31; 60–61.)

Kaiken kaikkiaan siis niin kielenulkoisen konteksti, kontekstin leksikko, kielioppi ja merkitys-piirteet sekä lekseemien morfologiset muodot ovat tekijöitä, jotka voivat erottaa synonyymeja toisistaan. Usein erot eivät kuitenkaan muodostu vain yhdestä tekijästä, vaan sitä vastoin monesta eri piirteestä. Sekä Jantunen (2004: 150–151) että Arppe (2008: 9–11) toteavatkin, että kontekstuaaliset piirteet eivät ole toisistaan erillisiä ilmiöitä, vaan eri piirteiden väliseen vuorovaikutukseen on kiinnitettävä huomiota. Eri tekijöiden välinen yhteisvaikutus korostuu erityisesti Arppen monimuuttujamenetelmiä hyödyntävässä tutkimuksessa.

Tässä tutkielmassa en tutkielman rajallisuuden vuoksi keskity eri tekijöiden välisen yhteisvaikutuksen tarkasteluun, vaikka se ehdottomasti onkin olennainen osa synonyymien valintaa. Edellä mainituista piirteistä kielenulkoisen konteksti ei ole tutkimuksen keskiössä, vaikka esimerkiksi rekisterikohtaista vaihtelua EHKÄ- ja MAHDOLLISESTI-sanojen käytössä saattaakin esiintyä; tutkimusaineistonani on ainoastaan sanomalehtikieli. En tarkastele kvantitatiivisesti myöskään kieliopillisten tekijöiden vaikutusta synonyymien valintaan. Samoin paradigmaattinen morfologinen preferenssi jää tutkimuksen ulkopuolelle – tosin sen vaikutusta modaalisanon valintaan ei voitane muutenkaan pitää merkittävänä, sillä näiden sanojen sisäinen morfologinen vaihtelu on vähäistä (ks. luku 3.3.3 muodoista *ehkä* ja *ehkäpä*).

2.4 Modaaliset partikkelit ja adverbis

2.4.1 Modaalisuus

Modaalisuus on kielen semanttinen alue, joka liittyy asiointilan todenmukaisuutta ja toteutumismahdollisuuksia koskeviin arvioihin. Modaalisilla ilmauksilla kielenpuhuja ilmaisee, onko asiointila varma, välttämätön, todennäköinen, mahdollinen, epävarma tai mahdoton, pakollinen tai luvallinen, toivottava tai epätoivottava tai riippuvainen ulkoisista tai sisäisistä edellytyksistä. (VISK § 1551.) Modaalisuutta voidaan ilmaista suomen kielessä useilla eri kieliopillisilla ja leksikaalisilla keinoilla, esimerkiksi modaaliverbeillä (SAADA, VOIDA, TÄYTYÄ), modaalisilla adverbeilla ja partikkeleilla (TAKUULLA, VARMASTI, TODENNÄKÖISESTI, KAI), verbimoduksilla (imperatiivi, potentiaali, konditionaali) ja erilaisilla lausekonstruktioilla (*X:n on tehtävä/on hyvä tehdä Y*). (VISK § 1558, 1562, 1601.) Tutkimani lekseemit EHKÄ ja MAHDOLLISESTI kuuluvat modaalisuuden eri ilmaisukeinoista modaalisten adverbien ja partikkeleiden kategoriaan.

Ison suomen kieliopin mukaan modaalisuutta voidaan katsoa olevan neljää eri lajia; niitä ovat episteeminen, deonttinen ja dynaaminen modaalisuus sekä praktinen välttämättömyys. Niistä episteemisellä modaalisuudella tarkoitetaan tietoon, uskoon tai päätelmiin perustuvaa (puhujan) arviota asiointilan mahdollisuudesta, todennäköisyydestä tai varmuudesta: *Hän lähtee ehkä ~ varmasti huomenna, Hän saattaa lähteä huomenna*. Episteemisen modaalisuuden alalaji on evidentiaalisuus, jossa on kyse sen ilmaisemisesta, miten puhuja on tietonsa saanut: toisen käden tietona (*kuulemma*), itse päättelemällä (*siis*), omana havaintonaan (*näköjään*) vai esimerkiksi muistiinsa perustuen (*muistaakseni*). Deonttisella modaalisuudella taas ilmaistaan puhujan tahdon tai puheyhteisön normien mukaista velvoitetta tai lupaa: *Koulut saavat itse päättää monista ainevalinnoista, Sinun on syytä tulla ajoissa*. Dynaaminen modaalisuus puolestaan ilmaisee tilanteen olosuhteista johtuvaa pakkoa tai fyysistä kykyä toimia: *Meidän täytyy ehtiä viimeiseen junaan, Jaksan vielä valvoa*. Praktisella välttämättömyydellä taas tarkoitetaan arkipäätelyn perusteella selittyvää välttämättömyyttä. Tätä modaalisuuden lajia esiintyy esimerkiksi erilaisissa toiminta- ja muissa ohjeissa: *Ovi on avattava sisältä käsin, koska kahva on pudonnut, Taivasta kannattaa kuitenkin katsoa vähintään puoli tuntia, sillä valoilmiot tulevat satunnaisesti*. (VISK § 1551–1552, 1555, 1557, määritelmät s.v. deonttinen modaalisuus, dynaaminen modaalisuus, episteeminen modaalisuus.)

Modaalisuuden eri lajien väliset merkityserot eivät kuitenkaan ole näin selviä, kun tarkastellaan aitoa kielenkäyttöä. Silti modaalisista adverbeista ja partikkeleista voidaan todeta, että ne rajoittuvat ilmaisemaan nimenomaan episteemistä modaalisuutta. (VISK § 667, 1552, 1614.) Esimerkiksi lauseiden ”*Ehkä siinä on jokin vika*” ja ”*Mahdollisesti olen väärässä*” (esimerkit KS s.v. *ehkä*,

mahdollisesti) modaalisuus liittyy nimenomaan arvioon siitä, mitä puhuja pitää todennäköisenä: modaalisen ilmauksen avulla ilmaistaan, että puheena olevassa asiassa saattaa puhujan arvion mukaan olla jokin vika ja että puhuja arvioi voivansa olla väärässä. Näillä modaalisilla sanoilla ei siis ilmaista esimerkiksi sitä, mikä on luvallista (deonttinen modaalisuus), mihin tilanteen osapuoli kykenee fyysisesti (dynaaminen modaalisuus) tai mikä on arkipäätelyn perusteella välttämätöntä (praktinen välttämättömyys). Pidän lekseemejä EHKÄ ja MAHDOLLISESTI siis episteemistä modaalisuutta ilmentävinä modaalisanoina.

Edellä on modaalisuuden ja erityisesti episteemisen modaalisuuden määrittelyssä käytetty sanaa ”puhujaa”, ja Jaakola (2011: 515) toteaa, että tutkimuksissa modaalisuutta lähestytään usein puhujakeskeisesti. Esimerkiksi Palmerin (1981: 43) mukaan modaalisuudella ilmaistaan puhujan mielipidettä tai asennetta lauseen ilmaisemaa asiointilaa kohtaan. Kahdenkymmenen viime vuoden aikana on kuitenkin nostettu esiin myös intersubjektiiivinen näkökulma: on osoitettu, että episteemiset ja evidentialiset kielenaineokset ovat myös intersubjektiiivisia (Lehtinen 1997: 234; Jaakola 2011: 515). Modaalisuutta ei tulisi määritellä vain puhujan suhtautumiseksi asiointilaan, vaan Nissin (2000: 12) mukaan pikemmin keinoiksi, ”joilla puhuja suhteuttaa esittämäänsä asiointilaa erilaisiin mahdollisiin maailmoihin” (ks. myös Orpana 1988: 34). Voidaan olettaa, että modaalisanat EHKÄ ja MAHDOLLISESTI eivät kaikissa tapauksissa ilmaise puhujan omaa mielipidettä tai näkemystä, vaan laajemmin puhujan tekemää arviota siitä, miten puheilla oleva asiointila suhteutuu erilaisiin mahdollisiin maailmoihin. Hakulinen (1996: 14) ja Lehtinen (1997: 241–242) toteavat, ettei epävarmuuden ilmauksia käyttävä kielenkäyttäjä ole välttämättä subjektiivisesti epävarma. Esimerkiksi Lehtisen aineistoesimerkissä puhuja käyttää EHKÄ-partikkelia lieventämään sanomaansa kritiikkiä: ”*Ehkä meilläkin on puutetta siitä yhteydestä*”, vaikka toisaalla puhujan puheenvuorossa käy ilmi, että hän nimenomaan on sitä mieltä, että ”meillä” on puutetta yhteydestä. Esimerkistä huomataan, että modaalisuudessa on siis kyse laajemmasta asiasta kuin vain yksinkertaisesti sen ilmaisemisesta, että puhuja suhtautuu epävarmasti tiettyyn asiointilaan. Modaalisanojen funktiot kielessä voivat siis olla moninaiset, mikä voi osaltaan vaikuttaa siihen, millaisia kollokaatteja EHKÄ ja MAHDOLLISESTI saavat aineistossa.

2.4.2 Modaalisanojen EHKÄ ja MAHDOLLISESTI sanaluokka

Kuten edellä mainittiin, sanat EHKÄ ja MAHDOLLISESTI kuuluvat modaalisten adverbien ja partikkelien kategoriaan. EHKÄ on modaalinen partikkeli ja MAHDOLLISESTI modaalinen adverbi (VISK § 680, 792). Sekä partikkelit että adverbit luokitellaan Isossa suomen kieliopissa taipumattomien sa-

nojen ryhmään erotukseksi taipuvista sanoista, joita ovat nominit ja verbit. Partikkelit ja adverbit kuitenkin eroavat toisistaan joiltain osin. Ensinnäkin partikkelit eivät saa määritteitä, mutta adverbit taas voivat saada. (VISK § 792.) Sellainen lauseke kuin **hyvin ehkä* on selkeästi kielitajun vastainen, mutta sen sijaan lauseketta *hyvin mahdollisesti* ei nähdäkseni ole pidettävä sellaisena, ja sitä esiintyykin Suomen kielen tekstikokoelman aineistoissa (esimerkiksi virkkeessä ”*Hän luopuu politiikasta, mutta hyvin mahdollisesti hänestä kuullaan vielä paljon*”), joskaan ei juuri tämän tutkielman aineistona olevissa nimenomaisissa vuosikerroissa. Toiseksi partikkelit ovat useimmiten yksimorfeemisia, kun taas adverbit saattavat sisältää erotettavan pääteaineksen, esimerkiksi *sti*-johtimen (VISK § 371–372, 1601). Tutkittavana oleva adverbi MAHDOLLISESTI on johdettu MAHDOLLINEN-adjektiivista juuri *sti*-adverbinjohtimen avulla. Kolmanneksi partikkeleita ja adverbeja erottaa toisistaan se, että partikkelit ovat aina täysin taipumattomia, muitta joillakin adverbeilla sen sijaan on osittainen taivutusparadigma (VISK § 792). Tutkimukseni kohteena olevilla sanoilla ei kuitenkaan ole eroa tässä suhteessa: sen lisäksi, että sanaa EHKÄ voidaan partikkelina pitää automaattisesti taipumattomana (ks. kuitenkin lukua 3.3.3 muodoista *ehkei* ja *ehkäpä*), myös adverbi MAHDOLLISESTI on aineistossani aina taipumaton.

Kaiken kaikkiaan voidaankin todeta, että vaikka EHKÄ ja MAHDOLLISESTI lähtökohtaisesti kuuluvat eri sanaluokkiin, niiden välillä ei ole kuitenkaan sellaista kieliopillista eroa, jonka vuoksi niiden toisiinsa rinnastaminen olisi epärelevanttia. Ylipäätään raja modaalisten partikkeleiden ja adverbien välillä on nimittäin epämääräinen: ne ovat funktioltaan hyvin samanlaisia keskenään, sillä molemmat toimivat lauseessa kommenttiadverbiaaleina eli sellaisina adverbiaaleina, joilla puhuja ilmaisee kantansa lauseen ilmentämään asiointilaan (VISK § 1000, 1601, määritelmät s.v. kommenttiadverbiaali). Samoin synonyymisanakirjat ja Kielitoimiston sanakirja mainitsevat sanat toistensa synonyymeina, kuten luvussa 3.2 esitän. Lekseemien EHKÄ ja MAHDOLLISESTI rinnastamista toisiinsa voidaan siis pitää relevanttina ja tarkoituksenmukaisena.

3 AINEISTO JA METODIT

3.1 Aineisto

Hyödynnän tutkimuksessa Tieteen tietotekniikan keskuksen, CSC:n, ylläpitämää Kielipankkia, joka koostuu useista kielentutkimuksen aineistoista. Niistä käytän hyödyksi Suomen kielen tekstikoelmaa. Se on noin 180 miljoonan saneen laajuinen sähköinen korpus, joka sisältää erilaisia alakorpuksia, kuten eri sanomalehtien vuosikertoja ja kustantamoiden eri vuosina julkaisemia kaunoja ja tietokirjallisuuden kokoelmia (CSCa). Useimmat alakorpuksat ovat 1990-luvulta. Käytän aineistoa Lemmie-käyttöliittymän kautta. Tutkin sanomalehtitekstejä, sillä sanomalehtiä luetaan paljon ja niiden kieli on miltei kaikille kielenpuhujille tuttua. Sanomalehdissä myös käsitellään laajasti eri aihepiirejä, joten voidaan ajatella, että niiden kielen tutkiminen antaa kattavan kuvan suomen yleiskielestä.

Suomen kielen tekstikokoelman alakorpuksista tutkielmani aineistona on 9,8 miljoonan saneen laajuinen aineisto, joka koostuu sanomalehti Karjalaisen vuoden 1998 vuosikerrasta sekä Aamulehden vuoden 1995 vuosikerrasta. Molemmat vuosikerrat ovat tutkimuksen tekohetkellä noin kaksikymmentä vuotta vanhoja, joten niiden kieli saattaa jossain määrin poiketa nykyisestä sanomalehtikielestä. Koivusalo (1992) käsittelee lehtikielen muuttumista: hän vertailee vuosien 1936 ja 1986 Uuden Suomen numeroita ja havaitsee, että esimerkiksi sanastossa on tapahtunut viidessäkymmenessä vuodessa muutoksia, sillä virallisesta ja muodollisesta tyylistä on siirrytty lähemmäs ihmisten arkista puhetta. Esimerkiksi tämän havainnon perusteella voidaan olettaa, että 1990-luvun lopun ja puolivälin kieli verrattuna nykyiseen, 2010-luvun puolivälin kieleen on erilaista. Toisaalta Koivusalon tutkima aikajänne on jopa viisikymmentä vuotta, joka on selvästi pitempi aika kuin tämän tutkimuksen aineiston ikä. Lisäksi Mäkinen (1992: 213–214) toteaa, että lehtikieli muuttuu yleensä hitaasti, koska sen tarkoitus on olla yhtenäistä ja luettavaa; muutokset herättävät lukijassa usein torjuntaa. Niinpä kahdenkymmenen vuoden ikäisen tutkimusaineiston kieli ei todennäköisesti kuitenkaan poikkea nykyisestä sanomalehtikielestä merkittävästi. Suomen kielen tekstikokoelman valintaa tutkielman aineistoksi puoltaa myös se, että korpusta käytetään Lemmie-käyttöliittymän avulla, jossa voi hyödyntää erilaisia korpustyökaluja. Tutkijan ei esimerkiksi tarvitse etsiä kollokatioita teksteistä itse, vaan se voidaan tehdä tarpeeseen kehitetyllä erillisellä sähköisellä toiminnolla.

Kun pyrkimyksenä ei ole tutkia esimerkiksi yhdelle, tietylle toimittajalle tyypillistä kieltä vaan ylipäätään sanomalehtikielelle tyypillistä fraseologiaa, tutkimuksen aineistona on nähdäkseni tarpeen käyttää useampaa kuin yhtä sanomalehteä. Yhden lehden vuosikerrassa on nimittäin toden-

näköisesti paljon samojen toimittajien kirjoittamia tekstejä, ja on mahdollista, että tietty kirjoittaja suosii kirjoittaessaan juuri tietynlaista fraseologiaa. Esimerkiksi Tarja Isokääntä (2014: 12, 69) tutkii pro gradu -työssään, millaiset kollokaatiot ovat tyypillisiä sanomalehti Keski-suomalaisen päätoimittajan Erkki Laatikaisen kirjoittamille kolumneille, ja hän löytääkin monia juuri Laatikaisen kielelle tyypillisiä, mutta muuten yleiskielelle epätavanomaisia kollokaatioita. Tutkielmassa pyrin nimenomaan välttämään sitä, että yhden, tietyn kirjoittajan kielenkäyttötapa vaikuttaisi merkittävästi tuloksiin. Siksi aineistona on kahden eri sanomalehden vuosikerrat. Tosin toki jo yhdessä sanomalehdessä on monien eri toimittajien kirjoittamia tekstejä, mutta mitä useampia lehtiä tutkitaan, sitä useampia eri kirjoittajia teksteillä kuitenkin on ja sitä vähemmän yhden toimittajan kieli siten painottuu tuloksissa.

Tutkin sitoutumattomien, tiettyyn aihepiiriin erikoistumattomien päivälehtien kieltä. Suomen kielen tekstikokoelmassa on tekstejä yhteensä yhdeksästä tällaisesta sanomalehdestä, ja kaikista näistä lehdistä on useita eri vuosikertoja tai muita alakorpuksia. Aineistot ovat keskenään hyvin erikokoisia: pienin alakorpus on noin 61 000 saneen laajuinen ja laajin puolestaan miltei 22 miljoonan saneen laajuinen. Kaikista alakorpuksista Aamulehden vuoden 1995 vuosikerta ja Karjalaisen vuoden 1998 vuosikerta ovat valikoituneet aineistokseni ensinnäkin siksi, että pidän niitä laajuudeltaan riittävän samanlaisina. Aamulehden vuosikerrassa on noin 3,8 miljoonaa sanetta ja Karjalaisen vuosikerrassa puolestaan noin 6 miljoonaa sanetta. Karjalainen on siis tosin selvästi laajempi, mutta koska Aamulehden osuus tutkimuksen koko aineiston sanemäärästä on kuitenkin noin 39 %, voidaan katsoa, ettei Karjalaisesta peräisin oleva kieli painotu aineistossa liian paljon. Mainittujen alakorpusten valintaan on vaikuttanut lisäksi se, että aineistoksi on syytä valita lehdet eri vuosilta. Eri sanomalehdissä käytetään nimittäin usein täsmälleen samoja uutisia. Saman uutisen esiintyminen kummassakin eri alakorpuksessa vaikuttaisi aineistoni sanojen frekvensseihin ja voisi mahdollisesti vääristää tuloksia, mutta tämä ongelma vältetään valitsemalla sellaiset alakorpuksat, jotka ovat peräisin keskenään eri vuosilta: eri vuosina sanomalehdissä ei oletettavasti koskaan esiinny täsmälleen samoja uutisia. Lisäksi nimenomaisten alakorpusten valikoitumiseen tutkimusaineistoksi on vaikuttanut se, ettei tutkittavan aineiston pidä olla liian laaja, ja se, että pyrin valitsemaan kaikista vaihtoehtoisista lehdistä mahdollisuuksien mukaan melko tuoreet vuosikerrat.

Aineiston koostaminen kahdesta eri lehdestä on tarpeen paitsi siksi, että vältettäisiin yhden toimittajan kielen piirteiden korostuminen, myös siksi, että kirjoittajien murretausta saattaa vaikuttaa sanomalehtikielenkin. Voidaan pitää mahdollisena, että esimerkiksi kolumneissa ja muissakin tekstilajeissa kirjoittajien murre vaikuttaa jossain määrin esimerkiksi tekstissä käytettyyn sanastoon, jota tutkielmani nimenomaan koskee. Kun murteita vertaillaan toisiinsa nimenomaan sanaston kannalta, suomen kielen murteet voidaan Leinon, Hyvösen ja Salmenkiven (2006: 34) mukaan jakaa

synkronisesti tarkasteltuna pohjois-, itä- ja länsimurteisiin. Käyttämästäni kahdesta sanomalehdestä Aamulehti ilmestyy Tampereella eli länsimurteiden alueella ja Karjalainen puolestaan Joensuussa eli itämurteiden alueella, joten mukana on aineistoa kahdelta eri murrealueelta. Voidaan olettaa, että tämä ehkäisee tietyn murrealueen sanaston painottumista tutkittavien sanojen kollokaateissa. Lisäksi on toisaalta selvää, että sanomalehtikielessä on kyse yleiskielestä, johon kirjoittajan murretausta ei loppujen lopuksi vaikuta oletettavasti kovin paljon. Kaiken kaikkiaan aineistoa voidaan siis pitää suomen sanomalehtikieltä kattavasti ja luotettavasti kuvaavana aineistona, vaikka 1990-luvun kielen tutkiminen ei tarjoakaan tietoa aivan tuoreimmista kielenkäytön tavoista.

Vaikka tutkin sanomalehtikieltä yhtenä yksikkönä, sanomalehdissä esiintyy useita erilaisia tekstejä. Eri juttutyypeikseen voidaan erottaa pääkirjoitus, arvostelu, reportaasi, taustajuttu, uutinen, pakina, mielipidekirjoitus, kommentti, kinalojuttu ja asiantuntija-artikkeli (Miettinen 1984: 74; Makkonen-Craig 2005: 210). Näitä kaikkia voidaan pitää keskenään eri tekstilajeina, ja on mahdollista, että tietyssä tekstilajissa sanat EHKÄ ja MAHDOLLISESTI saavat erilaisia kollokaatteja ja muita kontekstuaalisia piirteitä kuin jossakin toisessa. Tässä tutkimuksessa huomio ei kuitenkaan kohdistu siihen, miten sanojen kontekstuaaliset piirteet eroavat sanomalehden eri juttutyyppeiden tai muuten eri tekstilajien kesken, vaan käsittelen sanomalehtikieltä yhtenä kielimuotona. Näin voidaan tehdä, sillä sanomalehti muodostaa yhtenäisen, omana kulttuurituotteena toimivan kokonaisuuden, jota myös aiemmin on tutkittu yhtenä kokonaisuutena (ks. esim. Viinikainen 2010).

3.2 Perusteet tutkittavien lekseemien valinnalle

Valitsen tutkittavaksi lekseemeiksi eli tutkimuksen havaintoyksiköiksi sellaiset sanat, joilla ei ole kielenulkoista referenttiä. Kielenulkoisessa maailmassa usein yhdessä esiintyvät asiat esiintyvät todennäköisesti myös kielessä toistensa läheisyydessä (Sinclair 1991a: 110). Näin ollen voidaan olettaa, että tutkimukset saattavat paljastaa joidenkin sanojen tyypillisiksi kollokaateiksi erityisesti sellaisia sanoja, jotka jo kielenulkoisessa maailmassa esiintyvät usein kyseisen sanan yhteydessä. Esimerkiksi KOIRA-lekseemiä tutkittaessa saatettaisiin havaita sen tyypilliseksi kollokaatiksi vaikkapa verbi HAUKKUA. Tämä ei kuitenkaan nähdäkseni olisi erityisen kiinnostavaa, ja voidaan jopa ajatella, että tämäntyyppisissä tapauksissa tutkimustulokset kuvaisivat enemmän kielenulkoista maailmaa kuin kieltä. Oletettavasti erityisesti sisältösanoissa on tämäntyyppisiä sanoja. Sisältösanat ovat tyypillisesti substantiiveja, adjektiiveja ja verbejä, ja niille on luonteenomaista referentiaalinen merkitys eli juuri kyky viitata kielenulkoisessa maailmassa oleviin asioihin (VISK § 438).

Valitsenkin tutkimuskohteekseni mieluummin funktiosanoja, koska etenkin niitä tutkimalla voidaan olettaa saatavan tietoa nimenomaan kielestä sen sijaan, että ennen kaikkea kielenulkoisen maailman olosuhteet vaikuttaisivat tutkimustuloksiin. Funktiosanat ovat tavallisesti lyhyempiä kuin sisältösanat, niiden tehtävä on selkeämmin kieliopillinen tai keskustelun kulkuun liittyvä ja ne ovat sanaluokaltaan partikkeleita, apuverbejä, persoona-, demonstratiivi- ja relatiivipronomineja (VISK § 438). Sana EHKÄ modaalisena partikkelina ja sana MAHDOLLISESTI modaalisena adverbina voidaan luokitella funktiosanoihin. Joka tapauksessa niillä ei ole kielenulkoista referenttiä, joten sikäli ne täyttävät hyvin tutkimuskohteeni kriteerit. Sisältösanoja nimitetään toisinaan myös leksi-kaalisiksi sanoiksi ja niiden vastinparia, funktiosanoja, muoto- tai kieliopillisiksi sanoiksi (Kangasniemi 1997: 39). Tässä tutkielmassa käytän nimityksiä sisältösanat ja kieliopilliset sanat.

Suomen kielen keskenään synonyymisten kieliopillisten sanojen fraseologiaa on tutkittu sanojen HYVIN, KOVIN ja OIKEIN osalta (Jantunen 2004; sanojen kieliopillisiksi luokittelemisesta tarkemmin mts. 69) sekä sanojen VUOKSI ja TAKIA osalta (Suoraniemi 2011). Lisäksi Päiviö (2007) on tutkinut sanojen ASTI ja SAAKKA synonymiaa, joskaan ei fraseologisen tutkimuksen käsitteitä hyödyntäen; monet Päiviön havainnoista olisivat kuitenkin esitettävissä myös fraseologian näkökulmasta käsin. Modaalisten adverbien tai partikkelien synonymiaa ei ole tutkittu, mutta modaalisuus on keskeinen piirre kielessä, joten niiden tutkimiselle on sijaa kielentutkimuksen kentässä. Ylipäätään adverbien ja partikkelien valintaa tutkimuksen kohteeksi puoltaa se, että ne ovat yleensä taipumattomia. Taipuvienkin sanojen synonymiaa voidaan toki hyvin tutkia fraseologian näkökulmasta, mutta tällöin tutkimuksessa on usein hyvä ottaa huomioon lekseemin eri taivutusmuodot, mikä kasvat-
taa tutkimuksen laajuutta ja vähentää siten mahdollisuutta keskittyä sanojen piirteisiin perusteellisesti.

Vaikka tutkimusasetelma onkin luonteeltaan kahden sanan piirteitä vertaileva, kahden sanan vastakkainasettelua ei tule liiaksi korostaa: asetelmaa ei tule ymmärtää niin, että sanat EHKÄ ja MAHDOLLISESTI olisivat suomen kielen ainoat kyseistä merkitystä edustavat sanat. Sanat kuuluvat sitä vastoin laajempaan synonyymiryhmään. Tässä tutkielmassa ei ole kuitenkaan mahdollista tutkielman rajallisen laajuuden vuoksi tutkia kaikkia synonyymiryhmän sanoja, vaan on keskityttävä kahteen havaintoyksikköön. Olen käyttänyt lekseemien valinnassa hyväkseni ensinnäkin Ison suomen kieliopin modaalisten adverbien ja partikkelien luetteloa (VISK § 1601), sillä siitä voidaan poimia monta eri samaan merkityskenttään kuuluvaa, epävarmuutta ja epätodennäköisyyttä ilmaisevaa sanaa: MAHDOLLISESTI, KAI, KAIKETI, EHKÄ, KENTIES, KUKATIES, VARMAAN, ARVATENKIN. Näistä partikkeli EHKÄ on valikoitunut toiseksi havaintoyksiköksi siksi, että se on sanomalehtikielessä merkityskentän sanoista kaikkein yleisin. Myös esimerkiksi Jantunen (2004: 68) pitää synonymiaa koskevan fraseologisen tutkimuksensa yhtenä kriteerinä havaintoyksiköiden valinnassa sitä, että

sanojen on oltava tarpeeksi frekventtejä, jotta niiden kontekstuaalisia piirteitä voidaan kuvata riittävän luotettavasti. Suomen sanomalehtikielen taajuussanaston (CSCb) mukaan sana EHKÄ on kaikkien suomen sanomalehtikielessä esiintyvien sanojen yleisyysjärjestyksessä sijalla 335, ja sana on listassa korkeammalla kuin mikään sen edellä luetelluista synonyymeista. Muiden edellä lueteltujen EHKÄ-sanon synonyymien yleisyysjärjestys on seuraava: MAHDOLLISESTI (sijalla 1525), VARMAAN (sijalla 1670), KAI (sijalla 2069), KENTIES (sijalla 2340). Sanat KAIKETI, KUKATIES ja ARVATENKIN eivät ylipäätään ole mukana kyseisessä taajuussanastossa, johon on otettu mukaan suomen sanomalehtikielen 9996 yleisintä lekseemiä.

Toisen havaintoyksikön valinnassa voidaan hyödyntää edellä lueteltuja frekvenssitietoja. Sanat KAIKETI, KUKATIES ja ARVATENKIN on syytä jättää tutkimuksen ulkopuolelle, sillä ne eivät ole frekventtejä sanoja. Sanat MAHDOLLISESTI, VARMAAN, KAI ja KENTIES eivät nekään ole aivan kärkipäässä taajuusjärjestyksessä, mutta niitä voidaan kuitenkin pitää suhteellisen frekventteinä ja fraseologiseen tutkimukseen hyvin soveltuvina; esiintymiä suuressa aineistossa on joka tapauksessa paljon. Olennaista toisen havaintoyksikön valinnassa on tarkastella myös sanojen merkitystä, ja siinä voidaan hyödyntää sanakirjoja ja sanastoja. Ensinnäkin niistä voidaan löytää perusteita sille, että EHKÄ-sanon synonyymeja ovat pikemmin MAHDOLLISESTI ja KENTIES kuin VARMAAN ja KAI: sekä Antti Leinon ja Pirkko Leinon Synonyymisanastossa (1996), Kivimiehen Synonyymisanastossa (1974) että Kielitoimiston sanakirjassa (KS) mainitaan hakusanan EHKÄ sana-artikkelissa kaksi ensin mainittua sanaa, mutta ei lainkaan kahta viimeksi mainittua. Niinpä VARMAAN ja KAI voidaan jättää tutkittavien vaihtoehtojen joukosta pois – samoin harvinaiset ilmaisut KUKATIES, PIAN AIKAA, MENE TIEDÄ ja MITÄMAKS, jotka esimerkiksi Kivimies (mp.) mainitsee EHKÄ-sanon synonyymeina. Myös esimerkiksi adverbit LUULTAVASTI ja ILMEISESTI rajautuvat vaihtoehtojen joukosta pois, sillä niitä ei ylipäätään mainita synonyymisanastoissa EHKÄ-sanon synonyymeina. Näin ollen tutkittaviksi vaihtoehtoiksi EHKÄ-sanon ohella jäävät sanat MAHDOLLISESTI ja KENTIES. Olen valinnut niistä havaintoyksiköksi sanan MAHDOLLISESTI siksi, että se on edellä kuvattujen frekvenssitietojen mukaisesti yleisempi suomen sanomalehtikielessä kuin KENTIES. Ison suomen kieliopin, Kielitoimiston sanakirjan, synonyymisanastojen ja taajuussanaston tietojen perusteella tutkimukseni havaintoyksiköiksi ovat siis valikoituneet modaalisisista sanoista synonyymit EHKÄ ja MAHDOLLISESTI.

Äidinkielen suomenpuhujan kielelliseen intuitioon nojautuen voidaan todeta, että tutkittavat synonyymit eivät ole kuitenkaan täydellisiä synonyymeja, vaan niillä tuntuu olevan jonkinlaisia keskinäisiä merkityseroja. Esimerkiksi VUOKSI- ja TAKIA-sanojen väliseen synonymiaan verrattuna sanojen EHKÄ ja MAHDOLLISESTI synonymia tuntuu kenties hieman löyhemmältä, sillä intuitiivisesti saattaa ehkä tuntua siltä, että nämä kaksi modaalisanaa eivät ole yhtä usein korvattavissa lauseessa toisillaan kuin ensin mainitut synonyymit. Absoluuttista synonymiaa ei kuitenkaan ylipäätään lain-

kaan pidetä kielessä mahdollisena, vaan synonyymeina pidetyt sanat ovat siis aina lähisynonyymeja. Lähtökohta lekseemien EHKÄ ja MAHDOLLISESTI synonymian tutkimiselle on hedelmällinen juuri sillä tavoin, että sanat yhtäältä tuntuvat olevan synonyymisia ja toisaalta eroavan toisistaan. Eroja ei voitane selittää kovin tarkasti ja luotettavasti yksinomaan intuition perusteella: esimerkiksi Sinclair (1991a: 4) toteaa, että kielenkäyttäjän intuitiivinen oletus kielenkäytöstään on usein ristiriidassa sen kanssa, miten hän todellisuudessa käyttää kieltä (ks. kuitenkin luku 3.3.5 intuition roolista tutkimuksessa). Näin ollen työni tarkoitus on korpusmenetelmin selvittää, mitä kyseiset erot mahdollisesti ovat.

3.3 Fraseologisen yksikön tutkiminen korpuslingvistisin menetelmin

3.3.1 Tilastolliset testit

Kollokaatioita voidaan tutkia eri menetelmin. Yksinkertaisin menetelmä perustuu sanojen yhteisiintymien absoluuttiseen frekvenssiin. Aineistosta haetaan sanat, jotka esiintyvät toistensa läheisyydessä tiettyä frekvenssiarvoa useammin, ja ne poimitaan kollokaatiovaihtoehdoiksi. Kaikki niistä eivät kuitenkaan ole luettavissa tutkittavan sanan kollokaatioiksi. (Barnbrook 1996: 88; Metin & Karaođlan 2011: 177.) Esimerkiksi tähän menetelmään perustuva englannin kielestä tehty aineistohaku ehdottaa TIME-sanalle ('aika') kollokaateiksi muun muassa sanoja THE, AND, FOR ja THAT. Nämä sanat kieliopillisina sanoina kuitenkin esiintyvät kielessä tiheästi monien muidenkin sanojen kanssa kuin TIME-sanan kanssa. Toisilla menetelmillä sanalle saattaisi löytyä vaikkapa sellaisia kollokaatteja kuin WASTE 'tuhlata', DEVOTE 'käyttää', SPEND 'käyttää, viettää', SPARE 'ylimääräinen, vapaa' ja LIMIT 'rajoittaa'. (Baker, Hardi & McEnery 2006: s.v. *collocation*.) Ainoastaan frekvensseihin perustuva menetelmä ei siis paljasta kollokaatioita, jotka olisivat tyypillisiä juuri TIME-sanalle.

Toinen menetelmä kollokaatioiden tutkimiseen on tilastollinen testaus. Kollokaatio on tilastollisesti merkitsevä, jos on riittävän matala todennäköisyys sille, että sanat esiintyvät toistensa läheisyydessä sattumalta (Barnbrook 1996: 94). Niinpä tarkastelemalla kollokaatioiden tilastollista merkitsevyyttä voidaan löytää ne kollokaatiot, joiden esiintyminen ei johdu sattumasta. Useimmin kollokaatioanalyysissa käytettyjä tilastollisia testejä ovat Barnbrookin (mts. 94–100) mukaan MI-testi, z-testi ja t-testi. Sekä z-testillä että t-testillä kuvataan sitä, miten todellinen ja odotuksenmukainen frekvenssi eroavat toisistaan keskihajonnan pohjalta. Niiden avulla ilmaistaan, kuinka suurella varmuudella voidaan sanoa, että sanojen välillä on myötäesiintymä, joka ei johdu sattumasta.

MI-testi puolestaan ei perustu kollokaatioiden todellisten ja odotettujen frekvenssien vertailuun keskihajonnan pohjalta, vaan siinä verrataan noodin ja kollokaatin myötäesiintymän todennäköisyyttä siihen todennäköisyyteen, että noodi ja kollokaatti esiintyisivät sattumanvaraisesti, toisistaan riippumattomina. MI-testillä siis mitataan sen todennäköisyyden määrää, millä yhden sanan havaitseminen tekstissä tuottaa toisen todennäköisen havaitsemisen. (Clear 1993: 278–281; Barnbrook 1996: 94–100; Jantunen 2004: 78.) Testien tulokset ilmaistaan lukuina, ja analyysiin otetaan mukaan ne kollokaatiot, joilla testin tulos ylittää tilastollisen merkitsevyyden raja-arvon. Z-testissä raja-arvona voidaan pitää lukua 3, t-testissä lukua 2 ja MI-testissä lukua 1,58 (Barnbrook 1996: 96–99). Toisaalta eri tutkimuksissa käytetään joskus hieman eri raja-arvoja; esimerkiksi Jantunen (mts. 79–80) asettaa tutkimuksessaan MI-testin raja-arvoksi luvun 3. Tilastollisesti merkitsevistä kollokaatioista osa on merkitsevempiä kuin toiset, ja pyrin tarkastelemaan analyysissa tätäkin.

MI-testi eroaa siis kahdesta muusta testistä siinä, mitä asiaa se kuvaa. Testien kesken on kuitenkin toinenkin merkittävä ero. MI-testi ja z-testi nostavat herkästi kollokaateiksi sellaisia sanoja, jotka esiintyvät aineistossa melko harvoin, kun taas t-testin tuloksissa painottuvat sellaiset sanat, joiden frekvenssit ovat suhteellisen suuria (Barnbrook 1996: 101). Usein tämä näkyy esimerkiksi siinä, että MI-testi nostaa t-testiin verrattuna kollokaateiksi enemmän sisältösanoja, kun taas t-testi taas painottaa kieliopillisia sanoja (Clear 1993: 281; Stubbs 1995a: 12–15). Kaiken kaikkiaan mitään kuvatuista testeistä ei voida pitää yksiselitteisesti toista parempana kollokaatioiden tutkimuksessa. Pikemmin voidaan sanoa, että on hyvä käyttää kollokaatioista niin paljon tietoa kuin mahdollista. (Barnbrook 1996: 101.)

Tämän tutkielman analyysin perustana käytän kahta testiä, t-testiä ja MI-testiä. Tarkoitukseni on tutkia tyypillisiä, yleisiä kollokaatioita, mihin t-testi soveltuu hyvin. Toisaalta t-testi edustaa vain yhtä tilastollisen merkitsevyyden laskentatapaa, kollokaatioiden todellisten ja odotettujen frekvenssien vertailua keskihajonnan perusteella. Jotta myös MI-testin kuvaama, edellä selostettu ilmiö tulisi otetuksi analyysissa huomioon, käytän toisena testinä MI-testiä. Sekä t-testiä että MI-testiä hyödyntävää menetelmää käyttää myös Stubbs (1995a: 13) sekä tämän esittämää menetelmää tutkimuksessaan hieman soveltava Jantunen (2004: 79–80). Stubbsin menetelmässä analysoitaviksi otetaan kollokaatiot, jotka ylittävät raja-arvon molemmissa testeissä, ja Jantunen puolestaan ottaa niiden lisäksi analyysiinsa myös ne kollokaatiot, jotka ylittävät raja-arvon vain t-testissä.

Tässä tutkielmassa aineistoksi otetaan mukaan kolmenlaiset kollokaatiot: molempien testien mukaan merkitsevät, pelkän t-testin mukaan merkitsevät sekä pelkän MI-testin mukaan merkitsevät. Koska mukana ovat myös sellaiset kollokaatiot, jotka ylittävän raja-arvon vain toisessa testissä, aineisto on laajempi kuin Stubbsin menetelmän avulla saatava aineisto olisi; täten aineisto kuvaa monipuolisemmin modaalisanon kollokointia. Jantunen puolestaan jättää kuitenkin näistä kolmes-

ta kollokaatiotyypistä pois kollokaatiot, jotka vain MI-testi nostaa esille, sillä muuten esiin saattaisi nousta harvinaisia ja mahdollisesti idiosynkraattisia, tietyille tekstintuottajille ominaisia kollokaatioita (Jantunen 2004: 79). Onkin mahdollista, että koska tässä tutkielmassa MI-testi korostuu Jantusen menetelmään nähden, osa kollokaatioista saattaa olla harvinaisia tapauksia. Jantusen (mts. 66) aineistona on kuitenkin kokonaisia kirjoja. Tässä tutkielmassa aineistona on kahden eri sanomalehden vuosikerrat, joten tiettyjen kirjoittajien idiolekti oletettavasti ei sikälikään korostu (ks. luku 3.1). Toisekseen MI-testikään ei luonnollisesti nosta esille mitä tahansa sanoja, vaikka sanat saattavatkin olla harvinaisia, vaan sekin tuo esiin t-testin lailla myötäesiintymiä, joiden esiintyminen ei todennäköisesti johdu sattumasta. Lisäksi kaikenlaisten tilastollisten merkitsevien kollokaatioiden mukaan ottaminen kasvattaa aineiston laajuutta. Tutkielmassa tarkoituksena on luokitella sanoja semanttisiin ryhmiin jo kollokaatioanalyysissä, ja tämä onnistuu oletukseni mukaan parhaiten, kun aineistoon ei ole tehty merkittäviä rajauksia. Pois on jätetty tietenkin sellaiset kollokaatiot, jotka eivät ole merkitseviä kummankaan tilastollisen testin perusteella. Tutkielman luvussa 4.1 tarkastellaan, miten t-testin ja MI-testin antamat tulokset eroavat toisistaan tämän tutkielman aineiston osalta; tarkastelun kohteena siis on, nostavatko testit esiin keskenään erilaisia kollokaatioita.

Olen hakenut tilastollisesti merkitsevät kollokaatiot Lemmie-käyttöliittymän Collocation Table -hakutyökalun avulla. Olen asettanut ohjelman asetuksissa kollokaation vähimmäisfrekvenssiksi 5, jolloin kaikkein harvinaisimmat tapaukset suodattuvat tuloksista pois. Myös Jantunen (2004: 80) käyttää tutkimuksessaan minimifrekvenssiä 5. Olen hakenut EHKÄ- ja MAHDOLLISESTI-sanojen t-testin mukaan tilastollisesti merkitsevät kollokaatiot asettamalla raja-arvoksi 2, ja näin saatu aineisto on tutkielman liitteenä 1. MI-testissä olen puolestaan käyttänyt raja-arvoa 1,58, ja haun tuloksena saatu MI-testin mukaisten tilastollisesti merkitsevien kollokaatioiden luettelo on työn liitteenä 2.

3.3.2 Kollokaatiöväli

Kollokaatioita tutkittaessa on tärkeää määritellä tarkastelualue eli kollokaatiöväli. Sillä tarkoitetaan sitä kotekstin laajuutta, jonka sisällä kollokaatin on sijaittava noodiin nähden (Martin, Al & van Sterkenburg 1983: 84). Kollokaatiövälin määrittämiseen tulee kiinnittää huomiota, koska se, miltä alueelta kollokaatteja tarkastellaan, vaikuttaa selvästi tutkimuksen tuloksiin. Tutkijat eivät kuitenkaan ole yksimielisiä siitä, mikä on sopiva havaintoalue. (Stubbs 1995a: 8.) Sinclair (1991a: 170) määrittelee kollokaatiövälin viideksi sanaksi noodin molemmin puolin. Tätä ilmaistaan merkinnällä ”5: 5”, ja tällöin kollokaateiksi lasketaan siis vain ne sanat, jotka ovat enintään viiden sanan etäisyydellä noodista eli hakusanasta. Martinin ym. (1983: 84) mukaan tilastolliset testit osoittavatkin,

että yli 95 % kaikesta relevantista informaatiosta voidaan saada asettamalla kollokaatiöväliksi 5:5; välimerkkejä ei tällöin lasketa mukaan. Toisaalta Stubbsin (1995a: 8) mukaan käytetään usein tarkastelualueita 2:2 ja 3:3.

Se, mikä tarkastelualue on tutkimukseen optimaalisin, saattaa riippua siitä, mitkä sanat ovat tutkimuksen kohteena. Siksi kollokaatiövälän valinnassa voidaan pyrkiä ottamaan huomioon tutkimuksen kohteena olevien lekseemien erityispiirteet sen sijaan, että valinta perustettaisiin esimerkiksi vain aiempiin tutkimuksiin. Näin tekee ainakin Jantunen (2004: 83–84), joka rajaa tutkimuksensa tarkastelualueen laskemalla, kuinka monta sanaa keskimäärin esiintyy hänen tutkimiansa sanojen HYVIN, KOVIN ja OIKEIN ympärillä olevissa tietyissä kieliopillisissa rakenteissa. Myös sanoille EHKÄ ja MAHDOLLISESTI voidaankin pyrkiä määrittelemään kollokaatiöväli tarkastelemalla ensin, millaisissa kieliopillisissa rakenteissa tai lauseasemissa ne tyypillisesti esiintyvät.

Modaaliset adverbit ja partikkelit kuuluvat yhdessä asenneadverbien (esim. ONNEKSI, LUULTAVASTI, VALITETTAVASTI) kanssa kommenttiadverbien ja -partikkelien ryhmään. Niiden funktio ei ole luonnehtia esimerkiksi konkreettisia olioita, vaan tuoda esille puhujan asennoituminen tai kannanotto puheena olevaan asiaan. Kommenttiadverbit ja -partikkelit ovat lauseessa muihin adverbeihin verrattuna irrallisempia, ja niiden vaikutusala kattaa tavallisesti koko lauseen: esimerkiksi lauseen ”*Hän tulee ehkä vielä*” merkitys on ’ehkä on niin että hän tulee vielä’. Vaikutusala-termillä viitataan siihen, että tietynlaiset ilmaukset tulkitaan toisen vastaavanlaisen ilmauksen antamassa kehyksessä. (VISK § 667, määritelmät s.v. vaikutusala, lause-esimerkki § 963.) Kyse on modaalisten sanojen kohdalla siis siitä, mihin lauseen sanoihin epävarmuuden tai epätodennäköisyyden merkitys liitetään ja mihin taas ei. Esimerkiksi sanajärjestys voi määrittää vaikutusalaa: lauseessa *Syömme huomenna ehkä lohta* modaalisuuden merkitys kohdistuu substantiiviin LOHI, kun taas lauseessa *Syömme lohta ehkä huomenna* modaalisanan vaikutusala kohdistuu adverbiin HUOMENNA. Voitaisiin ehkä ajatella, että EHKÄ- ja MAHDOLLISESTI-sanat kollokoisivat vahvimmin vaikutusalaan olevien sanojen kanssa. Nähdäkseni kollokaatiöväliä ei kuitenkaan lopulta ole syytä perustaa siihen, kuinka laaja tutkittavien sanojen vaikutusala on, sillä tällöin kollokaatiövälän laskennassa ei otettaisi huomioon esimerkiksi monien kieliopillisten sanojen sijaintia lauseessa. Esimerkiksi MUTTA-konjunktio ei lähtökohtaisesti kuulu ollenkaan modaalisanojen vaikutusalaan, mutta pitää säilyttää se mahdollisuus, että sana kuitenkin kollokoi modaalisanojen kanssa. Myöskään virkettä ei ole syytä pitää kollokaatiövälän laskentaperusteena, sillä pitkissä virkkeissä tarkastelualueeksi tulisi usein laaja alue. Kollokaatilla tarkoitetaan siis sanaa, joka esiintyy noodin lähikontekstissa, eikä esimerkiksi tapauksissa (1) ja (2) ole syytä pitää modaalisanojen lähikontekstina virkkeen toisessa päässä olevia sanoja:

- 1) Kansallistunteen nostattaminen ei **ehkä** ollut Kontrapunktin tärkein tehtävä, mutta niin siinä vain kävi, että klassisen musiikin tietovisasta tuli aito kilpailu, jota seurasi innostuneemmin kuin Mika Häkkisen saavutuksia formula-auton ratissa.
- 2) Lentoaseman laitaa kulkeva tie Lempäälää kohti muutetaan väliaikaisesti yksisuuntaiseksi, sillä sitä käytetään varaparkkialueena muiden paikkojen **mahdollisesti** täyttyessä.

Jos siis tarkastelualue laskettaisiin virkkeen pituuden perusteella, siitä voisi tulla liian laaja. Perustankin kollokaatiovälin laskemisen siihen, kuinka pitkiä keskimäärin ovat ne lauseet, joissa sanat EHKÄ ja MAHDOLLISESTI esiintyvät. Esimerkiksi edellä olevissa esimerkeissä lähikontekstina voidaan nähdäkseni pitää pikemmin samaa lausetta kuin virkettä. Olen tehnyt laskelman tutkielman aineistosta systemaattisen otannan avulla, Jantusen (2004: 83–84) periaatteita mukaillen. Ensin olen luonut ohjelman avulla konkordanssin EHKÄ- ja MAHDOLLISESTI-sanoista ja sitten laskenut konkordanssilistan ensimmäiseltä riviltä, kuinka monta sanaa esiintyy yhtäältä noodin vasemmalla ja toisaalta oikealla puolella samassa lauseessa. Seuraavaksi olen laskenut vastaavat luvut sijalla 21 olevalta konkordanssiriviltä ja edennyt täten koko aineiston läpi laskemalla tiedot joka 20. konkordanssiriviltä. Lopuksi olen laskenut keskiarvon noodin oikealla puolella samassa lauseessa esiintyvien sanojen lukumäärästä ja vastaavan luvun vasemmalla esiintyvistä sanoista. Laskelman mukaan aineistossani esiintyy keskimäärin 3,28 sanaa noodin vasemmalla puolella samassa lauseessa kuin noodi. Noodin oikealla puolella samassa lauseessa puolestaan on keskimäärin 4,35 sanaa. Siksi kollokaatiovälinä tutkimuksessani on 3: 4, eli rajaan tarkastelualueen koskemaan noodista vasemmalla korkeintaan kolmen sanan päässä olevia sanoja sekä oikealla korkeintaan neljän sanan päässä olevia sanoja. Kyseessä on näin ollen epäsymmetrinen kollokaatioväli, jollaisesta puhuvat myös esimerkiksi Jantunen (mp.) ja Lindquist (2009: 73). Ohjelman asetuksissa säädän myös, että kollokaattien pitää esiintyä samassa virkkeessä. Täten siis aineistoksi ei oteta kollokaatteja, jotka ovat välillä 3:4, mutta kuuluvat eri virkkeeseen kuin noodi. Käytän kollokaatioväliä 3:4 sekä kollokaattien että semanttisen preferenssin tutkimisessa.

3.3.3 Kollokaatioanalyysin aineiston muodostaminen

Olen suorittanut tutkittavien saneiden haun Lemmien Concordance-hakutyökalulla asettamalla hakukriteeriksi sanan EHKÄ ja sanan MAHDOLLISESTI perusmuodot. EHKÄ-lekseemin esiintymiä on aineistossa yhteensä 3637 ja MAHDOLLISESTI-lekseemin esiintymiä 786. Näistä EHKÄ esiintyy 3462 kertaa muodossa *ehkä*, 174 kertaa muodossa *ehkäpä* ja yhden kerran muodossa *ehkÄ*. Viimeksi mainitussa esiintymässä on oletettavasti vain kirjoitusvirhe, eikä ole luonnollisestikaan syytä ajatella, että se olisi erillinen sananmuoto. Sen sijaan on pohdittava, tuleeko tutkimuksessa tehdä eroa

muotojen *ehkä* ja *ehkäpä* välille. On mahdollista käsitellä eri sananmuotoja erillisinä yksiköinä: esimerkiksi Sinclair (1991a: 8) esittää, että yksittäisiä sananmuotoja voisi jopa pitää omina leksi-kaalisina yksiköinä. Eri sananmuotojen kollokaattien käsittely erikseen veisi kuitenkin paljon tilaa tutkielmasta, ja koska aineistossani sananmuodon *ehkäpä* määrä kaikista lekseemin esiintymistä on kuitenkin melko vähäinen, vain 4,8 %, en tarkastele sananmuotoja erikseen. MAHDOLLISESTI-sanan osalta tätä ongelmaa ei lainkaan ole, sillä lekseemin 786 esiintymästä kaikki ovat sananmuotoja *mahdollisesti*.

Suomen kielessä esiintyy myös muoto *ehkei* ja sen eri taivutusmuotoja, kuten *ehken* ja *ehket*. Tässä tutkielmassa en luokittele näitä muotoja samoiksi lekseemeiksi partikkelin EHKÄ kanssa, sillä Ison suomen kieliopin mukaan kyse on syntagmaattisesta sulaumasta, jotka ovat ”yhdeksi sanaksi hahmotettavia kahden sanavartalon muodostamia kokonaisuuksia” (VISK § 139, 141). Syntagmaattisella sulaumalla *ehkei* on eri merkitys kuin sanalla EHKÄ: ensin mainittu sisältää kiellon merkityksen ja jälkimmäinen taas ei. Lisäksi käyttämässäni kieliopillisesti koodatussa korpusaineistossakaan muotoja ei ole koodattu samoiksi lekseemeiksi, joten jätän *ehkei*-muodon kokonaan tarkastelun ulkopuolelle. Myöskään kollokaatteja en tarkastele sananmuodoittain, vaan lekseemeittäin.

EHKÄ-sanan tilastollisesti merkitseviksi kollokaateiksi MI-testissä nousee myös kolme sanayhdistelmää, ”paras mahdollinen”, ”sen takia” ja ”jonkin verran”. Niiden nouseminen kollokaattien joukkoon johtuu Lemmie-ohjelmassa olevasta koodauksesta, mutta todellisuudessa ne eivät ole varsinaisia kollokaatteja, vaan klustereita (ks. käsitteestä luku 2.2.1). Tarkoitukseni ei ole tutkia klustereita, joten jätän nämä kaksisanaiset kollokaatit analyysin ulkopuolelle. Jätän analyysistä pois niin ikään sellaiset kollokaatit, joita kyllä esiintyy vähimmäisfrekvenssin mukaisesti vaaditut viisi kertaa, mutta joita ei ole vähintään viittä sellaista esiintymää, jotka olisivat peräisin keskenään erilaisista teksteistä. Korpuksessa nimittäin esiintyy jostain syystä joitakin tekstejä useita kertoja. Esimerkiksi sanojen EHKÄ ja USEASTI viidestä yhteisesiintymästä kaksi on peräisin keskenään samalaisesta Aamulehden tekstistä ja samasta virkkeestä ”*Ehkä pallo pomppi tänään useammin minulle, Sampras aprikoi*”. USEASTI-adverbin lisäksi jätän analyysistä pois adjektiivin KUULUISA ja substantiivit MÖKKI ja VAIHE. Aineiston luotettavuuden kannalta on ongelmallista, että jotkin tekstit esiintyvät useita kertoja. Kun tämän vaikutus kuitenkin poistetaan kaikkein harvimmin esiintyvien sanojen osalta jättämällä nämä sanat pois analyysistä, tutkimusaineistoa voidaan pitää luotettavana. Lisäksi ehdottomasti suurin osa aineiston kirjoituksista esiintyy kuitenkin vain kerran, mikä sekin osoittaa tutkimusaineiston luotettavuutta.

Kollokaatteja koskevassa analyysiluvussa 4 etenen siten, että ensin esitän t-testin mukaiset kymmenen merkitsevintä kollokaattia kummallekin modaalisanalle ja sen jälkeen MI-testin mukaiset vastaavat tulokset. Sen jälkeen vertailen lyhyesti t- ja MI-testin tuloksia. Sitten analysoin sana-

luokittain kummankin sanan kaikkia tilastollisesti merkitseviä kollokaatteja, siis myös muita kuin kymmentä merkitsevintä: ensin käsittelen substantiivikollokaatit ja sen jälkeen edelleen adjektiivivi-, numeraali-, pronomini-, verbi- ja adverbikollokaatit ja lopuksi vielä partikkeli- ja adpositiokollokaatit. Kuten Biber ym. (2006 [1998]: 5) toteavat, korpustutkimuksenkaan ei kuitenkaan pidä perustua ainoastaan ilmausten frevensseihin, vaan kvantitatiivista tietoa on selitettävä kvalitatiivisesti. Kvantitatiivisen tarkastelun lisäksi pohdinkin kollokaatioita myös kvalitatiivisemmalla otteella: tarkastelen muun muassa sitä, miksi juuri kyseiset sanat mahdollisesti ovat valikoituneet modaalisanojen kollokaateiksi. Luvussa 5 siirryn puolestaan analysoimaan sanojen semanttista preferenssiä, joten seuraavassa luvussa 3.3.4 esittelen sen tutkimiseen liittyviä metodeja.

3.3.4 Semanttisen preferenssin analyysin aineiston muodostaminen

Olen muodostanut aineiston semanttisen preferenssin analyysia varten Lemmie-ohjelman Concorde-toiminnon avulla. Olen hakenut ohjelman hakutoiminnolla ensin näkyviin kaikki 3637 EHKÄ-partikkelin esiintymää siten, että ne näkyvät yhtenä konkordanssilistana. Rivit on järjestetty V4-asemassa olevien sanojen eli noodista neljäntenä vasemmalla olevien sanojen mukaiseen aakkosjärjestykseen, sillä V4-asemassa olevat sanat eivät kuulu tarkastelualueeseen. Sen jälkeen olen poiminut systemaattisella otannalla hakutuloksista 60 konkordanssiriviä. Ensimmäiseksi riviksi olen valinnut rivin 1, ja sen jälkeen olen ottanut EHKÄ-sanana hakutuloksesta joka 125. rivin. Tällä tavalla saadaan 30 satunnaista konkordanssiriviä. Tämä aineisto osoittautui kuitenkin liian suppeaksi, joten lisäksi olen poiminut rivin 63 ja siitä eteenpäin joka 125. rivin. Tällä tavalla olen saanut 60 satunnaista riviä, joissa EHKÄ-sana esiintyy. Vastaavasti olen hakenut näkyviin MAHDOLLISESTI-adverbin kaikki 786 esiintymää. MAHDOLLISESTI-sanana hakutuloksesta olen ottanut ensimmäisen rivin ja sen jälkeen joka 27. rivin, mutta aineiston laajentamiseksi valinnut lisäksi rivin 14 ja siitä eteenpäin jälleen joka 27. rivin; täten myös MAHDOLLISESTI-sanana analyysi perustuu 60:een systemaattisella otannalla valittuun konkordanssiriviin. Tarkastelualueena semanttisen preferenssin analyysissa on kollokaatioanalyysin tavoin 3:4, eli tarkastelen analyysissa modaalisanojen kolmea vasemmanpuoleista ja neljää oikeanpuoleista kollokaattia. Lisäksi semanttisen preferenssin analyysi, kuten kollokaatioanalyysikin, perustuu vain niihin sanoihin, jotka esiintyvät modaalisanan kanssa samassa virkkeessä. Jos siis virke alkaa vaikkapa EHKÄ-sanalla, en ota analyysissa huomioon vasemmanpuoleisia kollokaatteja, vaikka kolme lähintä vasemmanpuoleista kollokaattia lähtökohtaisesti kuuluvatkin tarkastelualueeseen. Tarkastelualueen koko siis hieman vaihtelee konkordanssiriveittäin, ja siksi modaalisanoilla onkin keskenään eri määrä sanoja tarkastelualueella. EHKÄ-sanana tarkastelu-

alueella on 343 sanaa ja MAHDOLLISESTI-sanan vastaavalla alueella puolestaan 366 sanaa. Analyysin pohjana olevien kollokaattien keskenään erilainen määrä ei kuitenkaan ole ongelmallinen asia, sillä semanttisen preferenssin analyysi perustuu suhteellisiin osuuksiin, ei absoluuttisiin lukumääriin. Modaalisanojen suhteelliset osuudet ovat siis verrattavissa toisiinsa.

Semanttisen preferenssin tutkimista varten modaalisanojen kotekstissa tarkastelualueella olevat sanat on luokiteltava semanttisiin ryhmiin. Aineiston sanat ovat keskenään hyvin erityyppisiä, sillä joukkoon kuuluu eri sanaluokkien sanoja, niin substantiiveja, adjektiiveja, pronomineja, numeraaleja, verbejä kuin taipumattomiakin sanoja. Semanttisen preferenssin analyysissä pyritään jättämään sanaluokkajako huomiotta ja etsimään preferenssejä sanaluokkarajojen yli; olennaista on siis ryhmitellä sanat luokkiin nimenomaan niiden merkityksen perusteella (Sinclair 2004: 34; Jantunen 2004: 24–25, 133). Työssäni olen noudattanut semanttisten luokkien muodostamisessa osin korpusvetoista lähestymistapaa. Korpusvetoisessa tutkimuksessa aineisto on keskeisessä asemassa eikä sitä käytetä vain tukemaan esimerkiksi valmiiksi määriteltyä kategorisointia, vaan aineistosta esiin nousevia frekventtejä esimerkkejä voidaan käyttää pohjana kielellisille kategorioille (Tognini-Bonelli 2001: 84). Olenkin ryhmitellyt sanoja sen mukaan, millaisia semanttisia luokkia aineiston perusteella näyttää muodostuvan. Toisaalta olen sanoja luokitella osin myös nojannut aiemmassa tutkimuksessa käsiteltyihin erilaisiin semanttisiin luokkiin: olen käyttänyt luokituksessa apuna Pajusen (2001) luokittelua suomen kielen verbeistä, Jantusen (2004) semanttisen preferenssin analyysin luokkia sekä Ison suomen kieliopin esitystä lauseiden välisistä semanttisista suhteista (VISK § 1121).

Kuten Jantunen (2004: 135) toteaa, luokittelu voi kuitenkin olla ongelmallista ja monimutkaista. Modaalisanojen EHKÄ ja MAHDOLLISESTI tarkastelualueen sanasto onkin erittäin monenkirjavaa. Esimerkiksi EHKÄ-sanan tarkastelualueelle kuuluvat muun muassa seuraavat ilmaukset: ”*Jari juoksee suomalaisista ehkä terävimmän vauhdin ja sinkauttaa – –*” ja ”*Se on ehkä ainoa asia, joka ei – –*”. Ensin mainitusta lauseesta voidaan havaita ’urheiluun’ liittyviä sanoja JUOSTA, VAUHTI ja SINKAUTTAA, mutta jälkimmäisen ilmaisun lekseemit eivät ole merkitykseltään kovin spesifisiä, minkä vuoksi voi olla vaikeaa löytää merkitysryhmiä, joihin ne kuuluisivat. Toisaalta ensin mainitun lauseen urheiluun viittaavat sanatkaan eivät ole yksinkertaisia luokitella, kun otetaan huomioon koko muusta aineistosta muodostuvat semanttiset luokat. Aineistossa nimittäin on paljon liikettä ja muuta prosessia ilmaisevia verbejä, minkä vuoksi herääkin kysymys, olisiko verbit JUOSTA ja SINKAUTTAA syytä luokitella luokkaan ’urheilu, kilpailu’ vai ’prosessia’ kuvaavaan merkitysryhmään. Vastaavanlaisia ongelmallisia sanoja on aineistossa paljon. Sana SÄVELTÄJÄ voisi kuulua sekä luokkaan ’ihmiset’ että luokkaan ’kulttuuri’, sana EPÄILLÄ sekä ’mentalisuus’- että ’epävarmuus’-luokkaan ja niin edelleen.

Luokittelussa ensisijainen luokittelukriteeri on luonnollisesti semanttinen: ryhmät on muodostettava niin, että samoihin luokkiin luokitellaan sanat, jotka muistuttavat toisiaan semanttisesti ja eroavat vastaavasti muista semanttisista ryhmistä. Kunkin sanan merkitystä on arvioitava nimenomaan saneena, sen perusteella, minkä merkityksen sana saa nimenomaisessa kontekstissaan. Esimerkiksi kirjailija Väinö Linnaan viittaava sukunimi LINNA voitaisiin monessa kontekstissa luokitella 'ihmiset'-luokkaan, mutta sen esiintymän seuraavassa virkkeessä katson kuuluvan pikemmin luokkaan 'kulttuuri': ”*Mutta ehkä Linna ei sopinut sen hetkiseen elämäntilanteeseeni*”. Tässä lauseessa sana LINNA nimittäin viittaa metonymisesti Väinö Linnan teokseen *Täällä Pohjantähden alla*, ei Linnaan itseensä henkilönä. Etu- ja sukunimestä koostuvat henkilönnimet olen luokitellut yhdeksi lekseemiksi: esimerkiksi *Juha Lind* on aineistossa yksi tarkasteltava yksikkö.

Monet sanaesiintymät ovat merkitykseltään kahden tai jopa useamman eri luokan rajalla. Tällaiset sanat olen sijoittanut siihen luokkaan, kumpaa merkitystä sana kontekstissaan nähdäkseni vahvemmin edustaa. Jotta aineistosta saataisiin aikaan semanttinen luokittelu, ratkaisuja voidaan siis tehdä silläkin perusteella, mitä merkitysluokkaa sana kaikkein vahvimmin edustaa, vaikkei ero olisikaan aivan selkeä. Esimerkiksi substantiivi ROMAANI olisi luokiteltavissa joko 'kulttuuri'-luokkaan tai 'viestintä'-luokkaan, mutta olen katsonut sen ilmentävän vahvemmin kulttuurituotteen kuin viestintäkanavan merkitystä ja luokitellut sen siksi 'kulttuuri'-kategoriaan. Kyse on tulkinnoista, joista kaikki eivät ole ehdottomia. Olenkin noudattanut luokituksessa myös periaatetta, jonka mukaan merkitykseltään tarkat, kapea-alaiset ryhmät ovat luokittelussa ensisijaisia ja laaja-alaiset ryhmät vasta toissijaisia. Tällaisia laaja-alaisia, toissijaiseksi jääviä ryhmiä ovat 'prosessi' ja 'arvo, laatu'. Esimerkiksi 'prosessi'-ryhmään olisi luokiteltavissa moni sana, joka lisäksi olisi sijoitettavissa johonkin toiseen, merkitykseltään kapea-alaisempaan ryhmään. Esimerkiksi edellä mainitut verbit JUOSTA ja SINKAUTTAA siis ilmentävät semanttisesti niin 'urheilua, kilpailua' kuin 'prosessiakin', ja koska 'urheilu, kilpailu' on nähdäkseni merkitykseltään kapea-alaisempi kategoria kuin hyvin laaja-alainen 'prosessi'-kategoria, olen luokitellut nämä verbit 'urheilu, kilpailu'-luokkaan. 'Prosessi'-ryhmän lisäksi myös luokka 'arvo, laatu' jää luokittelussa toissijaiseksi, jos siihen luokiteltavissa oleva sana ilmentää samalla myös jotakin muuta merkitysryhmää. Esimerkiksi adjektiivin PIENI olen näin ollen luokitellut 'määrä, intensiteetti'-ryhmään, vaikka se ilmiselvästi ilmentää myös eräänlaista 'laatua'. Mitään sanaesiintymää en ole siis luokitellut kahteen eri luokkaan, vaikka siinä merkityksen puolesta niin voisikin tehdä – kyse on siis siitä, että sana tulee luokitelluksi merkitykseltään kapea-alaisempaan ryhmään silloin, kun sen voitaisiin nähdä kuuluvan kahteen eri luokkaan. Kaiken kaikkiaan tutkijan intuitio on väistämättä vahvasti mukana semanttisen preferenssin analyysissä. Luokittelu kuitenkin perustuu samalla myös sanakirjatietoon: kunkin sanaesiinty-

män analyysin tueksi olen tarvittaessa tarkastellut lekseemin merkitysryhmiä Kielitoimiston sanakirjan (KS) avulla.

Myös luokittelussa muodostuvien ryhmien koko on nähdäkseni jossain määrin otettava huomioon. Aiemmasta tutkimuksesta olisi löydettävissä enemmänkin semanttisia luokitteluja kuin edellä mainitut, tässä hyödyntämäni tutkimukset. Esimerkiksi Iso suomen kielioppi (VISK § 1121) mainitsee lauseiden välisistä semanttista suhteista paitsi tässä analyysissä hyödyntämäni additiivisen, kausaalisen, kontrastiivisen ja konsessiivisen suhteen, niin myös esimerkiksi vertailevan suhteen. Sana KUIN olisikin ryhmiteltävissä luokkaan 'vertailu', mutta koska kummallakin modaalisanalla on tarkastelualueellaan vain yksi KUIN-kollokaatti eikä nähdäkseni lainkaan muita 'vertailua' ilmentäviä sanoja, ei ole tarkoituksenmukaista ottaa semanttiseen luokitteluun mukaan kategoriaa 'vertailu'. Analyysi on myös paremmin hallittavissa, kun semanttisia ryhmiä on kohtuullinen määrä; tätä periaatetta soveltaa analyysissään myös Jantunen (2004: 135).

Luokittelun tarkemmat perusteet esitetään luvussa 5.1 ja koko analysoitava aineisto merkitysryhmiin luokiteltuna liitteessä 3. Muodostettujen merkitysluokkien perusteella tarkastelen luvussa 5.2, mitkä merkitysluokat EHKÄ- ja MAHDOLLISESTI-sanoilla ovat suuria eli mitä semanttisia preferenssejä modaalisanoilla on. Tarkastelen kummankin sanan preferenssejä erikseen merkitysryhmien suuruuden mukaisessa järjestyksessä. Sen jälkeen samassa luvussa keskityn vielä vertailemaan, missä merkitysluokissa sanojen prosentiosuuksilla on tilastollisesti merkitsevimmät erot eli miten synonyymit eroavat toisistaan semanttisten preferenssien osalta. Tutkin erojen tilastollista merkitsevyyttä käyttämällä hyödyksi suhteellisen osuuden testausta ja z-testiä.

Z-testin arvot kertovat, onko tulos tilastollisesti melkein merkitsevä, merkitsevä, erittäin merkitsevä vai ei mitään näistä. Merkitsevyytaso kertoo, kuinka suuri riski on, että saatu ero johtuu sattumasta. Kun z-arvo on vähintään 1,96, tulos on tilastollisesti melkein merkitsevä. Tällöin on 5 prosentin mahdollisuus ($p < 0,05$), että kyse on puhtaasta sattumasta. Z-arvon ollessa vähintään 2,58 kyseessä on tilastollisesti merkitsevä ero ja pelkän sattuman todennäköisyys on 1 prosenti ($p < 0,01$). Tilastollisesti erittäin merkitsevän eron kriittinen arvo on puolestaan 3,29, ja tällöin sattuman todennäköisyys on enää 0,1 prosenttia ($p < 0,001$). (Kanji 1993: 161; Heikkilä 2014a: 182–191; Heikkilä 2014b.) Täten z-testin avulla tutkitaan, missä määrin sattuma vaikuttaa eroihin, joita havaitaan modaalisanojen välillä sen suhteen, mitkä merkitysluokat niiden kotekstissa painottuvat.

3.3.5 Aineisto ja intuitio tutkimusmenetelmien perustana

Kuten edellä on todettu, korpuslingvistiikan hyötynä on tutkimuksen perustuminen kielen todelliseen käyttöön. Korpuukset tarjoavat myös laajan materiaalin tutkimuksen pohjaksi. Kvantitatiivisessa tutkimuksessa on tärkeää, että aineisto on riittävän laaja, jotta yleistyksiä voidaan tehdä. Etuna on myös aineistojen helppokäyttöisyys: kun hyödynnetään valmiita aineistoja, aikaa ei kulu aineiston keräämiseen. Korpustutkimus ei silti ole ainoa käyttökelpoinen ja varteenotettava menetelmä sanastontutkimukseen, kuten lähisynonyymien vertailuun, vaan myös esimerkiksi kielenkäyttäjän intuitioon perustuvien kyselytestimenetelmien avulla voidaan saada tietoa synonymiasta. Vanhatalo (2003) saakin kyselytestimenetelmään perustuvassa tutkimuksissaan selville lähisynonyymien KESKEINEN ja TÄRKEÄ eroista ja yhtäläisyyksistä sellaista tietoa, mitä esimerkiksi Jantusen (2001) korpustutkimus ei tavoita. Kyselytestimenetelmän etuna korpustutkimukseen verrattuna on muun muassa, että sen avulla voidaan ottaa huomioon myös koehenkilöiden kielentuntemus, kognitiiviset käsitykset ja muu kielentuotostilanteeseen liittyvä konteksti (Vanhatalo 2003: 353).

Aineistopohjainen menetelmä sekä intuitioon perustuva menetelmä, jota esimerkiksi kyselytestimenetelmä siis edustaa, on usein nähty toisilleen pitkälti vastakkaisina metodeina (vastakkainasettelusta ks. Jantunen 2009a). Kuten Jantunen (mts. 107) toteaa, näitä kahta ei kuitenkaan pitäisi nähdä keskenään vastakkaisina. Tämäkin korpustutkimus, kuten itse asiassa kaikki kielestä tehdyt tutkimukset, perustuvat nimittäin osin intuitioon. Vaikka todellisen kielenkäytön tutkiminen korpusmenetelmin onkin tärkeässä osassa tässä tutkimuksessa, intuition rooli tutkimuksessa ylipäätään on kuitenkin suuri eikä tätäkään tutkimusta tehdä ilman tutkijan intuitiivisia oletuksia. Esimerkiksi tutkimusaiheen ja tutkimusaineiston valinta pohjautuu aina tutkijan intuitiivisiin näkemyksiin kielestä (Stubbs 2002: 9; Jantunen 2009a: 107), niin tässäkin tutkimuksessa. Lisäksi kuten muun muassa Jantunen (mp.) toteaa, myös aineiston käsittely edellyttää tutkijan kielellistä intuitiota. Intuition merkitys on suuri tässä työssä etenkin semanttisen preferenssin analyysissa: olen luonut sanojen semanttisen luokittelun pitkälti siihen perustuen, mitkä merkitykset intuitioni pohjalta kuuluvat yhteen, ja kunkin yksittäisen sanan sijoittaminen tiettyyn merkitysryhmään niin ikään perustuu intuitiiviseen ymmärrykseeni sanojen merkityksistä. Myös Jantunen (2004: 221) havaitsee omassa tutkimuksessaan, että semanttisen preferenssin sekä etenkin semanttisen prosodian tutkimisessa tutkija on pitkälti oman intuitionsa varassa. Tämä voidaan kuitenkin nähdä pikemminkin semanttisen preferenssin tutkimukseen sekä laajemminkin kielentutkimukseen lähtökohtaisesti kuuluvana piirteenä, ei heikkoutena, josta pitäisi pyrkiä eroon. Vanhatalokin esittää (2003: 362), että kaikkeen sanan merkitystä koskevaan tietoon, myös intuitioon pohjautuvaan, tulisi suhtautua ensisijaisesti kiinnos-

tuneesti. Menetelmiä ei siis tulisi nähdä toisilleen vastakkaisina, vaan myös aineistoon vahvasti nojautuvassa korpustutkimuksessa voidaan hyödyntää tutkijan intuitiivista näkemystä kielestä.

4 KOLLOKAATIT

4.1 Tilastollisesti merkitsevimmät kollokaatit

4.1.1 T-testi

Liitteen 1 taulukossa on lueteltu molempien modaalisanojen kollokaatit, jotka ovat t-testin mukaan tilastollisesti merkitseviä. Tällaisia kollokaatteja on EHKÄ-sanalla jopa 69 ja MAHDOLLISESTI-sanalla 26. T-testin antamista kollokaateista kymmenen merkitsevintä on lueteltu taulukossa 1.

Taulukko 1. Modaalisanojen kymmenen merkitsevintä kollokaattia t-testissä.

EHKÄ			MAHDOLLISESTI		
kollokaatti	t-testin arvo	frekvenssi (vas./oik. puoleiset kollokaatit)	kollokaatti	t-testin arvo	frekvenssi (vas./oik. puoleiset kollokaatit)
mutta	9,18	242 (113/129)	jopa	5,57	37 (3/4)
se	8,63	424 (103/321)	jo	5,31	53 (4/49)
ei	8,07	434 (321/113)	tai	4,64	40 (19/21)
hieman	7,18	65 (4/61)	ensi	4,62	32 (5/27)
jokin	7,09	85 (16/69)	tulla	3,66	43 (20/23)
vähän	7,05	77 (8/69)	liittyä	3,56	17 (2/15)
liian	6,58	61 (1/60)	henkilö	3,14	13 (2/11)
paljon	6,45	120 (21/99)	voida	3,04	38 (21/17)
siksi	5,94	47 (9/38)	aiheuttaa	2,93	12 (1/11)
vielä	5,90	110 (18/92)	vielä	2,92	25 (6/19)

Kun vertaillaan taulukossa näkyviä merkitsevyysarvoja, havaitaan, että EHKÄ kollokoi voimakkaammin tilastollisesti merkitsevimpien kollokaattiansa kanssa kuin MAHDOLLISESTI omien vastaa-

vien kollokaattinsa kanssa. Sanan EHKÄ merkitsevin kollokaatti MUTTA saa merkitsevyysarvokseen jopa 9,18, kun taas MAHDOLLIESTI-sanana vastaavan kollokaatin JOPA merkitsevyysarvo on 5,57. Koska tilastollisen merkitsevyyden raja-arvo t-testissä on siis 2,00, luvut ilmentävät, että sekä EHKÄ ja MUTTA että MAHDOLLIESTI ja JOPA etsiytyvät voimakkaasti toistensa yhteyteen. EHKÄ-sanana kollokaatti ensimmäisellä sijalla olevan kollokaattinsa kanssa on kuitenkin siis selvästi voimakkaampaa kuin synonyymillaan. Itse asiassa sanan EHKÄ kaikki kymmenen merkitsevintä kollokaattia saavat suuremman merkitsevyysarvon kuin sanan MAHDOLLIESTI tilastollisesti kaikkein merkitsevin kollokaatti. Näin ollen voidaan sanoa, että EHKÄ kollokoiki erittäin voimakkaasti listassa olevien sanojen kanssa MAHDOLLIESTI-sanaan verrattuna.

Taulukon kolmansista sarakkeista on tarkasteltavissa, kuinka monta kertaa kollokaatti esiintyy aineistossa modaalisanan vasemmalla ja kuinka monta kertaa sen oikealla puolella. Useimmat taulukon kollokaateista esiintyvät huomattavasti useammin modaalisanan oikeanpuoleisina kuin vasemmanpuoleisina kollokaatteina. Poikkeuksen tekevät EHKÄ-sanana kollokaateista kuitenkin EI, joka esiintyy useammin vasemmanpuoleisena kollokaattina, sekä MUTTA, joka kollokoiki lähes yhtä usein noodin oikealla ja vasemmalla puolella. MAHDOLLIESTI-sanana kollokaateista TAI, TULLA ja VOIDA esiintyvät lähes yhtä usein vasemman- ja oikeanpuoleisena kollokaattina.

Modaalisanoilla on t-testin kymmenen tilastollisesti merkitsevimmän kollokaatin joukossa vain yksi yhteinen kollokaatti, VIELÄ. Se kollokoiki voimakkaammin EHKÄ-sanana (merkitsevyysarvo 5,90) kuin MAHDOLLIESTI-sanana (merkitsevyysarvo 2,92) kanssa. EHKÄ-sanana tilastollisesti merkitsevimmät kollokaatit t-testissä ovat adverbeja ja partikkeleita (MUTTA, HIEMAN, VÄHÄN, LIIAN, PALJON, SIKSI, VIELÄ), pronomineja (SE, JOKIN) sekä kieltosana EI, joka tavallisimmin on luettavissa verbiksi (mutta voi olla joskus tulkittavissa myös partikkeliksi; kieltoisanan luokittelusta tarkemmin ks. alaluku 4.2.2). Näitä kaikkia voidaan sanaluokkiensa vuoksi pitää kieliopillisina sanoina. MAHDOLLIESTI-sanalla puolestaan on merkitsevimpinä kollokaatteina t-testissä niin ikään adverbeja ja partikkeleita (JOPA, JO, TAI, VIELÄ), ja nekin kaikki ovat tulkittavissa kieliopillisiksi sanoiksi. Lisäksi sillä on merkitsevimpien kollokaattien joukossa, toisin kuin synonyymillaan, adjektiivi (taipumaton adjektiivi ENSI), verbejä (TULLA, LIITTYÄ, VOIDA, AIHEUTTAA) ja substantiivi (HENKILÖ). Näistä etenkin verbit ja substantiivi ovat selkeästi sisältösanoja ja viittaavat kielenulkoiseen tarkoitteeseen. Kymmenen merkitsevimmän kollokaatin listasta voidaan siis havaita, että t-testin mukaan sana MAHDOLLIESTI näyttää kollokoivan taajemmin sisältösanojen kanssa kuin EHKÄ.

EHKÄ-sanana merkitsevimpien kollokaattien joukossa on neljä määrää ilmaisevaa sanaa, kvanttoriadverbit HIEMAN, VÄHÄN ja PALJON sekä intensiteettimäärite LIIAN (määrää ilmaisevien sanojen luokittelusta ks. VISK määritelmät sv. intensiteettimäärite, kvanttoriadverbi). Myös indefiniittisellä kvanttoripronominikollokaatilla JOKIN on samankaltainen merkityspiire. Kvanttoripronomit ni-

mittäin ilmaisevat, että tarkoitettua entiteettiä tai joukkoa ei ole identifioitu, vaan se on yksi monista mahdollisista; toisissa yhteyksissä korostuu määrä ('ainakin yksi') ja toisissa tarkoitteen identifioimattomuus (VISK § 746). Lisäksi VIELÄ esiintyy toisinaan vahvistamassa komparatiivista ilmausta (KS s.v. *vielä*). Näin ollen jopa viidellä EHKÄ-partikkelin kymmenestä merkitsevimmästä kollokaatista on ainakin joissain yhteyksissä määrään tai intensiteettiin viittaava merkitys. MAHDOLLISESTI-sanan kollokaatteina ei ole kvanttoriadverbeja tai intensiteettimääritteitä, mutta sen kollokaateista VIELÄ ja JOPA liittyvät 'määrän' merkitykseen (ks. luku 4.2.4) ja JO, ENSI ja joissakin konteksteissa myös VIELÄ puolestaan 'ajan' merkitykseen. T-testin merkitsevimmistä kollokaateista voidaan näin ollen tehdä alustava oletus siitä, että EHKÄ-sanalla ja samoin kenties sen synonyymilla MAHDOLLISESTI olisi semanttinen preferenssi 'määrä' ja MAHDOLLISESTI-sanalla puolestaan semanttinen preferenssi 'aika'. Semanttisiin preferensseihin syvennytään tarkemmin luvussa 5.

4.1.2 MI-testi

Liitteessä 2 on lueteltu MI-testin mukaiset tilastollisesti merkitsevät kollokaatit sanoille EHKÄ ja MAHDOLLISESTI. Niitä on EHKÄ-sanalla on 51 ja MAHDOLLISESTI-sanalla 28. Molempien modaalisanojen MI-testissä kymmenen merkitsevintä kollokaattia on lueteltu taulukossa 2.

MI-testin mukaan sana EHKÄ kollokoi kaikkein voimakkaimmin adverbien HANAKASTI kanssa ja MAHDOLLISESTI puolestaan verbin AIHEUTUA kanssa. Kumpikin kollokaatti esiintyy aineistossa vain alle kymmenen kertaa, HANAKASTI viisi kertaa ja AIHEUTUA kahdeksan kertaa. Näiden kollokaattien merkitsevyysarvot (4,87 ja 4,47) eivät poikkea merkittävästi toisistaan, joten molempia kollokaatiosuhteita voidaan pitää suurin piirtein yhtä voimakkaina. Voidaan sanoa, että muutenkin MI-testissä EHKÄ-sanan merkitsevimpien kollokaatioiden merkitsevyysarvot ovat suurin piirtein samalla tasolla kuin puolestaan MAHDOLLISESTI-sanan omien vastaavien kollokaatioiden merkitsevyysarvot. T-testissä huomattiin, että EHKÄ-sanan merkitsevimmät kollokaatiot ovat tilastollisesti huomattavasti merkitsevempiä kuin MAHDOLLISESTI-sanan vastaavat, mutta MI-testissä ei siis ole näin. Samoin kuin t-testin myös useimmat MI-testin mukaisen listan kollokaateista esiintyvät huomattavasti useammin modaalisanan oikealla kuin vasemmalla puolella. Kuitenkin partikkeli NO ja substantiivi HYLKY esiintyvät EHKÄ-sanaan nähden ainoastaan vasemmalla puolella. Lisäksi MAHDOLLISESTI-sanan kollokaateista MYÖHEMMIN ja VAI esiintyvät useammin modaalisanan vasemmalla kuin oikealla puolella.

Taulukko 2. Modaalisanojen kymmenen merkitsevintä kollokaattia MI-testissä.

EHKÄ			MAHDOLLISESTI		
kollokaatti	MI-testin arvo	frekvenssi (vas.-/oik.puoleiset kollokaatit)	kollokaatti	MI-testin arvo	frekvenssi (vas.-/oik.puoleiset kollokaatit)
hanakasti	4,87	5 (0/5)	aiheutua	4,47	8 (1/7)
aavistus	4,60	6 (0/6)	havainto	4,08	5 (1/4)
kaikkein	4,21	36 (0/36)	jopa	3,58	37 (3/34)
turhan	4,12	18 (0/18)	henkilö	2,96	13 (2/11)
sittenkin	4,11	17 (4/13)	liittyä	2,87	17 (2/15)
hiukan	3,40	22 (2/20)	2000	2,74	6 (1/5)
no	3,27	15 (15/0)	aiheuttaa	2,69	12 (1/11)
hylky	3,22	5 (5/0)	ensi	2,44	32 (5/27)
viisas	3,21	11 (3/8)	myöhemmin	2,40	7 (4/3)
hieman	3,19	65 (4/61)	vai	2,38	6 (5/1)

Sanoilla EHKÄ ja MAHDOLLISESTI ei ole ainuttakaan yhteistä sanaa MI-testin kymmenen merkitsevimmän kollokaatin luetteloissaan. MI-testin mukaan sanat siis eroavat toisistaan huomattavasti kärkipään kollokaattien osalta. Sanan EHKÄ tilastollisesti merkitsevimmät kollokaatit MI-testissä ovat suurimmaksi osaksi adverbeja ja partikkeleita (HANAKASTI, KAIKKEIN, TURHAN, SITTENKIN, HIUKAN, NO, HIEMAN), mutta joukossa on myös substantiiveja (AAVISTUS, HYLKY) ja adjektiivi (VIISAS). Sanan MAHDOLLISESTI tilastollisesti merkitsevimpiin kollokaatteihin puolestaan kuuluu niin ikään adverbeja ja partikkeleita (JOPA, MYÖHEMMIN, VAI) – joskin vähemmän kuin synonyymillaan –, substantiiveja (HAVAINTO, HENKILÖ) ja adjektiivi (taipumaton adjektiivi ENSI). Lisäksi joukkoon kuuluu myös verbejä (AIHEUTUA, LIITTYÄ, AIHEUTTAA) ja numeraali (2000), joita kumpiakaan ei ole EHKÄ-sanan vastaavassa listassa.

Modaalisanat siis eroavat toisistaan kärkipään kollokaattien osalta MI-testissä erityisesti siten, että EHKÄ-sanan kollokaatit painottuvat synonyymiaan enemmän adverbeihin ja partikkeleihin ja että synonyymisista modaalisanoista vain MAHDOLLISESTI-sanan kollokaateissa esiintyy verbejä ja

numeraali. Kiinnostava havainto on myös se, että kaikki MAHDOLLISESTI-sanan listassa olevat verbikollokaatit ovat merkitykseltään suhteellisen lähellä toisiaan: verbit AIHEUTUA, LIITTYÄ ja AIHEUTTAA kuvaavat kaikki jollakin tavalla kausaalisia suhteita tai muita abstraktien asioiden välisiä suhteita.

4.1.3 Eri testien antama kuva modaalisanojen kollokoinnista

Edellä on lueteltu taulukoissa t-testin ja MI-testin mukaiset kymmenen merkitsevintä kollokaattia. Taulukossa 3 puolestaan on koottu yhteen kummastakin testistä jo edellä mainitut EHKÄ-sanan kymmenen merkitsevintä kollokaattia havainnollistamaan sitä, millaisia sanoja kumpikin testi nostaa kymmenen tilastollisesti merkitsevimmän kollokaatin joukkoon. Taulukossa 4 on sama vertailu MAHDOLLISESTI-sanan osalta.

Taulukko 3. EHKÄ-sanan kymmenen merkitsevintä kollokaattia t- ja MI-testeissä.

T-TESTI				MI-TESTI			
kollokaatti	t-testin arvo	kollokaatin frekvenssi (vas./oik.puol. kollokaatit)	sanan frekvenssi	kollokaatti	MI-testin arvo	kollokaatin frekvenssi (vas./oik.puol. kollokaatit)	sanan frekvenssi
mutta	9,18	242 (113/129)	38205	hanakasti	4,87	5 (0/5)	66
se	8,63	424 (103/321)	94880	aavistus	4,60	6 (0/6)	95
ei	8,07	434 (321/113)	102413	kaikkein	4,21	36 (0/36)	747
hieman	7,18	65 (4/61)	2752	turhan	4,12	18 (0/18)	400
jokin	7,09	85 (16/69)	7577	sittenkin	4,11	17 (4/13)	379
vähän	7,05	77 (8/69)	5844	hiukan	3,40	22 (2/20)	801
liian	6,58	61 (1/60)	3713	no	3,27	15 (15/0)	601
paljon	6,45	120 (21/99)	18994	hylky	3,22	5 (5/0)	207
siksi	5,94	47 (9/38)	2424	viisas	3,21	11 (3/8)	458
vielä	5,90	110 (18/92)	18525	hieman	3,19	65 (4/61)	2752

Kuten luvussa 3.3.1 huomautettiin, t-testin on usein todettu korostavan frekventtejä, kieliopillisiä sanoja ja MI-testin puolestaan melko harvoin esiintyviä sisältösanoja. EHKÄ-sanalla osalta voidaan havaita, että t-testin mukaiset kollokaatit ovat keskimäärin frekventimpiä sanoja kuin MI-testin esiin nostamat kollokaatit. Esimerkiksi t-testin mukainen merkitsevin kollokaatti MUTTA esiintyy tutkielman koko aineistossa kaiken kaikkiaan peräti 38205 kertaa, mutta MI-testin vastaava kollokaatti HANAKASTI ainoastaan 66 kertaa; nämä luvut kuvaavat siis sanojen esiintymiä ylipäättään, eivät niiden kollokointia modaalisanan kanssa. Samoin ero tulee esille siinä, kuinka monta kertaa sanat esiintyvät modaalisanan kotekstissa: MUTTA esiintyy modaalisanan kollokaattina jopa 242 kertaa ja HANAKASTI puolestaan ainoastaan viisi kertaa. Edelleen testien välinen ero ilmenee esimerkiksi siinä, että t-testin mukaisista kollokaateista suurin absoluuttinen frekvenssi aineistossa on sanalla EI, jolla esiintymiä on jopa 102 413. Sen sijaan MI-testin mukaisista kymmenestä merkitsevimmästä kollokaatista useimmin koko aineistossa esiintyy HIEMAN, joka esiintyy EI-sanaan verrattuna huomattavasti harvemmin, 2752 kertaa. Testit siis poikkeavat toisistaan huomattavasti siltä osin, kuinka frekventtejä sanoja ne nostavat tilastollisesti merkitseviksi.

T-testi korostaa hypoteesin mukaan siis kieliopillisiä sanoja ja MI-testi sisältösanoja. EHKÄ-sanalla kaikki kymmenen tilastollisesti merkitsevintä kollokaattia ovatkin luettavissa kieliopillisiksi sanoiksi. MI-testin mukaisista kollokaateista puolestaan seitsemän on kieliopillisiä sanoja ja kolme sanaa (AAVISTUS, HYLKY, VIISAS) sisältösanoja. Sisältösanoista etenkin HYLKY on erittäin selkeästi sisältösana. Sillä on konkreettinen tarkoite, johon se viittaa, ja kiinnostavaa on, että sanalla HYLKY on myös erittäin spesifinen merkitys. HYLKY on MI-testin mukaan EHKÄ-sanalla kahdeksanneksi merkitsevin kollokaatti, ja kun verrataan sanan merkitystä t-testin vastaavalla sijalla olevaan kollokaattiin, PALJON, voidaan havaita selkeä ero. T-testin kahdeksanneksi merkitsevin kollokaatti esiintyy aineistossa kaiken kaikkiaan jopa 18994 kertaa ja MI-testin vastaava kollokaatti puolestaan vain 207 kertaa. Ero ilmenee myös sanojen frekvensseissä kollokaatteina: HYLKY esiintyy kollokaattina ainoastaan viisi kertaa ja PALJON puolestaan jopa 120 kertaa.

Kun tarkastellaan puolestaan taulukkoa 4, MAHDOLLISESTI-sanalla osalta havaitaan sama asia kuin EHKÄ-sanalla osalta havaittiin: t-testi nostaa merkitseviksi enemmän suurifrekvenssisiä kollokaatteja kuin MI-testi. Esimerkiksi t-testin kärkikollokaatti JOPA esiintyy aineistossa yhteensä 5513 kertaa ja MI-testin kärkikollokaatti AIHEUTUA puolestaan selvästi harvemmin, 644 kertaa. MAHDOLLISESTI-sanalla erot eivät kuitenkaan keskimäärin ole yhtä suuria kuin EHKÄ-sanalla. MAHDOLLISESTI-sanalla kollokaateissa ei esimerkiksi ole yhtäkään sanaa, jonka absoluuttinen frekvenssi aineistossa olisi alle sata: harvimminkin esiintyvä kollokaatti HAVAINTO esiintyy sekä 526 kertaa aineistossa.

Taulukko 4. MAHDOLLISESTI-sanan kymmenen merkitsevintä kollokaattia t- ja MI-testeissä.

T-TESTI				MI-TESTI			
kollokaatti	t-testin arvo	kollokaatin frekvenssi (vas.-/oik.puol. kollokaatit)	sanan frekvenssi	kollokaatti	MI-testin arvo	kollokaatin frekvenssi (vas.-/oik.puol. kollokaatit)	sanan frekvenssi
jopa	5,57	37 (3/4)	5513	aiheutua	4,47	8 (1/7)	644
jo	5,31	53 (4/49)	25630	havainto	4,08	5 (1/4)	526
tai	4,64	40 (19/21)	18969	jopa	3,58	37 (3/34)	5513
ensi	4,62	32 (5/27)	10493	henkilö	2,96	13 (2/11)	2970
tulla	3,66	43 (20/23)	33858	liittyä	2,87	17 (2/15)	4157
liittyä	3,56	17 (2/15)	4157	2000	2,74	6 (1/5)	1601
henkilö	3,14	13 (2/11)	2970	aiheuttaa	2,69	12 (1/11)	3325
voida	3,04	38 (21/17)	34309	ensi	2,44	32 (5/27)	10493
aiheuttaa	2,93	12 (1/11)	3325	myöhemmin	2,40	7 (4/3)	2369
vielä	2,92	25 (6/19)	18525	vai	2,38	6 (5/1)	2059

MAHDOLLISESTI-sanalla t-testin mukaisista kymmenestä merkitsevimmästä kollokaatista viisi, ENSI, TULLA, LIITTYÄ, HENKILÖ ja AIHEUTTAA, ovat luettavissa sisältösanoiksi ja muut viisi kieliopillisiksi sanoiksi. MI-testin mukaisista kärkikollokaateista puolestaan seitsemän sanaa, AIHEUTUA, HAVAINTO, HENKILÖ, LIITTYÄ, 2000, AIHEUTTAA ja ENSI, ovat sisältösanoja ja muut kolme kieliopillisia sanoja. MI-testi siis osoittaa myös MAHDOLLISESTI-sanalla osalta hienoista taipumusta painottaa sisältösanoja. Testien välinen ero ilmenee kuitenkin selkeämmin EHKÄ-sanalla, jonka t-testin kärkikollokaateista kaikki kymmenen ovat kieliopillisia sanoja.

Testien antamia tuloksia voidaan vertailla toisiinsa myös siitä näkökulmasta, ovatko merkitsevimmät kollokaatit samoja sanoja eri testien mukaan. T- ja MI-testi antavat MAHDOLLISESTI-sanalle viisi keskenään samaa kollokaattia (JOPA, ENSI, LIITTYÄ, HENKILÖ, AIHEUTTAA). Tämä kertoo ensinnäkin siitä, että nämä kollokaatit ovat tärkeitä MAHDOLLISESTI-sanalla kollokaatteja. Toi-

sekseen huomiota voidaan verrata siihen, että t- ja MI-testi antavat puolestaan EHKÄ-sanalle kymmenen merkitsevimmän kollokaatin joukossa vain yhden keskenään saman kollokaatin, HIEMAN. Muut modaalipartikkelin kollokaatit ovat eri sanoja eri testien kesken. Näin ollen MAHDOLLIESTI-sanalla on kymmenen kärkikollokaatin joukossa EHKÄ-sanaa enemmän sellaisia kollokaatteja, jotka nousevat merkitseviksi molemmissa testeissä. EHKÄ-sanan kymmenen merkitsevimmän kollokaatin kirjo on siten laajempi kuin synonyymillaan. Lisäksi kollokaatioiden variaatioon liittyvästä havainnosta voidaan päätellä, että testit siis voivat nostaa merkitseviksi monia samoja kollokaatteja, mutta voi myös olla, että eivät nosta. Tämä riippunee itse tutkittavien sanojen ominaisuuksista, ei niinkään tilastollisten testien ominaisuuksista.

Kuten jo luvuissa 4.1.1 ja 4.1.2 todettiin, t-testin mukaan kymmenen merkitsevintä kollokaatiota ovat EHKÄ-sanalla huomattavasti tilastollisesti merkitsevempiä kuin MAHDOLLIESTI-sanalla, mutta MI-testin mukaan kymmenen merkitsevintä kollokaatiota taas ovat suurin piirtein yhtä voimakkaita EHKÄ- ja MAHDOLLIESTI-sanoilla. Kun vertaillaan kummankin testin tilastollisesti kaikkein merkitsevimpiä kollokaatiosuhteita, havaitaan siis seuraavaa: t-testin kärkikollokaatiosuhde EHKÄ <MUTTA> on huomattavasti voimakkaampi kuin kärkikollokaatio MAHDOLLIESTI <JOPA>, mutta MI-testissä kärkikollokaatiot EHKÄ <HANAKASTI> sekä MAHDOLLIESTI <AIHEUTUA> ovat tilastollisesti suunnilleen yhtä merkitseviä. Kahdesta tilastollisesta testistä t-testi siis ilmentää suurempaa eroa siinä, kuinka voimakkaita modaalisanon kollokaatiosuhteet ovat verrattuna toisiinsa. MI-testissä modaalisanon välisten kollokaatioiden merkitsevyyserot puolestaan ovat pienemmät.

4.2 Eri sanaluokkien kollokaatit

Edellä tarkasteltiin modaalisanon kymmentä tilastollisesti merkitsevintä kollokaattia, mutta on huomattava, että molemmilla modaalisanoilla kollokaatteja on kuitenkin huomattavasti enemmän. Niitä on EHKÄ-sanalla t-testin mukaan yhteensä siis 69 ja MI-testin mukaan puolestaan 51. Vastavasti MAHDOLLIESTI-sanalla on t-testin mukaan 26 tilastollisesti merkitsevää kollokaattia ja MI-testin mukaan 28. Molempien sanojen t-testin mukaiset kollokaatit on lueteltu liitteessä 1 ja MI-testin mukaiset liitteessä 2. Seuraavaksi tarkastelen sanojen kaikkien kollokaattien kirjoa. Tarkoituksena on tehdä katsaus siihen, millaisia erilaisia kollokaatteja esiintyy toisaalta sanan EHKÄ, toisaalta sanan MAHDOLLIESTI kotekstissa tarkastelualueella 3:4. Etenen analyysissä sanaluokittain. Molemmilla modaalisanoilla on tilastollisesti merkitseviä kollokaatteja kaikissa eri sanaluokissa: kollokaatteina on niin substantiiveja, adjektiiveja, numeraaleja, pronomineja, verbejä, adverbeja, adpositioita kuin partikkeleitakin.

Etenen kollokaattianalyysissä siten, että ensin esittelen nominikollokaatit: ensimmäiseksi substantiivit, toiseksi adjektiivit, sen jälkeen numeraalit ja lopuksi pronominit. Nominien jälkeen analysoin verbikollokaatteja, sitten adverbikollokaatteja ja viimeiseksi sellaisia kollokaatteja, jotka ovat adpositioita ja partikkeleita. Kunkin ryhmän kollokaatit esitän ensin taulukossa, ja sen jälkeen analysoin niitä yksityiskohtaisemmin. Osa kollokaateista on tilastollisesti merkitseviä vain toisen testin mukaan ja osa molempien testien mukaan. Kollokaatit on järjestetty taulukkoon t-testin merkitsevyysarvon mukaiseen suuruusjärjestykseen, ja niiden jälkeen ne kollokaatit, jotka ovat merkitseviä vain MI-testin perusteella, on esitetty MI-testin mukaisessa suuruusjärjestyksessä. Merkitsevyysrajan ylittävät arvot (t-testissä 2,00 ja MI-testissä 1,58) on kussakin taulukossa lihavoitu. Lisäksi ne kollokaatit, jotka ovat yhteisiä kummallekin sanalle, on alleviivattu. Kollokaatioanalyysin tarkoitus on osoittaa, mitkä nimenomaiset sanat esiintyvät tyypillisimmin modaalisanojen kontekstissa. Tarkastelen kollokaatteja kuitenkin myös siitä näkökulmasta, millaisia semanttisia luokkia ne muodostavat, ja täten luon hypoteeseja varsinaista semanttisen preferenssin analyysia varten.

4.2.1 Nominikollokaatit

4.2.1.1 Substantiivikollokaatit

Sanalla EHKÄ on kahdeksan tilastollisesti merkitsevää substantiivikollokaattia ja sanalla MAHDOLLI-SESTI puolestaan viisi, kuten taulukosta 5 nähdään. Modaalisanojen ainoa yhteinen substantiivikol- lokaatti on TULEVAISUUS, joka on merkitsevä EHKÄ-sanalla t-testin mukaan ja MAHDOLLI-SESTI-sanalla puolestaan MI-testin mukaan. Sen kollokointia modaalisanojen kanssa voidaan ainakin jos- sain määrin selittää sillä, että tulevaisuuteen liittyy aina epävarmuuden merkitys. Kukaan ei tiedä varmasti, mitä tulevaisuus pitää sisällään, minkä eräänlaisena seurauksena voidaan pitää sitä, että kielessä sana TULEVAISUUS kollokoi episteemisten modaalisanojen kanssa. Virkkeissä (3) ja (4) on esimerkkejä siitä, miten epävarmuutta ilmentävä modaalisana sekä sana TULEVAISUUS kollokoivat.

- 3) Toisaalta lieksalaiset ovat tottuneet ajamaan kaarisillalla, ja se *ehkä* onnistuisi myös **tule-
vaisuudessa** nykymitoissaan.
- 4) 19-vuotias Pyykkönen on toiminut kansalaisjärjestöissä ennenkin ja haluaa *mahdollisesti* **tulevaisuudessa** nuoriso- tai kulttuurialalle.

Taulukko 5. Modaalisanojen EHKÄ ja MAHDOLLISESTI substantiivikollokaatit aineistossa.

EHKÄ			MAHDOLLISESTI		
kollokaatti	t-testin arvo	MI-testin arvo	kollokaatti	t-testin arvo	MI-testin arvo
syy	4,35	1,60	henkilö	3,14	2,96
<u>tulevaisuus</u>	2,91	1,34	syksy	2,71	2,19
aavistus	2,35	4,60	havainto	2,10	4,08
hylky	2,00	3,22	tulo	1,93	1,88
kuulija	1,69	2,02	<u>tulevaisuus</u>	1,91	1,84
hahmo	1,68	2,00			
tahto	1,58	1,78			
siirto	1,57	1,74			

Taulukosta 5 nähdään, että EHKÄ-sana kollokoi seitsemän sellaisen sanan kanssa, joiden kanssa MAHDOLLISESTI ei kollokoi: SYY, AAVISTUS, HYLKY, KUULIJA, HAHMO, TAHTO ja SIIRTO. Näistä kaikkein merkitsevimpinä voidaan pitää kahta ensin mainittua sanaa, SYY ja AAVISTUS, sillä ne ovat merkitseviä molempien testien mukaan. Niiden molempien merkitys on luokiteltavissa merkitykseltään abstraktiksi, kuten myös virkkeistä (5) ja (6) nähdään. Sanan AAVISTUS kollokointi EHKÄ-modaalisanan kanssa näyttää liittyvän samaan piirteeseen, joka tuli osin esille jo kymmenen kärkikollokaatin luettelosta: EHKÄ kollokoi voimakkaasti määrää ilmaisevien sanojen kanssa. Lekseemin AAVISTUS ensisijainen merkitys on ”hämärä (ennakko)tunne, -tuntemus”, mutta aineistossa se esiintyy useimmin, viidessä tapauksessa kuudesta, genetiivimuodossa ja merkityksessä ”hiven, hitunen” (merkitykset KS s.v. *aavistus*). Nominin genetiivimuodot toimivat usein intensiteettimääritteinä, kuten myös monet adverbis ja partikkelit (VISK § 664). AAVISTUS esiintyykin siis juuri tässä funktiossa virkkeessä (6) ja muistuttaa siten adverbia. Substantiivikollokaattilistan viimeisinä olevat TAHTO ja SIIRTO ovat nekin merkitykseltään abstrakteja, kuten huomataan myös esimerkeistä (7) ja (8).

5) *Ehkä* tärkein **syy** tähän on se, että kaksi maailman suurinta lelutehdasta on perustanut Suomeen omat myyntiorganisaationsa.

6) Leppoisa 15 asteen lämpötila oli *ehkä aavistuksen* ihanteellista lämpimämpi, mutta juoksemista helpotti tuulettomuus.

7) Jakaja on *ehkä* vasten **tahtoaan** joutunut lähtemään myöhästyneelle jakokierrokselle toiseen piiriin, kun sijaista on ollut vaikea saada.

8) Hän ei **siirtoa** suunnitellessaan *ehkä* osaa ottaa huomioon, että vastustaja saattaa yllättää kiskaisemalla peliin laskuvarjosissin varanappuloista, Lähdesmäki sanoo.

Abstrakteihin ilmiöihin viittaavien kollokaattien lisäksi EHKÄ-sanalla on myös muunlaisia kollokaatteja. KUULIJA viittaa inhimilliseen tekijään, kuten virkkeessä (9), ja HAHMO esimerkiksi taiteessa esiintyviin henkilöihin, kuten virkkeessä (10). Substantiivi HYLKY puolestaan viittaa konkreettiseen olioon. HYLKY-sanalla on varsin spesifinen merkitys, ja sen tilastollisesti merkitsevää kollokointia EHKÄ-sanan kanssa voidaan selittää sillä, että aineistona olevissa lehdissä uutisoidaan useita kertoja samasta Tampereella sattuneesta vedenalaisesta hylkylöydöstä. Täten se, mikä milloinkin on uutisissa ajankohtaisista, voi vaikuttaa siihen, mitkä kollokaatit nousevat sanomalehtiaineistossa tilastollisesti merkitseviksi. Virke (11) esittää yhden otsikkona toimivan kontekstin, jossa EHKÄ ja HYLKY kollokoivat.

9) *Ehkä* laulun haikeus vetoaa **kuulijoihin**, Tapio Selin sanoo.

10) – – ja vaikka maalauksissa **hahmoja** on *ehkä* Tikan aikaisempaan tuotantoon verrattuna niukemmin, kokonaan abstraktiksi hän ei ole vielä heittäytynyt.

11) Mustanlahden **hylyt** *ehkä* proomuja

MAHDOLLISESTI-sanalla on neljä kollokaattia, joita ei ole sen synonyymilla: HENKILÖ, SYKSY, HAVAINTO ja TULO. Substantiivien SYKSY ja TULO kollokointia modaalisanan kanssa voitaneen ainakin osin selittää samalla seikalla kuin TULEVAISUUS-sanan kollokointia molempien modaalisanojen kanssa: nämäkin kollokaatit viittaavat aikaan. Etenkin tuleva aika sisältää aina epävarmoja tekijöitä. Esimerkiksi voidaan ajatella, että virkkeessä (12) MAHDOLLISESTI- ja SYKSY-sanojen kollokointi liittyy siihen, että tulevaan aikaan, ensi vuoden syksyyn, liittyvät tapahtumat eivät ole vielä aivan varmoja, vaan tulevaan on viitattava episteemisellä modaalisanalla. TULO puolestaan liittyy niin ikään 'tulevaisuuden' merkitykseen. Sitä käytetään lähes kaikissa tapauksissa rakenteessa ”olla tulossa”, joka on OLLA-verbin ja verbikantaisen substantiivijohdoksen liitto. Tällaiset rakenteet esittävät tapahtuman käynnissä olevana tai olleena, mutta usein myös ennakoivana eli tulevaisuuteen liittyvänä (VISK § 1519). Käynnissä olevaa tai tulevaisuuteen liittyvää sanan käyttöä kuvaa virke (13). Sen sijaan lekseemi ei esiinny kertaakaan esimerkiksi merkityksessä ”ansiona, palkkana, voittona tms. saatava rahamäärä” (merkitys KS s.v. *tulo*). Kaiken kaikkiaan sana MAHDOLLISESTI yhdis-

tyykin aineiston perusteella hieman herkemmin aikaan viittaaviin substantiiveihin kuin synonyyminsa EHKÄ.

Tilastollisesti merkitseviä MAHDOLLISESTI-sanan substantiivikollokaatteja ovat myös HENKILÖ ja HAVAINTO, jotka nousevatkin merkitseviksi molempien testien mukaan. Kuten virkkeistä (14) ja (15) havaitaan, nämä kollokaatiot esiintyvät usein sellaisissa lauseissa, joissa on puhe siitä, että poliisi on selvittämässä jotakin rikosta. Näin ollen voidaan tehdä alustava hypoteesi, että sanan MAHDOLLISESTI yksi semanttinen preferenssi olisi 'rikokset'. Tähän ja muihin mahdollisiin semanttisiin preferensseihin syvennyttään tarkemmin luvussa 4.2.

12) Rakentaminen alkaa *mahdollisesti* jo ensi vuoden **syksyllä**.

13) Kiteen kaupunki on *mahdollisesti* **tulossa** mukaan esisuunnitteluvaiheen rahoituksen hakijana.

14) Tuhotöitä tutkiva keskusrikospoliisi kertoo, että palot ovat tahallaan sytytettyjä ja *mahdollisesti* asialla olivat samat **henkilöt**.

15) Tampereen poliisi on kiinnostunut kaikista ryöstöön *mahdollisesti* liittyvistä **havainnoista**.

4.2.1.2 Adjektiivikollokaatit

Modaalisanat poikkeavat toisistaan huomattavasti, kun verrataan niiden saamien tilastollisesti merkitsevien adjektiivikollokaattien määrää: EHKÄ-sanalla näitä kollokaatteja on jopa 11 ja MAHDOLLISESTI-sanalla vain kaksi. Täten paljastuu samalla modaalisanoja erottava kolligaatio: EHKÄ <adjektiiv> on useammin esiintyvä kolligaatio kuin MAHDOLLISESTI <adjektiiv>. Taulukosta 6 havaitaan, että modaalisanojen ainoa yhteinen tilastollisesti merkitsevä adjektiivikollokaatti on sana ENSI. Se voi viitata joko aikaan, paikkaan tai arvo-, tärkeys- tai muuhun järjestykseen (KS s.v. *ensi*), mutta aineistossa se esiintyy molemmilla modaalisanoilla yksinomaan aikaan viittaamassa. Aikaan viittavan ENSI-sanan ja modaalisanan kollokointia ilmentävät esimerkit (16) ja (17). Mielenkiintoista on myös, että ENSI viittaa nimenomaan tulevaan aikaan, aivan kuten myös monet sanojen substantiivikollokaatit.

16) *Ehkä* se menee jo **ensi** viikolla Ruotsissa pidettävissä keihäskarnevaaleissa, Mari kertoi.

17) Unkarista tulee Naton jäsen *mahdollisesti* **ensi** vuoden alussa.

Taulukko 6. Modaalisanojen EHKÄ ja MAHDOLLISESTI adjektiivikollokaatit aineistossa.

EHKÄ			MAHDOLLISESTI		
kollokaatti	t-testin arvo	MI-testin arvo	kollokaatti	t-testin arvo	MI-testin arvo
hyvä	4,78	0,87	<u>ensi</u>	4,62	2,44
turhan	4,00	4,12	uusi	2,01	0,76
tärkeä	3,00	0,89			
viisas	2,96	3,21			
suuri	2,77	0,57			
kova	2,55	1,06			
paha	2,54	1,32			
<u>ensi</u>	2,15	0,59			
koko	2,12	0,59			
oleellinen	1,86	2,57			
mielenkiintoinen	1,63	1,91			

Modaalisanan EHKÄ adjektiivikollokaattien suuri määrä MAHDOLLISESTI-sanaan nähden paljastaa, että sana osoittaa synonyymiaan voimakkaampaa kollokointia laatua ilmaiseviin sanoihin. Sen adjektiivikollokaatit edustavat useita eri Dixonin (2004: 3–5) adjektiiviryhmiä: kollokaatteina on niin arvoon (HYVÄ, *turhan*, TÄRKEÄ, PAHA, MIELENKIINTOINEN), inhimilliseen taipumukseen (VIISAS), ulottuvuuteen (SUURI), kvantiteettiin (KOKO) kuin fyysiseen ominaisuuteen (KOVA) liittyviä adjektiiveja. Näistä painokkaimpina voidaan pitää arvoon viittaavia sanoja, sillä sellaiseksi luokiteltavia on kollokaatteina siis viisi, ja niistä HYVÄ on t-testin arvollaan 4,78 suhteellisen selvästi muita kollokaatteja merkitsevempi t-testissä. Adjektiivikollokaattien suuresta määrästä voidaan näin päätellä, että EHKÄ-sanaa käytetään usein arvioitaessa asioiden laatua. Kun vastaavasti MAHDOLLISESTI-sanalla ei ole ainoatakaan varsinaisesti arvoa merkitsevää sanaa – vaikka toki UUSI voi joskus saada myös sellaisen merkityksen –, niin hyvyttä, huonoutta tai muuta arvoa ilmaisevat adjektiivit painottuvat nimenomaan EHKÄ-sanana kollokaatteina.

EHKÄ-sanana adjektiivikollokaatit esiintyvät aineistossa usein superlatiivimuodossa, kuten virkkeet (18)–(22) ilmentävät. Superlatiivimuotoisuutta voidaan kenties selittää sillä, että superlatiivi ilmaisee kaikkein äärimmäisimmän adjektiivin asteen, ja modaalisanalla tätä tietynlaista jyrkkyyttä voidaan ikään kuin hieman pehmentää. Toisaalta osa adjektiivien esiintymistä on myös

komparatiivimuodossa tai perusasteessa, kuten virkkeet (23)–(24) ilmentävät. Lisäksi yhtenä adjektiivikollokaattina on taipumaton adjektiivi KOKO, kuten virke (25) osoittaa. Aineistossa esiintyy kollokaattina myös sananmuoto *turhan*, joka korpuksessa on luokiteltu adverbiksi, mutta jonka olen kuitenkin luokitellut Ison suomen kieliopin mukaisesti (VISK § 550, 658) TURHA-adjektiivin geneetiivimuodoksi, joka toimii astemääritteenä. Esimerkiksi virkkeessä (26) *turhan* määrittää VAISUSTI-adverbia.

- 18) Ja joukkueella on *ehkä* maailman **paras** kaksikko kärjessä (Ivan Zamorano ja Marcelo Salas).
- 19) *Ehkä tärkeintä* on se, että oppilaat saadaan näin paremmin sitoutumaan oppimiseen.
- 20) *Ehkä suurin* vaara on sellaisella valtiolla, joka ei taloudellisista syistä pysy ydinasekilpailussa mukana – –
- 21) *Ehkä oleellisin* keksintö oli täyteen käyttöön päässyt turvevankkuri.
- 22) *Ehkä mielenkiintoisin* lopputulos onkin, että sarjakuva on liian omalaatuinen väline, jotta kirjallisuuden teoria riittäisi sen avaimiksi.
- 23) Mutta kun julkisen sektorin työpaikat vähenevät kuten viime vuonna 600 työpaikan vuosivauhdilla – jatkossa *ehkä* vielä **kovemmallakin** – pitää teollisustyöpaikkoja syntyäkin – –
- 24) Asiantuntijat arvelevat, ettei tästä talvesta *ehkä* tule erityisen **pahaa** myyrävuotta.
- 25) Punamultalukko oli suppatutkijoiden mukaan Suomen ja *ehkä* **koko** maailman syvin.
- 26) Hiihto sujui *ehkä turhan* vaisusti, varsinkin viimeinen puolikilometrinen tuolta mäen päältä oli vaikeaa, Kosunen kommentoi suoritustaan.

MAHDOLLISESTI-sanana kollokaateista ENSI on niin ikään EHKÄ-sanana kollokaatti, kuten edellä esitettiin, mutta UUSI sen sijaan kollokoi kahdesta modaalisanasta ainoastaan MAHDOLLISESTI-sanana kanssa. Merkillepantavaa on, että MAHDOLLISESTI-sanana molemmat adjektiivikollokaatit liittyvät merkitykseltään aikaan. Sana UUSI esiintyy modaalisen adverbina kollokaattina usein sellaisissa lauseissa, joissa on puhe suunnitteilla olevista yhteiskunnallisista uudistuksista; virkkeessä (27) on tästä esimerkki. Modaalinen adverbi ei sitä vastoin kollokoi lainkaan esimerkiksi arvoa merkitsevien adjektiivien kanssa, kuten synonyyminsa EHKÄ.

- 27) Ylitarkastaja Hiltunen tarkoittaa kunnassa *mahdollisesti* tehtäviä **uusia** määräyksiä.

4.2.1.3 Numeraalikollokaatit

Taulukon 7 mukaisesti EHKÄ-sana osoittaa tilastollisesti merkitsevää kollokointia neljään numeraaliin ja MAHDOLLISESTI-sana yhteen. EHKÄ kollokoi numeraalien PARIKYMMENTÄ, KYMMENEN, SATA ja 400 kanssa. Usein modaalisanan vaikutusala on juuri numeraali eli sanan ilmentämä epävarmuus kohdistuu numeraalin ilmentämään määrään, kuten virkkeissä (28)–(30). Tällaisissa virkkeissä määrää siis arvioidaan modaalisanan avulla, ja sillä osoitetaan, ettei ilmaistu määrä ole ehkä aivan tarkka, vaan vain arvio. Toisaalta joskus modaalisanan ja numeraalikollokaatin suhde on muunlainen ja modaalisanan ilmentämä epävarmuus kohdistuu virkkeessä laajemmalle kuin pelkkään numeraaliin. Tästä on esimerkkinä virke (31), jossa modaalisanalla ei arvioida numeraalin ilmaisemaa määrää, vaan sitä, onko pääkaupunkiseutu Joensuuta turvattomampi. Toisaalta kyse saattaa esimerkkivirkeessä olla myös eräänlaisesta sanottavan pehmentelystä eikä varsinaisesta epävarmuudesta.

Taulukko 7. Modaalisanojen EHKÄ ja MAHDOLLISESTI numeraalikollokaatit aineistossa.

EHKÄ			MAHDOLLISESTI		
kollokaatti	t-testin arvo	MI-testin arvo	kollokaatti	t-testin arvo	MI-testin arvo
kymmenen	2,49	0,97	2000	2,08	2,74
parikymmentä	2,43	2,10			
sata	2,35	1,16			
400	1,86	1,74			

- 28) – Pirkanmaalta löytyy *ehkä* **parikymmentä** maatilaa, jossa paimenkoirat ovat jokapäiväisessä työkäytössä, Maria tietää.
- 29) – Meitä oli silloin huipulla *ehkä* **kymmenen** naista, eikä taso vielä yltänyt kansainväliseen kärkeen.
- 30) Vepsäläisen mukaan kyse on *ehkä* **sadasta** uudesta sijoituspaikasta.
- 31) *Ehkä* pääkaupunkiseutu yli **400** jutullaan saattaa olla kuitenkin turvattomampi kuin Joensuu tilastoiduilla 20 tapauksellaan.

Sana MAHDOLLISESTI puolestaan osoittaa kollokointia numeraalin 2000 kanssa. Se esiintyy kaikissa tapauksissa viittaamassa vuoteen 2000, ei sitä vastaavaan lukumäärään, joten kyse on nimenomaan ajankohdasta; esimerkkinä tästä on virke (32). Tämä vahvistaa jo edellä esitettyä MAHDOLLISESTI-sanana ja aikaan liittyvän merkityksen kollokointia. Lisäksi koska vuosi 2000 oli aineiston sanomalehtien kirjoittamishetkellä vielä tulevaisuutta, kollokaatti osoittaa osaltaan, että modaalisana esiintyy teksteissä usein juuri tulevaisuudesta puhuttaessa.

32) Kirkon ikonostaasin rakentaminen, ikonien maalaaminen ja muu sisustaminen tehdään ensi vuoden aikana, ja kirkko vihitään käyttöön *mahdollisesti* vuonna **2000**.

4.2.1.4 Pronominikollokaatit

Taulukosta 8 havaitaan, että EHKÄ-sanalla on kahdeksan tilastollisesti merkitsevää pronomini-kollokaattia ja MAHDOLLISESTI-sanalla neljä. Näistä kummallekin modaalisanalle yhteisiä kollokaatteja on kaksi, indefiniittiset kvanttoripronomininit JOKIN ja JOKU. Erityisen painokas kollokaatio on t-testissä sanapari EHKÄ - JOKIN, joka saa arvokseen jopa 7,09. Indefiniittiset kvanttoripronomininit ilmaisevat, että ”puheena oleva entiteetti tai joukko on yksi monista mahdollisista, ei mikään tietty” (VISK määritelmät s.v. indefiniittinen kvanttoripronomini). Episteemisten modaalisanojen kollokointia näiden pronominiin kanssa ei voidakaan pitää yllättävänä: episteemisyys ja indefiniittisyyden ilmentämä useiden vaihtoehtojen olemassaolo liittyvät merkityksiltään yhteen, kuten virkkeistä (33)–(36) havaitaan. Esimerkiksi virkkeessä (33) ilmaistaan JOKIN-pronominilla, että järjestettävä tapahtuma ei ole mikään tietty, ja modaalisanalla EHKÄ, ettei järjestäminen ylipäättään ole täysin varmaa. Samoin esimerkiksi virkkeessä (36) pronomini JOKU ilmentää, ettei puheena olevaa henkilöä ole identifioitu, ja modaalisana MAHDOLLISESTI tuo lauseeseen ylipäättään epävarmuuden merkitystä: on epävarmaa, onko raunioissa ylipäättään ketään.

33) - Ensi vuonna yritetään *ehkä* järjestää **jotain** suurempaa tapahtumaa, mutta suunnitelmat ovat ihan alkuvaiheessaan.

34) *Ehkä joku* saattaa tykätä, että hän on saanut jo liikaa elämässään: kiinnostavan työn ja titeleitä.

35) Kirijenkön mukaan talous- ja valtiovarainministeriöihin on *mahdollisesti* tulossa **joitakin** henkilövaihdoksia.

36) Palokunta oli paikalla raivaustöissä vielä myöhään illalla eikä tiennyt, oliko raunioissa *mahdollisesti joku*.

Taulukko 8. Modaalisanojen EHKÄ ja MAHDOLLISESTI pronomnikollokaatit aineistossa.

EHKÄ			MAHDOLLISESTI		
kollokaatti	t-testin arvo	MI-testin arvo	kollokaatti	t-testin arvo	MI-testin arvo
se	8,63	0,78	muu	2,82	1,23
<u>jokin</u>	7,09	2,11	<u>jokin</u>	2,24	1,50
<u>joku</u>	4,65	2,08	ne	2,17	0,84
me	3,74	0,95	<u>joku</u>	1,93	1,89
minä	2,99	0,97			
he	2,44	0,59			
tämä	2,09	0,31			
muutama	2,18	0,82			

EHKÄ-partikkelilla on kuusi sellaista pronomnikollokaattia, jotka puolestaan eivät kollokoi tilastollisesti merkitsevästi MAHDOLLISESTI-sanalla: SE, ME, MINÄ, HE, TÄMÄ ja MUUTAMA. Kiinnostavana modaalisanoja erottavana piirteenä voidaan pitää erityisesti persoonapronomnikollokaatteja. MAHDOLLISESTI ei siis kollokoi tilastollisesti merkitsevästi ainoakaan persoonapronominin kanssa, mutta EHKÄ-sanalla tällaisia kollokaatteja on kolme, ME, MINÄ ja HE. Niistä itse asiassa kaksi ensin mainittua, 1. persoonan pronominit, ovat tilastollisesti hieman merkitsevempiä (t-testin arvot 3,74 ja 2,99) kuin MAHDOLLISESTI-sanalla koko pronomnikollokaattilistan kärjessä oleva MUU (2,82). Täten etenkin 1. persoonan pronominit esiintyvät tyypillisesti nimenomaan modaalisen partikkelin kontekstissa. Esimerkiksi virkkeessä (37) pronominilla ME viitataan joukkoon, johon puhuja itse kuuluu, ja virkkeessä (38) puhuja viittaa MINÄ-pronominilla itseensä. Toisaalta ME- ja MINÄ-pronominien lisäksi myös demonstratiivipronomini TÄMÄ ilmaisee eräänlaista puhujalähtöisyyttä. TÄMÄ ilmaisee nimittäin, että tarkoite kuuluu puhujan huomiopiiriin, siinä missä esimerkiksi SE kuuluu puhuteltavan huomiopiiriin (VISK § 720). Virkkeessä (39) TÄMÄ-pronominilla määritetään KYSYMYSSUBSTANTIIVIA, joka viittaa tekstissä aiemmin puheena olleeseen ja siten puhujan huomiopiirissä olevaan kysymykseen.

37) *Ehkä me* kuitenkin tehdään vielä yksi levy, ennen kuin aletaan tällaisia miettimään.

38) Sitten ajattelin, että koska olen itse niin paljon saanut, *ehkä minä* voin sitä jakaa myös muille.

39) *Ehkä tämä* kysymys toistuu siksi, että tapahtuma sai alkunsa poliittisen laululiikkeen vahvoina vuosina.

EHKÄ-sanalla siis voidaan sanoa olevan preferenssi sellaisiin pronominkollokaatteihin, jotka ilmaisevat jollakin tavalla puhujalähtöisyyttä. MAHDOLLISESTI-sanalla tällaista preferenssiä ei ole. Sillä on kaksi pronominkollokaattia, joita sen synonyymilla ei ole, NE ja MUU. Sana MUU viittaa esimerkiksi johonkin 'jäljelle jäävään' tai 'toiseen, erilaiseen' (KS s.v. *muu*). Esimerkiksi virkkeessä (40) mainitaan erikseen Ruotsi ja MUU-pronominilla sitten viitataan jäljelle jääviin Pohjoismaihin. Pronomini NE puolestaan osoittaa tarkoitteen olevan nimenomaan puhuteltavan, ei puhujan, huomiopiirissä (VISK § 722); näin on esimerkiksi virkkeessä (41).

40) Presidentti Martti Ahtisaari toivoo, että helikopterien hankinnassa on yhä mahdollista toimia yhteistyössä Ruotsin ja *mahdollisesti muiden* Pohjoismaiden kanssa.

41) Kun robotit jakavat informaationsa, *niistä* voi *mahdollisesti* tulla yhtä älykkäitä kuin ihmiset.

Toisaalta tulee huomata, että modaalisanat eivät eroa toisistaan pronominkollokaattien osalta siten, että EHKÄ-sanalla painottuisivat ainoastaan puhujalähtöisyyttä ilmaisevat sanat ja MAHDOLLISESTI-sanalla muut pronominit. Sen sijaan EHKÄ osoittaa preferenssiä myös sellaisiin pronomineihin (SE, HE, MUUTAMA sekä modaalisanojen yhteiset pronominkollokaatit JOKU ja JOKIN), jotka eivät ilmennä puhujalähtöisyyttä. Näitä ovat SE ja HE, edellä esitellyt modaalisanojen yhteiset pronominkollokaatit, indefiniittiset kvanttoripronominit JOKU ja JOKIN sekä niin ikään indefiniittinen kvanttoripronomini MUUTAMA. Esimerkiksi SE viittaa puhuteltavan huomiopiiriin yhtä lailla kuin NE (VISK § 722), joka kollokoi MAHDOLLISESTI-sanalla kanssa, ja sanan SE merkitsevyysarvo t-testissä on jopa 8,63. Se onkin t-testin mukaan itse asiassa sanan kaikkein merkitsevin pronominkollokaatti. Virkkeet (42)–(44) osoittavat, että siis myös muut kuin puhujalähtöisyyttä ilmaisevat sanat kollokoivat EHKÄ-partikkelin kanssa.

42) *Ehkä se* on vain keskittymiskyvyn puutetta.

43) Menestyksen tuomassa huumassa *he ehkä* muuttavat suuntansa ja alkavat harjoitella kovasti.

44) Kysymykseen tulevat siis yliopistosairaalat ja *ehkä muutama* keskussairaala.

4.2.2 Verbikollokaatit

Kummallakin modaalisanalla on runsaasti verbikollokaatteja: EHKÄ-sanalla 18 ja MAHDOLLISESTI-sanalla 16. Sanoilla on kuitenkin ainoastaan kaksi yhteistä verbikollokaattia, TULLA ja VOIDA. Molemmat niistä nousevat merkitseviksi t-testin perusteella, mutteivät MI-testin perusteella. Aineistosta on havaittavissa, että verbi TULLA esiintyy EHKÄ-sanan kollokaattina yleensä finiittimuodoissa (45), mutta MAHDOLLISESTI-sanan kollokaattina puolestaan usein infiniittimuodoissa (46). Kiinnostavaa on, että usein TULLA-verbiin liittyy tulevaisuuden ja siten ajankohdan merkitys, kuten siis useisiin muihinkin modaalisanojen kollokaatteihin.

45) *Ehkä* juhlista **tulee** nelipäiväiset?

46) Hienopaperin hintoihin *mahdollisesti* **tulevien** korotusten suuruutta Vaajoki ei vielä halunnut ennakoita.

Modaalisanojen toinen yhteinen verbikollokaatti, VOIDA, on modaalinen sana, aivan kuten EHKÄ- ja MAHDOLLISESTI-sanatkin. Modaaliverbin kollokointia modaalisanojen EHKÄ ja MAHDOLLISESTI kanssa voidaan ainakin osittain selittää modaaliharmonialla. Modaaliharmonialla tarkoitetaan sitä, että samassa lauseessa tai lausumassa käytetään useampaa kuin yhtä modaalisuuden ilmaisukeinoa, jotka vaikuttavat samansuuntaisesti ja usein täydentävät toisiaan. Modaaliharmonia ilmenee esimerkiksi käytettäessä modaaliverbiä VOIDA potentiaalimuodossa (*voinee*), mutta myös juuri esimerkiksi modaaliverbin (VOIDA) ja modaalisen adverbien tai partikkelien esiintyessä samassa lauseessa. (VISK § 1560.) Virkkeessä (47) epävarmuuden ja episteemisyyden merkitystä ilmaistaan sekä verbillä VOIDA että partikkelilla EHKÄ. Virkkeessä (48) tätä merkitystä taas ilmaistaan sekä VOIDA-verbillä että adverbilla MAHDOLLISESTI.

47) *Ehkä* hän **voi** ajatella, että ongelma olisi miedompi kuin miten me se nähdään.

48) Ruuanjakeluun **voidaan** kyllä saada *mahdollisesti* parannuksia.

Taulukko 9. Modaalisanojen EHKÄ ja MAHDOLLISESTI verbikollokaatit aineistossa.

EHKÄ			MAHDOLLISESTI		
kollokaatti	t-testin arvo	MI-testin arvo	kollokaatti	t-testin arvo	MI-testin arvo
ei ¹	8,07	0,71	<u>tulla</u>	3,66	1,18
johtua	4,29	2,66	liittyä	3,56	2,87
yllättää	4,27	2,62	<u>voida</u>	3,04	0,98
ymmärtää	3,41	1,79	aiheuttaa	2,93	2,69
tuntea	3,18	1,25	aiheutua	2,70	4,47
<u>voida</u>	2,90	0,44	seurata	2,60	1,87
<u>tulla</u>	2,77	0,43	myydä	2,28	2,07
odottaa	2,67	0,99	perustaa	2,21	2,20
vaikuttaa	2,28	0,99	siirtyä	2,08	1,93
ajatella	2,24	1,24	pyytää	1,73	2,14
luulla	2,17	1,86	etsiä	1,88	2,09
kaivata	2,11	1,46	käsitellä	1,61	1,83
huomata	2,08	1,24	tapahtua	2,00	1,78
osata	2,08	1,06	osallistua	1,86	1,75
sopeutua	1,88	2,65	lisätä	1,92	1,63
hymyillä	1,74	2,17	johtaa	1,91	1,62
odotella	1,53	1,67			
kuvitella	1,65	1,62			

Yhteisten verbikollokaattien lisäksi sanalla EHKÄ on 16 sellaista kollokaattia, joita ei ole sen synonyymilla MAHDOLLISESTI, ja MAHDOLLISESTI-sana puolestaan kollokoi 14 sellaisen sanan kanssa, joiden kanssa EHKÄ ei kollokoi. Nämä kollokaatit ovat luettavissa taulukosta 9. Aineistoa voidaan tarkastella Pajusen (2001) verbiluokittelun mukaisesti, jossa suomen kielen verbit jaotellaan primääri-A-verbeihin, primääri-B-verbeihin sekä sekundaariverbeihin. Modaalisanoja erottaa aineiston perusteella toisistaan etenkin se, että partikkelilla EHKÄ on synonyymiaan suurempi taipumus esiintyä primääri-B-verbien yhteydessä. Pajunen määrittelee primääri-B-verbit mentaalisia

¹ Kielto sana EI on luettu tässä verbiksi, sillä se kongruoi subjektin kanssa ja taipuu esimerkiksi indikatiivin preesensissä (*en, et, ei, emme, ette, eivät*). Toisinaan, esimerkiksi kieltävän vastauksen alussa, se voi olla tosin tulkittavissa myös partikkeliksi. (Laitinen 2004: 185–186; VISK § 1623–1625.) Tällaisia tapauksia on aineistossa kuitenkin vain vähän, joten käsitellen kieltosanaa tässä verbinä.

tiloja tai prosesseja kuvaaviksi verbeiksi erotuksena primääri-A-verbeistä, jotka eivät ole merkitykseltään mentaalisia (Pajunen 2001: 56–57). Näitä verbejä voidaan nimittää myös mentaalisiksi verbeiksi, kuten tehdään Isossa suomen kielioipissa. Mentaalisia verbejä kuvaillaan siinä niin, että ne kuvaavat mielen sisäisiä tiloja ja toimintoja ja edellyttävät siten elollista, tajunnallista osallistujaa. Niillä onkin yleensä elollistarkoitteinen täydennys, useimmiten subjekti. (VISK § 445.)

EHKÄ-partikkelin 18 kollokaatista seitsemän voidaan luokitella mentaalisten verbien ryhmään: YMMÄRTÄÄ, TUNTEA, AJATELLA, LUULLA, KAIVATA, HUOMATA ja KUVITELLA. Sanan MAHDOLLISESTI kollokaattien joukossa puolestaan on verbi PYYTÄÄ, joka voidaan niin ikään luokitella primääri-B-verbiksi eli mentaaliseksi verbiksi, vaikka se ilmentääkin puheaktia; Pajunen (2001: 56) verbiluokituksessa puheaktiverbit kuuluvat nimenomaan primääri-B-verbeihin. Esimerkissä (49) esiintyy t-testin mukainen EHKÄ-sanan merkitsevin mentaaliverbikollokaatti YMMÄRTÄÄ ja esimerkissä (50) MI-testin mukainen merkitsevin mentaaliverbikollokaatti LUULLA.

49) Tätä taustaa vasten uudistuksen arvoa ei *ehkä* vielä täysin **ymmärretä** Itä-Suomessa.

50) Kysymys on Pasilan pommi-iskusta, jonka tutkinta on osoittautunut vaikeammaksi – ja kalliimmaksi – kuin aluksi *ehkä* **luultiin**.

Mentaaliverbien suuri määrä paljastaa siis, että EHKÄ-sanana kontekstissa esiintyy tyypillisesti elollinen, tajunnallinen osallistuja, muun muassa *suomalaiset* virkkeessä (51) ja ilmipanematon subjekti *minä* virkkeessä (52). Verbi TUNTEA esiintyy muita mentaaliverbejä useammin infiniittimuodossa, jolloin tajunnallinen osallistuja ei tule virkkeen merkityksessä vahvasti esiin, kuten virkkeestä (53) huomataan. Toisin kuin korpuksessa, muotoa TUNNETTU pidetään Kielitoimiston sanakirjassa (KS) itse asiassa omana, adjektiivistuneena lekseminään, jonka merkitys on ”joka tunnetaan, tiedetään yleisesti, tuttu; huomattava, maineikas, kuulu, nimekäs”. Tällöin merkityksessä korostuu pikemminkin yleisesti vallitseva tunnettuus kuin se, että toimijana olisi mikään yksittäinen elollinen, tajunnallinen osallistuja, kuten muilla mentaaliverbikollokaateilla on.

51) *Ehkä* suomalaiset silloin **ajattelevat**, että on perusteltua jatkaa tukea ohjelmalle.

52) Solistin pianismi on monipuolista ja soinniltaan rikasta – *ehkä* pienimpiä pianissimoja jäin **kaipaamaan** – ja hän musiikin mukaan tempautuessaan saa ilmaisuunsa syvyyttä –

53) Rääkkylän Kihaus on tällä hetkellä *ehkä* Pohjois-Karjalan **tunnetuin** festivaali.

Kummallakin modaalisanalla on kollokaattilistassaan myös primääri-A-verbejä. Primääri-A-verbeillä ilmaistaan konkreettista, ei siis mentaalista, tilaa tai prosessia. EHKÄ-sanalla tällaisia kol-

lokaatteja on kahdeksan: JOHTUA, YLLÄTTÄÄ, TULLA, ODOTTAAN, VAIKUTTAA, SOPEUTUA, HYMYILLÄ ja ODOTELLA. MAHDOLLISESTI-sanalla puolestaan näitä kollokaatteja on jopa neljätoista: TULLA, LIITTYÄ, AIHEUTTAA, AIHEUTUA, SEURATA, MYYDÄ, PERUSTAA, SIIRTYÄ, ETSIÄ, KÄSITELLÄ, TAPAHTUA, OSALLISTUA, LISÄTÄ ja JOHTAA. Pajunen (2001: 94) luokittelee primääri-A-verbit kahteen aspektityyppiin, tilaa ja prosessia ilmaiseviin verbeihin. Koska yksi ja sama verbilekseemi voi edustaa molempia tyyppisiä (mts. 45), ratkaisevaa on yksittäisen sanaesiintymän merkitys. Tässä analyysissä en kuitenkaan tarkastele, miten verbit jakautuvat tilan ja prosessin verbityyppisiin, sillä en tarkastele verbejä sanaesiintymittain, vaan lekseemeittain. Useimmat mainitut verbit näyttävät kuitenkin ilmaisevan ennen kaikkea prosessia eivätkä staattista tilaa, kuten SOPEUTUA virkkeessä (54), MYYDÄ virkkeessä (55) ja PERUSTAA virkkeessä (56).

54) *Ehkä* ruotsalaistenkin on vähitellen **sopeuduttava** yrittäjän riskeihin.

55) Jos pipoja tulee vuori kuten tuli sukia, *mahdollisesti* osa **myydään** – –

56) Postivirkamiesliitto vaatii, että syksyllä *mahdollisesti* **perustettava** Postin ja Postinpankin verkostoyhtiö tulisi PT-konsernin eli entisen Posti-Telen alaisuuteen.

Kiinnostava havainto on lisäksi, että molemmilla modaalisanoilla on taipumus esiintyä kausaalista suhdetta ilmaisevien verbien seurassa: EHKÄ-sanalla tällaisia ovat usein lekseemit JOHTUA (57) ja VAIKUTTAA (58) ja MAHDOLLISESTI-sanalla AIHEUTTAA (59), AIHEUTUA ja JOHTAA (60) sekä joissain tapauksissa myös SEURATA. Tämä piirre tuli MAHDOLLISESTI-sanalla esiin MI-testin kärkikollokaateista, mutta kaikkien kollokaattien analyysi osoittaa, että myös EHKÄ-sana esiintyy tällaisissa konteksteissa. Voidaan ajatella, että syy-seuraussuhteiden toteutuminen on usein hieman epävarmaa eikä niitä haluta esittää varmoina, ja siksi niitä ikään kuin pehmennellään modaalisanoilla.

57) *Ehkä* se **johtuu** koulusuunnistuksesta, jota ei koeta kovin innostavaksi, Romppanen arvelee.

58) *Ehkä* kehitykseen **vaikuttaa** myös kolmas seikka, kuten uusimmassa Lieksan tapauksessa, kymmenpaikkaisen hoitokoti Oma Tuvan aloittamisessa.

59) – – ettei siihen lääninoikeuden näkemyksen mukaan sisältynyt riittäviä määräyksiä laitoksen toiminnan ja sen *mahdollisesti* **aiheuttamien** päästöjen tarkkailemiseksi.

60) Diabetekseen *mahdollisesti* **johtava** geeni voidaan havaita jo napaverestä, jolloin selviää, onko lapsella suuri, normaali vai pieni riski sairastua diabetekseen.

Kolmantena verbiryhmänä primääri-A-verbien sekä primääri-B-verbien lisäksi voidaan pitää sekundaariverbejä. Sekundaariverbit ovat verbejä, jotka eivät varsinaisesti koodaa itse asiointiloja, vaan niiden avulla asiointiloja kvalifoidaan tai suhteutetaan toisiinsa. Sekundaariverbejä käytetään usein infinitiivien kanssa (esim. *alkaa itkeä*, *ruveta satamaan*, *saattaa antaa*). (Pajunen 2001: 51–56.) Aineiston sekundaariverbeistä VOIDA kollokoi molempien modaalisanojen kanssa ja OSATA sekä EI kollokoivat EHKÄ-sanalla. Kieltoisanan EI arvo t-testissä on huomattavan suuri, 8,07 – esimerkiksi seuraavaksi merkitsevimmän verbikollokaatin JOHTUA saama arvo on 4,29. Modaalisanoista nimenomaan EHKÄ-sanalla on siis voimakas kollokaatiosuhde kieltosanaan; esimerkkejä EHKÄ-sanalla ja kieltosanan kollokoinnista on virkkeissä (61) ja (62). Muita sekundaariverbejä sanoilla ei ole tilastollisten merkitsevien kollokaattien joukossa.

61) Näky **ei** ole *ehkä* kaunis, mutta tarkoitus on hyvä.

62) Pelvon mukaan matkakirjallisuudella **ei** *ehkä* ole enää samanlaista informatiivista merkitystä kuin ennen, mutta toisaalta elämyksellinen matkakirjallisuus kukoistaa taas.

Kaiken kaikkiaan verbikollokaattianalyysi osoittaa siis, että molemmat modaalisanat kollokoivat tilastollisesti merkitsevästi hyvin monen verbin kanssa. Siinä missä verbien ilmaisema merkitys on MAHDOLLESTI-sanalla tavallisesti konkreettista, myös EHKÄ-sanalla se on usein konkreettista, mutta toisaalta usein mentaalistakin. Modaalisanojen yhteydessä esiintyy myös sekundaariverbejä. Lisäksi kolligaatiopiirteinä verbikollokaateista nousee esiin, että MAHDOLLESTI-sanalla on voimakas taipumus esiintyä infiniittimuotoisten verbien yhteydessä. Tämä tuli esiin jo TULLA-verbien yhteydessä, mutta näkyy myös esimerkiksi virkkeissä (56), (59) ja (60). Esimerkin (55) tavoin sanalla on toki myös finiittimuotoisia verbikollokaatteja, mutta huomattavan suuri osa sanan verbikollokaateista esiintyy kuitenkin infiniittimuodoissa. EHKÄ-sanalla tällainen kolligaatio ei ole yhtä voimakas.

4.2.3 Adverbikollokaatit

Kun verrataan toisiinsa sanojen EHKÄ ja MAHDOLLESTI tilastollisesti merkitseviä adverbikollokaatteja, havaitaan merkittävä ero jo kollokaattien määrässä. Sanalla EHKÄ on 26 adverbikollokaattia ja sanalla MAHDOLLESTI ainoastaan kolme. Näin ollen yksi modaalisanojen keskeisistä kollokatio-naalisista eroista on, että EHKÄ kollokoi huomattavasti useammin adverbien kanssa kuin synonyyminsa MAHDOLLESTI. Kyseessä on toisaalta samalla myös kolligaatiopiirre. Tilastollisesti merkitsevät adverbikollokaatit on esitetty taulukossa 10.

Taulukko 10. Modaalissanojen EHKÄ ja MAHDOLLISESTI adverbikollokaatit aineistossa.

EHKÄ			MAHDOLLISESTI		
kollokaatti	t-testin arvo	MI-testin arvo	kollokaatti	t-testin arvo	MI-testin arvo
hieman	7,18	3,19	jo	5,30	1,88
vähän	7,05	2,34	<u>vielä</u>	2,92	1,27
liian	6,58	2,66	<u>myöhemmin</u>	2,14	2,40
paljon	6,45	1,28			
siksi	5,94	2,90			
<u>vielä</u>	5,90	1,19			
kaikkein	5,68	4,21			
joskus	4,77	2,56			
liikaa	4,31	3,05			
hiukan	4,25	3,40			
hyvin	4,20	1,03			
sitten	3,81	0,95			
koskaan	3,76	2,10			
niin	2,96	0,61			
<u>myöhemmin</u>	2,63	1,47			
silloin	2,62	1,18			
tässä	2,53	1,53			
osittain	2,27	2,03			
pian	2,18	1,54			
hanakasti	2,16	4,87			
täällä	2,09	1,06			
tarpeeksi	2,02	1,80			
toisin	1,79	2,33			
jotenkin	1,71	2,08			
osaltaan	1,84	2,01			
osin	1,65	1,93			

Modaalisanoilla on kaksi yhteistä adverbikollokaattia, MYÖHEMMIN ja VIELÄ, joista ensin mainittu voidaan luokitella aikaa ilmaisevaksi adverbiksi. Esimerkiksi virkkeessä (63) MYÖHEMMIN- ja EHKÄ-sanojen kollokointia voidaan selittää sillä, että MYÖHEMMIN viittaa tulevaan aikaan ja EHKÄ tulevan ajan epävarmuuteen, siihen, että ei ole täysin varmaa, onko myöhemmin tosiaan kyse toisesta viidestä miljoonasta kuutiosta. Samoin virkkeessä (64) MYÖHEMMIN viittaa tulevaan aikaan ja MAHDOLLISESTI ilmentää tulevaisuuteen kohdistuvaa epävarmuutta: vielä ei voida varmasti sanoa, tulevatko puheena olevat kiinteistöt myyntiin. Samoin kollokaatti VIELÄ ilmentää aikaa monissa konteksteissa, kuten esimerkeissä (65) ja (66). Sana VIELÄ voi toisaalta myös esimerkiksi toimia komparatiivijohdoksen määrittänä ja osoittaa, että vertailtavalla asiallakin on kyseessä olevaa ominaisuutta suhteellisen paljon (VISK § 640). Esimerkiksi virkkeessä (67) VIELÄ-sanalla ilmaistaan, että myös toinen ajankohta saattaa olla ratkaiseva, mutta ehkä sitäkin ratkaisevampi ajankohta on 16. elokuuta.

63) Ylöjärven Takamaalla on vireillä viisi murskauslupaa, ensin alkuun viisi miljoonaa kuutiota, **myöhemmin** ehkä toinen mokoma lisää.

64) Mielenkiintoinen yksityiskohta myyntiin **myöhemmin** *mahdollisesti* tulevista kiinteistöistä on keskustan läheisyydessä Pielisen rannalla oleva Monolanniemi.

65) *Ehkä* hänellä ei sittenkään **vielä** tällä kaudella tule liigadebyyttiä.

66) Sinne rakenteilla oleva linja kytketään pääverkkostoon *mahdollisesti* **vielä** tämän vuoden aikana.

67) Vapaa-ajan kalastajan kannalta *ehkä* **vielä** ratkaisevampi ajankohta on Joensuussa elokuun 16. päivä, jolloin kielto koskee myös viehekalastusvälineitä.

68) Rakentaminen alkaa *mahdollisesti* **jo** ensi vuoden syksyllä.

MAHDOLLISESTI-sanalla on kaiken kaikkiaan kolme adverbikollokaattia, joista MYÖHEMMIN- ja VIELÄ-adverbien lisäksi myös kolmas, sana JO, esiintyy tavallisesti ajan adverbina. Tätä kuvaa esimerkkivirke (68). Kaikki MAHDOLLISESTI-sanana adverbikollokaatit siis ilmentävät jollakin tavalla aikaa, joten sanalla voidaan tämän perusteella esittää olevan semanttinen preferenssi 'aika'. Sama oletus semanttisesta preferenssistä pätee myös EHKÄ-sanaan: se niin ikään kollokoi siis MYÖHEMMIN- ja VIELÄ-sanojen kanssa – ei kuitenkaan JO-sanana kanssa, ja lisäksi sillä on jopa viisi muuta aikaan viittaavaa adverbikollokaattia. Niitä ovat JOSKUS, SITTEN, KOSKAAN, SILLOIN ja PIAN. Näin ollen voidaan sanoa, että molemmille modaalisanoille tyypillinen piirre on, että niiden yhteydessä esiintyy aikaa ilmaisevia adverbeja. Virkkeessä (69) epävarmuutta ilmaisevan EHKÄ-sanana kollokaattina esiintyy adverbi JOSKUS, joka sekin ilmaisee eräänlaista epävarmuutta, nimittäin ajankoh-

taan kohdistuvaa epämääräisyyttä. Virkkeessä (70) EHKÄ-sanan kanssa kollokoiva adverbipuolesta ei ilmennä epämääräisyyttä, vaan sen sijaan eräänlaista ehdottomuutta. Adverbilla KOSKAAN kielletään kaikki ajankohdat (VISK § 649).

69) *Ehkä* se arvoitus **joskus** ratkeaa.

70) Monet kehitykset ovat niin paljon velkaa teollisuusmaille, että eivät *ehkä* **koskaan** selviä veloistaan.

MAHDOLLISESTI-sana ei siis kollokoi tilastollisesti merkitsevästi muiden adverbien kuin ensisijaisesti aikaa ilmaisevien kanssa, mutta EHKÄ-sanalla tilastollisesti merkitseviä adverbikollokaatteja on sitä vastoin paljon muitakin. Kaiken kaikkiaan suomen kielen adverbeja voidaan luokitella ajan adverbien lisäksi paikan, tavan ja määrän adverbeihin sekä kommenttiadverbeihin; lisäksi on erilaisia merkityksiä ilmaisevia pronominkantaisia adverbeja eli proadverbeja (VISK § 647, 715, määritelmät s.v. proadverbi). EHKÄ-sanan kollokaatit TÄSSÄ ja TÄÄLLÄ ilmaisevat puheyhteydestä ilmenevää paikkaa tai muuta tilaa tai tilannetta (KS s.v. *tässä, täällä*), kuten esimerkissä (71) konkreettista sijaintia ja esimerkissä (72) puheyhteydestä ilmenevää tilannetta. Usein EHKÄ-sanan kontekstissa ilmaistaan myös tekemisen tapaa: HANAKASTI, HYVIN, NIIN, TOISIN ja JOTENKIN ilmaisevat usein tapaa. Esimerkiksi virkkeessä (73) kuvataan adverbien avulla, miten ihmiset suhtautuvat humoristen lahjojen hankkimiseen ja virkkeessä (74), miten puun erottelu onnistuu eri sijainneista käsin. Virkkeessä (75) puolestaan ETTÄ-konjunktilla alkava sivulause ilmaisee tapaa, johon adverbipuolesta NIIN viittaa. Tosin sana NIIN toimii usein myös määritteen kaltaisena sanana, kuten esimerkissä (76). HYVIN puolestaan on usein intensiteettimäärite (VISK § 605; KS s.v. *hyvin*), kuten virkkeessä (77). SIKSI-sana ei varsinaisesti ilmaise paikkaa, tapaa tai määrää, vaan syytä tai seurausta; sana voidaan luokitella proadverbiksi (VISK § 668). Esimerkiksi virkkeessä (78) SIKSI ilmaisee syytä.

71) Rotuvähemmistöt eivät *ehkä* **täällä** keskisessä Suomessa ole vielä nousseet niin paljon esille kuin etelässä.

72) *Ehkä* myös kokemattomuus näkyi **tässä**, kommentoi synkkänä ViU:n pelinjohtaja Jukka Varonen.

73) Winqvist arvelee, että *ehkä* aiempaa **hanakammin** isien paketteihin on tänä vuonna käärity humoristisia kravatteja.

74) Erikoispuiden erottelu käy *ehkä* **paremmin** puun juurelta kuin koneen ohjaamosta.

75) *Ehkä* siinä on käynyt **niin**, että vesi ei ole vaihtunut riittävästi – –

76) Kilpailua käydään myös Joensuun seudulla, joskaan ei *ehkä niin* rajuna kuin ruuhka-Suomessa.

77) Itä-Suomi-yhteistyötä voidaan toteuttaa siellä, *ehkä hyvin* kapeillakin alueilla, missä se on mahdollista ja omaa elinkeinoelämää hyödyttävää – –

78) Turvallisuusneuvosto haluaa *ehkä siksi* tehdä omat päätelmänsä tässä suhteessa selontekoni pohjalta, Annan mainitsi.

Edellä esitellyn HYVIN-sanan lisäksi moni muukin EHKÄ-sanan kollokaateista ilmaisee määrää ja intensiteettiä. Näistä tilastollisen merkitsevyyden kärkipäässä on etenkin määrän suuruutta ja tavan intensiteettiä osoittavia sanoja: HIEMAN, VÄHÄN, LIIKAA ja HIUKAN ovat tilastollisesti merkitseviä molempien testien mukaan, ja niistä HIEMAN (t-testin arvo 7,18) ja VÄHÄN (7,05) ovat aivan t-testin kärkikollokaatteja. Määrän suuruutta ja tavan intensiteettiä ilmaisevat sanat toimivat tyypillisesti astemääritteinä eli adjektiivin tai adverbien määritteinä (VISK määritelmät s.v. astemäärite), kuten esimerkeissä (79) ja (80). Toisaalta ne voivat myös liittyä verbiin ja toimia adverbiaaleina, kuten esimerkeissä (81) ja (82). Niin ikään sanat PALJON ja TARPEEKSI ilmaisevat määrän suuruutta tai tavan intensiteettiä. PALJON esiintyy usein nimenomaan komparatiivimuodossa astemääritteenä, kuten esimerkissä (83).

79) Tulevista mökeistä tehdään *ehkä hieman* suurempia.

80) Helsinkiläisnäyttelijä tulkitsee taitavasti ja persoonallisestikin hempeitä rakkauslauluja, mutta *ehkä vähän* yksiulotteisesti.

81) *Ehkä* muut yrittivät *liikaa* ja minä taas juoksin sen minkä osaan – –

82) *Ehkä* Finlandiaa pitäisi *hiukan* tuulettaa, jotta kirjallisuuskeskustelu viriäisi jälleen.

83) Vaikka Suomessa on nykyisin puuta *ehkä enemmän* kuin koskaan, talousmetsien yksi-puolinen käsittely on merkinnyt luonnon monimuotoisuuden vähenemistä.

Sanaa LIIAN voidaan pitää partikkelina tai adverbina (VISK § 658, 664). Se kuitenkin joka tapauksessa kuuluu merkitykseltään yhteen määrän adverbien kanssa, sillä se ilmaisee liiallisuutta. Lisäksi sitäkin käytetään astemääritteenä, aivan kuten edellä mainittuja määrän suuruutta ja tavan intensiteettiä osoittavia sanoja. Kaiken kaikkiaan astemääritteiden ja EHKÄ-sanan yhteisesiintymistä havaitaan, että EHKÄ esiintyy usein rakenteessa EHKÄ + astemäärite + adjektiivi (79, 84) ja rakenteessa EHKÄ + astemäärite + adverbi (80).

84) Meille tuli *ehkä liiankin* hyvä tunne välillä ja vastustaja pääsi takaisin mukaan peliin.

Muita määrän adverbija ovat OSITTAIN, OSALTAAN ja OSIN, joilla ei ilmaista niinkään määrän suuruutta, vaan pikemmin osoitetaan täyteen astetta ja likimäärää (määrän adverbien luokittelusta ks. VISK § 656–657). Esimerkiksi virkkeessä (85) todetaan, että museoiden vähäinen mainostus on osasy, ei välttämättä ainoa syy, puheena olevaan asiaan. OSITTAIN, OSALTAAN ja OSIN ilmaisevatkin eräänlaista epävarmuutta. Sana KAIKKEIN puolestaan sekin muistuttaa määrän adverbija: se tähdentää adjektiivin superlatiivimuotoisuuden yksiselitteisyyttä (VISK § 645), kuten virkkeessä (86). Sillä ei kuitenkaan siis ilmaista epävarmuutta, vaan pikemminkin siis varmuutta, yksiselitteisyyttä ja siten epävarmuudelle päinvastaista ilmiötä. Kiinnostavaa onkin, että EHKÄ kollokoi siis keskenään tavallaan päinvastaisia merkityspiirteitä ilmaisevien sanojen kanssa.

85) *Ehkä* tämä johtuu **osittain** siitä, etteivät museot juuri mainosta olemassaoloaan.

86) Työvaiheista *ehkä* **kaikkein** mielenkiintoisin on maalausvaihe.

Kaiken kaikkiaan tilastollisesti merkitsevien adverbien analyysi osoittaa, että modaalisanalla EHKÄ on erittäin paljon tilastollisesti merkitseviä adverbikollokaatteja. MAHDOLLISESTI-sanallakin niitä on muutama, mutta ero adverbikollokaattien määrissä on kuitenkin niin suuri, että sanalla EHKÄ voidaan olettaa olevan jonkinlainen synonyymiaan selkeämpi semanttinen preferenssi paikan, tavan ja etenkin määrän ilmaisemiseen. Tätä tarkastellaan lisää semanttista preferenssiä koskevassa luvussa 5.

4.2.4 Partikkeli- ja adpositiokollokaatit

Partikkelit ja adpositiot muodostavat keskenään eri sanaluokat, mutta tässä tutkielmassa niitä käsitellään yhtenä ryhmänä. Modaalisanojen kollokaateista partikkeli- ja adpositiokollokaattien joukossa on EHKÄ-sanalla kymmenen kollokaattia ja MAHDOLLISESTI-sanalla viisi, ja ne ovat nähtävissä taulukossa 11. Näistä kollokaateista molemmille modaalisanoille yhteiset kollokaatit ovat JOPA ja TAI. Niistä JOPA on fokuspartikkeli, joka nostaa fokukseen jonkin osan lauseesta ja jättää tällöin lauseen muun osan taustalle (VISK § 792, määritelmät s.v. fokuspartikkeli). Se on MAHDOLLISESTI-sanan kaikkein merkitsevin kollokaatti t-testissä koko aineistossa; sen arvo on t-testissä 5,57. Molempien modaalisanojen kollokaattina sana JOPA nostaa lauseessa fokukseen usein nimenomaan jonkin määrän ilmauksen, kuten esimerkissä (87), mutta toisaalta joskus myös esimerkiksi substantiivin, kuten esimerkissä (88). TAI puolestaan on konjunktio ja ilmaisee eräänlaista epävarmuutta,

useiden vaihtoehtojen olemassaoloa. EHKÄ-sanan kollokaattina se esiintyy tyypillisesti virkkeen alussa ikään kuin irrallisen lisäyksen aloituksena, kuten esimerkissä (89), ja MAHDOLLISESTI-sanan kollokaattina puolestaan luettelossa kahta tai useampaa vaihtoehtoa erottamassa, kuten esimerkissä (90).

87) Hasista on *mahdollisesti* tuotu maahan **jopa** 200 kiloa.

88) Kriisin diplomaattinen sopiminen on koko arabialueen ja *ehkä* **jopa** maailmanrauhan kannalta toivottavaa.

89) Niin Mäkinen kuin hänen Ford Sierransakin oli ollut ennen kisaa viisi vuotta kilpailematta, mutta tauko ei tuloksissa näkynyt. **Tai** *ehkä* hieman, sillä Mäkisen ensimmäinen startti meni poskelleen ja se maksoi monta kallista sekuntia.

90) Yllään hänellä on *mahdollisesti* farkkuposakka **tai** tummansininen kansitakki.

Taulukko 11. Modaalisanojen EHKÄ ja MAHDOLLISESTI partikkeli- ja adpositiokollokaatit aineistossa.

EHKÄ			MAHDOLLISESTI		
kollokaatti	t-testin arvo	MI-testin arvo	kollokaatti	t-testin arvo	MI-testin arvo
mutta	9,18	1,29	<u>jopa</u>	5,57	3,58
<u>tai</u>	4,77	0,96	<u>tai</u>	4,64	1,91
<u>jopa</u>	4,58	1,65	ja	2,88	0,34
sittenkin	3,88	4,11	sekä	2,18	1,00
no	3,47	3,27	vai	1,98	2,38
kuin	3,06	0,55			
kuluttua	2,88	2,12			
kuitenkin	2,73	0,66			
lukuunottamatta	2,39	2,29			
sentään	1,92	1,87			

Yhteisten partikkelikollokaattien lisäksi modaalisanoilla on omia, erillisiä partikkeli- tai adpositiokollokaattejaan. EHKÄ-sanan omista kollokaateista ensinnäkin MUTTA, SITTEINKIN, KUITENKIN, LUKUUNOTTAMATTA ja SENTÄÄN ilmaisevat kukin eräänlaista vastakkaisuutta. Tätä ilmentävät virkkeet (91)–(93). Silmiinpistävää on, että MUTTA-konjunktion t-testin arvo on peräti 9,18. MUTTA onkin EHKÄ-sanan kaikkein merkitsevin kollokaatti tutkielman koko aineistossa t-testin mukaan. Sanan yhtenä semanttisena preferenssinä voidaan täten alustavasti pitää vastakkaisuutta.

91) Vaikutus tuntuu *ehkä* hetkellisesti, **mutta** en usko että se näyttelee jatkossa kovin suurta osaa, vaan tilanne tasoittunee nopeasti.

92) En tiedä, olisi *ehkä* kannattanut **sittenkin** mennä itse avaamaan se ura, Virtanen pohtii.

93) Vaikka pyynti on nykyisin tehokkaampaa, ei merkittävästi laskevaa trendiä ole näkyvissä *ehkä* ahventa **lukuunottamatta**.

Muut EHKÄ-sanan omat partikkeli- ja adpositiokollokaatit ovat NO, KUIN ja KULUTTUA. Näistä NO on puhekielelle tyypillinen partikkeli. Se on puhutussa kielessä kaikkein yleisimpiä partikkeleita (VISK § 808). Voidaankin pitää osin yllättävänä, että se nousee sanomalehtiaineistossa EHKÄ-sanan tilastollisesti merkitseväksi kollokaatiksi. Se lisäksi ylittää raja-arvon molemmissa tilastollisissa testeissä. Sanaa voidaan käyttää lausuman tai vuoron alussa lausumapartikkelina (VISK § 796), missä tehtävässä se aineistossa esiintyykin. NO-sanan tyypillisestä käytöstä virkkeen alussa on esimerkkinä virke (94). NO on yksi harvoista sanoista, jotka kollokoivat ainoastaan vasemmalla puolella modaalisanaan nähden. Sen sijaan EHKÄ-sanan synonyymi MAHDOLLISESTI ei kollokoi NO-partikkelin kanssa, ja tämän havainnon modaalisanojen eroista voidaan ajatella tukevan hypoteesia, jonka mukaan MAHDOLLISESTI on sävyltään virallisempi ja EHKÄ hieman puhekielisempi. KUIN-konjuktiolla puolestaan ilmaistaan vertailua: vaikkapa esimerkissä (95) Makhtinen soittoa verrataan muiden soittajien suoritukseen. KULUTTUA taas on temporaalinen postpositio (VISK § 697), jolla on merkityksensä puolesta yhteneväisyys ajan adverbeihin. Esimerkiksi virkkeessä (96) sen avulla ilmaistaan, että onnettomuuden ja ambulanssin saapumisen välillä on kulunut tietty aika.

94) Karuselli-kilpailun menestys ei tuntunut juuri missään. **No**, *ehkä* esiintymispyyntöjä tuli hieman lisää, mutta ei niin paljon kuin olisi voinut luulla.

95) Nevanlinnan kappaleen Makhtine soitti *ehkä* paremmin **kuin** kukaan muu.

96) Ambulanssin pillit rupesivat soimaan *ehkä* kolmen minuutin **kuluttua** tapahtumasta, hän arvioi.

MAHDOLLISESTI-sanallakin on joitakin omia partikkelikollokaatteja. Se kollokoi paitsi edellä esitellyn TAI-konjunktion kanssa, myös toisen vaihtoehtoista eli disjunkttiivista suhdetta ilmaisevan konjunktion, VAI-sanan, kanssa (disjunkttiivisesta suhteesta ks. VISK § 1098), kuten virkkeessä (97). Niin ikään se kollokoi kahden additiivista suhdetta ilmaisevan konjunktion kanssa, joita ovat JA ja SEKÄ. Sen voidaan siten sanoa ilmaisevan hienoista preferenssiä additiivisiin suhteisiin synonyymiinsa nähden: sana EHKÄ ei kollokoi additiivisten konjunktioiden kanssa tilastollisesti merkitsevästi. Virkkeessä (98) konjunktiolla JA lisätään virkkeen sisälle toinen verbi-ilmaus, ja virkkeessä (99) konjunktiolla SEKÄ puolestaan lisätään lauseeseen uusi luettelon osa, tässä tapauksessa yksi mauste.

97) Vähintä mitä viranomaiset voivat nyt tehdä, on selvittää perusteellisesti, onko ongelma todellinen **vai** onko kysymys *mahdollisesti* vain väärinkäsityksestä.

98) Tarkoitus on korvata Venäjän tuontia paikallisella tukilla **ja** *mahdollisesti* lisätä sahausmääriä.

99) Leivänpalat paistetaan tietenkin oliiviöljyssä ja maustetaan suolalla ja pippureilla **sekä** *mahdollisesti* yrteillä.

Liitteestä 1 on nähtävissä, että Lemmie-korpusohjelma nostaa yhdeksi EHKÄ-sanan tilastollisesti merkitseväksi kollokaatiksi välimerkin, pilkun. Pilkku voi esiintyä päälauseiden välissä, päälauseen ja sivulauseen välissä sekä useissa muissa yhteyksissä, kuten luettelon osien välissä, lainauksen ja johtolauseen välissä ja lisähuomautusten kaltaisten jaksojen yhteydessä (Kielitoimiston ohjepankki 2015a; 2015b). Välimerkit eivät ole kollokaatteja, vaikka ohjelma nostaakin pilkun kollokaatiksi. Pilkku kollokaattilistassa kuitenkin voi tietyissä määrin paljastaa, millaisia kolligatiopiiirteitä EHKÄ-sanalla on. Partikkeli EHKÄ on usein osa sivulauseita (100) tai sellaista päälauseita, jota seuraa toinen päälause (101). Sanan kotekstissa voi myös esiintyä sivulause (102). Toisaalta partikkeli voi myös aloittaa muusta kokonaisuudesta irrallisen lisähuomautuksen, kuten virkkeessä (103). Lisäksi EHKÄ voi esiintyä esimerkiksi lainauksen ja johtolauseen välissä, kuten esimerkissä (104). Pilkku esiintyy EHKÄ-sanaan nähden niin vasemmalla puolella (697 tilastollisesti merkitsevää esiintymää) kuin oikeallakin puolella (785 tilastollisesti merkitsevää esiintymää). Sanan MAHDOLLISESTI kollokaattilistassa pilkkua ei ole, joten on mahdollista, että esimerkiksi sivulauseita, pilkulla toisistaan erotettavia päälauseita, lainauksia tai irrallisia lisähuomautuksia ei esiinny sen kotekstissa yhtä frekventisti kuin sanan EHKÄ kotekstissa. En kuitenkaan syvenny tähän tarkemmin tässä tutkielmassa, sillä tutkielman varsinaisena tutkimuskohteena ovat nimenomaan kotekstin leksikaaliset ja semanttiset piirteet.

- 100) Samana päivänä kilpaileva Kauppalehti pysyttelee niiden parissa, jotka *ehkä* huonoiten ymmärtävät niin Topeliusta ja Juteinia kuin kettutyttöjäkin.
- 101) Siemeniä innostumiseen on *ehkä* olemassa, mutta ne eivät nykytilanteessa jotenkin pääse itämään.
- 102) Suomessa ei vielä *ehkä* oikein ymmärretä sitä, että kilpailujen merkitys on hirveästi vähentynyt 20 vuoden aikana.
- 103) Valot meillä oli, *ehkä* joku sähkölevykin. Helloja tai jääkaappeja ei vielä ollut.
- 104) Äiti ei kyllä siitä *ehkä* tykkää, Paavo arvelee.

5 SEMANTTINEN PREFERENSSI

5.1 Luokittelun perusteet

Edellä on tarkasteltu modaalisanojen tilastollisesti merkitsevien kollokaattien koko kirjoa, käsitelty näiden kollokaattien merkityksiä ja muodostettu samalla myös hypoteeseja siitä, millaiset semanttiset ryhmät missäkin sanaluokassa näyttävät olevan tyypillisiä toisaalta sanan EHKÄ, toisaalta sanan MAHDOLLISESTI kontekstissa. Analyysin pohjalta on siis syntynyt joitakin oletuksia modaalisanojen semanttisista preferensseistä. Varsinaisesti semanttisiin preferensseihin syvennytään tässä luvussa.

Luvussa 3.3.4 kuvailtujen periaatteiden pohjalta olen päätenyt luokittelemaan aineiston sanat semanttisen preferenssin analyysin pohjaksi yhteensä 19 semanttiseen luokkaan. Niitä ovat 'tila', 'prosessi', 'mentaalisuus', 'viestintä', 'aika', 'paikka', 'määrä, intensiteetti', 'ihmiset', 'terveydentila', 'urheilu, kilpailu', 'konkreettiset oliot', 'kulttuuri', 'yhteiskunnallisuus, institutionaalisuus', 'epävarmuus', 'additiivisuus', 'kausaalisuus', 'vastakkaisuus', 'arvo, laatu' ja lisäksi ryhmä 'luokittelemattomat'. Liitteessä 3 on esitetty, mitä sanoja mihinkin luokkaan kuuluu. Merkityskategorioista neljä perustuu Pajusen (2001) verbiluokitukseen: 'tila', 'prosessi', 'mentaalisuus' ja 'viestintä'. Luokitteluni ei kuitenkaan täysin mukaile Pajusen verbiluokkia. Seuraavaksi ensinnäkin selostan, millä tavalla luokitukseni neljä kategoriata pohjautuvat Pajusen verbiluokkiin ja millaisia sanoja näihin merkitysryhmiin luokittelussani kuuluu. Sen jälkeen esittelen luokitukseni muut 15 semanttista ryhmää ja perusteluja sille, miten olen päätenyt ottamaan juuri nämä merkitysryhmät semanttisen preferenssin analyysin pohjaksi. Esittelen myös esimerkkejä siitä, mitkä sanat ovat olleet luokittelussa ongelmallisia ja millaisiin ratkaisuihin olen päätenyt niiden ryhmittelyssä.

Kuten jo luvussa 4.2.2 verbikollokaattien yhteydessä tuli ilmi, Pajunen (2001) luokittelee suomen kielen verbit primääri-A-verbeihin, primääri-B-verbeihin sekä sekundaariverbeihin. Primääriverbien ja niitä vastaavien asiaintilojen välillä on rakenteellinen samankaltaisuus: kielenulkoinen asiointi sekä sitä kielentävä tyyppi vastaavat toisiaan. Primääri-A-verbeissä vastaavuus on suurempaa kuin primääri-B-verbeissä. Primääri-A-verbit ilmentävät tilaa (ASUA), tekoa (SIIVOTA), tapahtumaa (SYNTYÄ) tai liikettä (KÄVELLÄ). Primääri-B-verbit puolestaan koodaavat itse asiointilan lisäksi myös jonkin suhteen tai asenteen kyseessä olevaa asiointilaa kohtaan: ne ilmentävät kognitioita ja emootioita (AJATELLA, VIHATA), perkeptiota (havaitsemista; NÄHDÄ) sekä puheaktia ja kielellistä kommunikaatiota (PUHUA). Sekundaariverbit taas eivät varsinaisesti koodaa itse asiointiloja, vaan niiden avulla asiointiloja kvalifioidaan tai suhteutetaan toisiinsa. Sekundaariverbejä käyt-

tetään usein infinitiivien kanssa (esim. *alkaa itkeä*, *ruveta satamaan*, *saattaa antaa*). (Pajunen 2001: 51–56.)

Tämän tutkielman semanttisen ryhmittelyn ’tila’-luokka vastaa Pajusen verbiluokituksen tilaverbien ryhmää. Pajusen luokittelussa tilaverbit ovat vastakkaisia prosessia ilmaiseville verbeille: teko-, tapahtuma- ja liikeverbeille. Tilaverbeihin ei liity kausatiota, vaan ne ilmaisevat olioiden välisiä suhteita, kuten tilaa tai omistusta. (Pajunen 2001: 94–96.) Suurin osa semanttisen preferenssin tarkastelualueen ’tila’-ryhmän sanoista on OLLA-verbejä: predikatiivilauseiden (”*Mustan siirto oli ehkä ylisuojausta*”) tai omistuslauseiden (”*oliko karhulla mahdollisesti rabies*”) predikaatteja tai infiniittimuotoisia OLLA-verbejä (”– – *selviäisi ilmassa mahdollisesti olevien sienitiöiden määrää*”). Apuverbinä toimivat OLLA-verbit sitä vastoin olen luokitellut sekundaariverbeiksi, jotka puolestaan olen jättänyt ’luokittelemattomien’ kategoriaan. OLLA-verbin lisäksi ’tila’-luokkaan kuuluu erilaisia muitakin staattista tilaa ja olemista ilmaisevia verbejä, kuten ESIINTYÄ (”– – *esiintyisi ehkä 30 ylimääräistä kilpirauhassyöpätapausta*”) ja PIILLÄ (”*siinä piilee ehkä syvin viehäytys*”), staattista tilaa ilmaisevasta ELÄÄ-verbistä johdettu substantiivi ELO sekä virheellisesti yhteen kirjoitettu, niin ikään olemista ilmaiseva sana JÄLJELLÄOLEVA. Lisäksi koska etenkin omistuslauseiden OLLA-verbin esiintymien avulla ilmaistaan habitiivisuutta, olen luokitellut myös habitiivisuutta ilmaisevan adjektiivin OMA tähän luokkaan. Se ei tosin ilmennä tilaa samalla tavalla kuin muut luokan sanat, ja luokan nimeksi kenties sopisikin paremmin ’tila, habitiivisuus’. Toisaalta habitiivisuutta niin ikään ilmaisevat verbit SAADA ja ANTAA kuuluvat luokittelussani muutosta koodaavina verbeinä ’prosessi’-merkitysryhmään, eivät ’tila’-luokkaan, joten olen tyytynyt tiivistämään luokan nimen muotoon ’tila’.

’Prosessi’-luokka perustuu niin ikään Pajusen verbiluokitukseen, ja luokkaan kuuluukin erityisesti verbejä (SIIRTYÄ, TEROITTA, MYYDÄ, KUIVUA), mutta toisaalta myös sellaisia verbeistä johdettuja substantiiveja, joista on tulkittavissa prosessin merkitys (KATOAMINEN, TOILAILU). Sen sijaan esimerkiksi verbeistä johdettu substantiivi TUTKIMUSSAAVUTUS kuuluu tässä ryhmittelyssä ’luokittelemattomiin’, sillä sen voidaan ajatella ilmentävän pikemminkin prosessin päätepistettä kuin itse prosessia. ’Prosessi’-kategoriaan kuuluvat verbit ovat niitä verbejä, jotka Pajusen verbiluokituksessa kuuluvat tapahtumaverbeihin, liikeverbeihin ja teko-verbeihin; näiden taustalla on aina jonkinlaisia kausatiota (Pajunen 2001: 120). Mentaalista prosessia ilmaisevat sanat kuitenkin kuuluvat jaotelussa ’mentaalisuus’-luokkaan, kuten ne Pajusen luokittelussakin kuuluvat mentaaliverbeihin. Esimerkiksi KIINNOSTUA ja HYVÄKSYÄ ovat tällaisia sanoja. Lisäksi monet aineiston verbit ja verbeistä johdetut substantiivit ovat sellaisia, että ne ilmaisevat paitsi prosessia, myös jotakin muuta merkityspiirrettä: esimerkiksi LYHETÄ liittyy myös ’määrään’ ja YHTIÖITTÄÄ ’yhteiskunnallisuuteen’. En ole luokitellut näitä ’prosessi’-ryhmään, vaan aina kulloiseenkin toiseen vaihtoehtoon,

esimerkiksi siis 'määrä, intensiteetti'- tai 'yhteiskunnallisuus, institutionaalisuus' -luokkiin, sillä nämä ovat merkitykseltään kapea-alaisempia luokkia kuin 'prosessi'. Kyseessä on siis merkitykseltään laajojen luokkien toissijaisuuden periaate, joka esiteltiin luvussa 3.3.4.

'Mentaalisuus'- sekä 'viestintä'-luokka pohjautuvat molemmat edellä mainittuihin Pajusen verbiluokittelun primääri-B-verbeihin. Tällaiset verbit ovat mentaaliverbejä, ja niillä koodataan referentiaalisiin asiointiloihin kohdistuvaa tietoa, sen lähdettä ja varmuusastetta. Ero primääri-A-verbeihin on siinä, että A-verbeillä leksikalisoidaan asiointiloja mutta B-verbeillä puolestaan leksikalisoidaan sitä, miten ihminen kokee tai tulkitsee nämä A-verbeillä leksikalisoidut asiointilat. Primääri-B-verbit koodaavat siis kognitioita ja emootioita, perkeptiota (aistein havaitsemista) sekä puheaktia ja kielellistä kommunikaatiota. (Pajunen 2001: 56, 296.) Luokittelussani 'mentaalisuus'-luokkaan kuuluvat ne verbit, jotka ilmaisevat kognitioita (YMMÄRTÄÄ, HYVÄKSYÄ), emootioita (HARMITTAA, MASENTAA) ja perkeptioita (NÄHDÄ, KATSOA). Puheaktia ja kielellistä kommunikaatiota ilmentävät verbit, kuten ILMAISTA, SANOA ja VASTATA, puolestaan kuuluvat semanttisessa jaottelussani 'viestintä'-luokkaan. Mentaaliset verbit ARVELLA (*on arvellut*), LUULLA (*luullakseni*) ja EPÄILLÄ (*epäillään*) sopisivat merkityksensä puolesta 'epävarmuus'-luokkaan, mutta olen kuitenkin luokitellut ne 'mentaalisuus'-luokkaan, sillä nähdäkseni on tarkoituksenmukaista ryhmitellä sekä tällaiset kognitioihin viittaavat sanat että emootioihin viittaavat sanat keskenään samaan luokkaan.

'Mentaalisuus'-luokkaan kuuluu verbien lisäksi muihin sanaluokkiin kuuluvia sanoja, joiden keskeinen semanttinen piirre on mentaalisiin toimintoihin liittyvä. Esimerkiksi KUVA merkityksessä 'mielikuva', VÄÄRINYMMÄRRYS, HENKISESTI ja TYYTYMÄTÖN ovat tällaisia sanoja. Luokkaan kuuluu ainoastaan mielensisäisiä toimintoja ja ilmiöitä kuvaavia sanoja, eikä lainkaan sellaisia sanoja, jotka esimerkiksi kuvaavat aistein havaittavia ominaisuuksia tai asioita, vaikka tällaiset sanat osaltaan liittyvätkin perkeptioon. Siten esimerkiksi KAIKU- ja MUSTA-sanoja en ole luokitellut tähän luokkaan; KAIKU on 'viestintä'-luokassa ja MUSTA 'arvo, laatu'-luokassa. HALUTA-verbi muistuttaa kielipölyisesti sekundaariverbejä, sillä sitä käytetään infinitiivien kanssa, mutta mentaaliseen toimintoon viittaavana verbinä olen luokitellut sen 'mentaalisuus'-luokkaan.

Kuten edellä mainittiin, 'viestintä'-luokkaan kuuluu puheaktia ja kielellistä kommunikaatiota ilmentäviä verbejä, kuten ILMAISTA, SANOA ja VASTATA. Lisäksi luokassa on kommunikointiin ja viestintään liittyviä substantiiveja (SITAATTI, MEDIA, ILMOITUS, ANOMUS, OHJE). Etenkin monen substantiivin osalta raja 'viestintä'- ja 'yhteiskunta'-luokkien välillä on epätarkka: moni sana olisi sijoitettavissa molempiin luokkiin. Olen arvioinut kunkin yksittäisen sanan erikseen: esimerkiksi RAUHANNEUVOTTELU ja NEUVOTTELU kuuluvat luokittelussani luokkaan 'yhteiskunnallisuus, institutionaalisuus', sillä arvioin niiden kontekstissaan ilmentävän erityisesti yhteiskunnallista toimintoa ja vasta toissijaisesti viestintää. Luokittelu ei ole ehdoton, ja muutkin ratkaisut olisivat perusteltavissa.

Myös substantiivin KAIKU olen sijoittanut 'viestintä'-luokkaan, vaikka siinä kyse onkin hieman erilaisesta viestinnästä kuin suoraan ihmiseltä ihmiselle tapahtuvasta.

'Aika'-luokkaan kuuluu sanoja, jotka ilmentävät ajankohtaa (2003, SILLOIN, TIISTAI, NYT) ja muita ajankohtaan tai ajan kulumiseen liittyviä asiointiloja (PITKÄJÄNTEINEN, JOUTAA, LYKÄTÄ). 'Aika'-luokka on ilmeinen jo aineiston pohjalta, mutta toisaalta tämän kategorian luokitteluun ottaessani olen noudattanut samalla Jantusen luokittelua (2004: 245–246), jossa on yhtenä semanttisena ryhmänä 'aika, ikä'. Luokittelussani olen kuitenkin nimennyt ryhmän 'aika'-nimiseksi, sillä ikä ei ole olennainen piirre tämän ryhmän sanoilla. Esimerkiksi ryhmän ainoa VANHA-lekseemin esiintymä on lauseessa ”*Yleistetään ehkä hyvinkin vanhojen luutumien varassa*”, ja näin ollen kontekstissaan LUUTUMA-substantiivin määritteenä toimivan adjektiivin ei voida ajatella kuvaavan ikää. Substantiivi NUORI sitä vastoin kyllä liittyy merkitykseltään ikään, mutta ihmiseen viittaavana sanana se kuuluu ryhmittelyssä 'ihmiset'-luokkaan. *Pieniä lapsia* -ilmauksen adjektiivin PIENI olen puolestaan luokitellut 'aika'-luokkaan.

Merkitysryhmään 'paikka' kuuluu fyysistä paikkaa tai sijaintia merkitseviä sanoja (TÄÄLLÄ, ASUNTO, TEHDASALUE). Tässä luokassa on esimerkiksi fyysiseen paikkaan viittaavia paikannimiä, kuten seuraavan lauseen paikannimet LIEKSA ja SAARIJÄRVI: ” – – *kaiut Lieksan Saarijärvellä mahdollisesti oleilevasta idänkäestä tulivat lintupiippariin* – –”. Kuitenkin sellaiset paikannimet, joiden katson kontekstissaan viittaavan ensisijaisesti poliittiseen toimijaan fyysisen paikan sijaan, olen luokitellut myöhemmänä esiteltävään 'yhteiskunnallisuus, institutionaalisuus' -luokkaan. Esimerkiksi lauseen ” – – *ilmaista kantansa Kiinalle mahdollisesti yhdessä muiden Pohjoismaiden kanssa*” paikannimet KIINA ja POHJOISMAAT olen luokitellut poliittisia toimijoita ilmaisevina sanoina 'yhteiskunnallisuus, institutionaalisuus' -luokkaan 'paikka'-luokan sijasta. Vastaavasti sanan HAKIJA-KAUPUNKI olen luokitellut samoin, sillä se merkitsee pikemminkin yhteiskunnallista toimijaa kuin fyysistä sijaintia. Eräänlaista abstraktia sijaintia merkitsevät sanat *parissa, piirissä, [syytettyjen] penkille* ja *taustalla* puolestaan olen luokitellut muihin luokkiin – esimerkiksi viimeiseksi mainitun 'kausaalisuus'-kategoriaan –, sillä 'paikka'-kategoriassa on kyse nimenomaan fyysisestä sijainnista.

'Määrä, intensiteetti' -luokkaan kuuluu nimensä mukaisesti määrää ilmaisevia sanoja (VÄHÄN, MUUTAMA, OSITTAIN). Myös erilaiset intensiteettiä (ERITTÄIN, RIITTÄVÄSTI) ja ulottuvuutta (SYVÄ, PIENI) ilmaisevat sanat kuuluvat tähän luokkaan. Nämä sanat olisi mahdollista jaotella myös erillisiksi kategorioikseen 'määrä', 'intensiteetti' ja 'ulottuvuus', kuten Jantunen on luokittelussaan (2004: 245–246) tehnyt. Kaikilla näillä merkitysryhmillä on kuitenkin keskinäinen yhtäläisyys, jonka perusteella ne ovat ryhmiteltävissä samaan kategoriaan: kaikki tällaiset sanat ilmaisevat eräänlaista määrään tai suuruuden tai voimakkuuden asteeseen liittyvää ominaisuutta. Monet 'määrä, intensiteetti' -kategorian sanat ovat kuitenkin merkitykseltään päällekkäisiä muiden semanttisten

ryhmien kanssa. Sanan AINAKIN olen sijoittanut 'epävarmuus'-luokkaan, vaikka sen toisaalta voi ajatella ilmaisevan myös 'määrää'. Lisäksi moni aineiston 'määrää' ilmaisevista sanoista, esimerkiksi adjektiivinen PIENI ("*pieniä korotuksia voi mahdollisesti tulla*") olisi luokiteltavissa myös 'arvo, laatu' -luokkaan. 'Arvo, laatu' -luokkaan liittyvän toissijaisuuden periaatteen mukaisesti olen kuitenkin luokitellut tällaiset sanat 'määrä, intensiteetti' -kategoriaan. 'Määrä, intensiteetti' -kategoria on osin päällekkäinen myös 'additiivisuus' -luokan kanssa, kuten jäljempänä mainitaan.

'Ihmiset'-luokkaan olen luokitellut useimmat sanat, joiden referentti on ihminen. Esimerkiksi MIES, HÄN ja HENKILÖ kuuluvat tähän luokkaan, samoin ihmisiin viittaavat erisnimet, kuten RIIKKA. Kuitenkin Väinö Linnaan viittaavan sukunimen LINNA olen luokitellut ryhmään 'kulttuuri' aiempaan mainituin, metonymisyyteen liittyvin perustein. Sana LEIJONA (*leijonien*) puolestaan kuuluu luokkaan 'urheilu, kilpailu'. Yksittäisen 'leijonan' voidaan toki nähdä viittaavan yhteen Suomen maajoukkueen jääkiekkoilijaan, mutta monikkomuotoisena sanalla on kontekstissaan kollektiivinen, urheilujoukkueeseen viittaava merkitys, jolle 'urheilu, kilpailu' -luokka on tulkintani mukaan oikeampi kategoria kuin 'ihmiset'. Sitä vastoin jääkiekkoilijoiden nimet JERE LEHTINEN, JUHA LIND ja ESA TIKKANEN olen puolestaan luokitellut 'ihmiset'-luokkaan, vaikka kyseessä olevat henkilöt ovatkin ammatiltaan juuri urheilijoita. Substantiivit ÄÄNESTÄJÄ, ASIANOSAINEN, PÄÄMINISTERI ja kollektiivista, yhteiskunnallista merkitystä ilmentävä substantiivi VÄESTÖ puolestaan ilmaisevat kontekstissaan nähdäkseni vahvemmin yhteiskunnallisuuden kuin ihmisen merkitystä, joten olen sijoittanut ne luokkaan 'yhteiskunnallisuus, institutionaalisuus', vaikka myös 'ihmisten' luokkaan sijoittaminen olisi perusteltavissa. Kontekstissaan ihmisiin viittaavaa pronomini JOKA ei jaotellussa kuulu 'ihmiset'-luokkaan, vaan 'luokittelemattomiin', sillä lekseemillä itsessään ei ole ihmiseen viittaavaa merkitystä.

'Terveystila'-luokka sisältää terveyteen, sairauteen ja vammaisuuteen viittaavia sanoja, kuten TAUTI, LÄÄKE, VAMMA ja TERVE. Luokka on hyvin pieni: siihen kuuluu vain joitakin sanoja kummankin modaalisanan kontekstissa. Kategorian sanat olisi ollut mahdollista myös jättää 'luokittelemattomat'-luokkaan, mutta ne ovat silti erillisenä ryhmänään, sillä ne muodostavat yhtenäisen ryhmän merkityksensä puolesta. Nähdäkseni ei olisi tarkoituksenmukaista myöskään sijoittaa terveydentilaan viittaavia sanoja mihinkään muuhun luokkaan, sillä niiden merkitys ei ole kovin lähellä minkään muun luokan sanoja.

'Urheilu, kilpailu' -luokkaan kuuluu erilaisia urheiluun liittyviä sanoja (OLYMPIAKISAT, MAAJOUKKUEVALINTATAPA), mutta myös muunlaisia kilpailuun viittaavia sanoja (shakin pelaamiseen liittyvät SIIRTO ja YLISUOJAUS ja kontekstissaan pilkkikisaan viittaava TASAPELI). Substantiivi SAUVARINNE olisi 'urheilu, kilpailu' -luokan lisäksi luokiteltavissa myös 'paikka'-kategoriaan, mutta sana viittaa tulkintani mukaan vahvemmin urheilusuoritukseen kuin varsinaiseen paikkaan: "*kan-*

*naisi [kannattaisi] ehkä painua mielummin **sauvarinteeseen** kuin sujutella tavallinen kahden tunnin harjoitus*". Sana IZVESTIJA puolestaan viittaa venäläisen sanomalehden nimeen, mutta kontekstissaan se tarkoittaa kuitenkin lehden mukaan nimensä saanutta jääkiekkoturnausta ja kuuluu siten 'urheilu, kilpailu' -kategoriaan. JUOSTA- ja SINKAUTTAA-verbit samoin kuin urheiluun liittyvät, processiverbeistä johdetut substantiivit, kuten PUTOAMINEN, olen sijoittanut 'urheilu, kilpailu' -luokkaan aiempina kuvatuin, 'prosessi'-luokan toissijaisuuteen pohjautuvin perustein.

'Konkreettiset oliot' -luokka muodostuu sanoista, jotka viittaavat konkreettisiin, ei-inhimillisiin olioihin (KUORMA-AUTO, SÄHKÖKIUAS, HORNET). Myös eläimet olen luokitellut tähän luokkaan (HIRVI, KARHU, IDÄNKÄKI). Ryhmän sanat ovatkin merkitykseltään monilta osin varsin kaukana toisistaan, mutta niitä kuitenkin yhdistää se, että kyse on jostakin konkreettisesta entiteetistä, ei abstraktista. Olennaista onkin verrata luokan merkitystä muiden luokkien merkityksiin, ei ainoastaan tarkastella ryhmän sisäisiä yhtäläisyyksiä. Lähes kaikki sanat muissa luokissa ovat nimitäin merkitykseltään abstrakteja. Toki kuitenkin esimerkiksi 'urheilu, kilpailu' -luokan substantiivi SAUVARINNE ja 'paikka'-luokan substantiivi TEHDASALUE ovat merkitykseltään konkreettisia, mutta ne kuitenkin ilmentävät nähdäkseni ensisijaisesti 'urheilua' ja 'paikkaa', joten olen luokitellut ne konkreettisesta tarkoitteestaan huolimatta näihin kategorioihin. Vielä monitulkintaisempina tapauksina voidaan pitää sanoja HORNET, HÄVITTÄJÄLENTÄJÄ ja (virheellisesti erilleen kirjoitettu) SB2 POMMITUSKONE, sillä ne olisivat maanpuolustukseen viittaavina sanoina luokiteltavissa myös 'yhteiskunnallisuuteen, institutionaalisuuteen'. Olen kuitenkin luokitellut ne 'konkreettiseksi olioiksi', sillä 'yhteiskunnallisuus, institutionaalisuus' -luokan sanoille on ominaista nimenomaan abstrakti, ei konkreettinen merkitys.

'Kulttuuri'-luokkaan olen ryhmitellyt kirjallisuuteen, musiikkiin ja muuhun taiteeseen sekä kulttuuriperinnön vaalimiseen viittaavia sanoja (KUORO, ROMAANI, MUSEAALINEN, PERINNEMAIMA). 'Kulttuuri'-luokka on merkitykseltään osittain päällekkäinen 'viestintä'-luokan kanssa: kirjallisuuden lajien nimet ROMAANI ja SATU liittyvät myös viestintään, mutta olen kuitenkin luokitellut ne 'kulttuuri'-luokkaan, sillä nähdäkseni ne ilmaisevat ennen kaikkea kulttuurituotteita eivätkä yhtä vahvasti kommunikaatiota ja viestintää. Uskonnolliseen yhteisöön viittaavan adjektiivin ORTODOKSINEN olisi puolestaan voinut luokitella 'yhteiskunnallisuus, institutionaalisuus' -luokkaan, mutta koska se kontekstissaan (*Tiettävästi saarella on ollut kalmisto ja ehkä **ortodoksinen** tsasouna*) viittaa nähdäkseni pikemminkin kulttuurisen perinnön vaalimiseen, se kuuluu jaotellussa kategoriaan 'kulttuuri'.

Luokkaan 'yhteiskunnallisuus, institutionaalisuus' kuuluu sellaisia sanoja, jotka viittaavat esimerkiksi poliittiseen toimintaan (KESKUSTALAINEN, PUOLUSTUSPOLITIIKKA), lakiin ja oikeuteen (SYYTETTY, LAILLINEN), talouteen ja työelämään (PÖRSSI, INFLAATIO-ODOTUS, TYÖ), maanpuolus-

tukseen (RINTAMATEHTÄVÄ) ja erilaisiin yhteiskunnallisiin instituutioihin (KOULU, KIINTEISTÖLAITOS). Luokkaan kuuluu siis sanoja melko laajan merkityskentän alueelta, mutta kaikissa niissä on kyse jonkinlaisesta yhteiskunnallisesta tai institutionaalisesta toiminnasta. Voidaan pohtia, voisiko tämän laajan luokan jakaa useaan, esimerkiksi edellä mainitun mukaiseen, merkitykseltään yksityiskohtaisempaan luokkaan. Joistain kategorioista muodostuisi kuitenkin silloin kooltaan varsin pieniä, joten olen päätenyt ottamaan luokitteluun laajan semanttisen luokan 'yhteiskunnallisuus, institutionaalisuus'. Verbit YHTIÖITTÄÄ ja LEIKATA sekä jotkin prosessiverbeistä johdetut substantiivit, kuten KIRISTÄMINEN, olisivat luokiteltavissa 'prosessi'-kategoriaan, mutta 'prosessi'-luokan toissijaisuuden periaatteen vuoksi olen sijoittanut ne 'yhteiskunnallisuus, institutionaalisuus' -luokkaan. Jotkin sanat, esimerkiksi RAUHANSOPIMUS ja YHTEISTYÖSOPIMUS, olisivat puolestaan luokiteltavissa myös 'viestintä'-kategoriaan, mutta katson niiden kontekstissaan kuitenkin ilmentävän ensisijaisesti poliittista ja yhteiskunnallista entiteettiä enkä niinkään viestintää. Lisäksi esimerkiksi sellaiset maiden nimet, jotka viittaavat maihin poliittisina toimijoina, olen luokitellut tähän luokkaan, kuten jo aiemmin todettiin.

'Epävarmuus'-luokan sanat ilmaisevat episteemistä modaalisuutta (VOIDA, SAATTAA), useiden vaihtoehtojen olemassaoloa (TAI), indefiniittisyyttä (JOKU, JOKIN), konditionaalisuutta (JOS, MIKÄLI) sekä muuta epävarmuutta ja epämääräisyyttä. Lauseessa ” – – *pieniä korotuksia voi mahdollisesti tulla*” modaalisen MAHDOLLISESTI-adverbin lisäksi myös modaaliverbi VOIDA ilmaisee episteemistä modaalisuutta, puhujan uskoa tai tietoa asiointilan mahdollisuudesta tai varmuudesta (ks. episteemisestä modaalisuudesta luku 2.4.1). Kyse on modaaliharmoniasta, joka tuli esille jo luvussa 4.2.2 verbikollokaattien yhteydessä. Muut 'epävarmuus'-ryhmään kuuluvista sanoista kuin VOIDA- ja SAATTAA-verbit eivät kuitenkaan ole varsinaisia modaalisanoja. Niissäkin kuitenkin on kyse eräänlaisesta epävarmuuden merkityksestä. Esimerkiksi konjunktiolla TAI ilmaistaan useiden vaihtoehtojen olemassaoloa: vaikkapa lauseessa ” – – *oliko karhulla rabies tai muita tauteja*” tuodaan TAI-konjunktiolla ilmi, että ”rabies” ja ”muut taudit” ovat keskenään vaihtoehtoja toisilleen kyseessä olevassa tilanteessa. Edelleen indefiniittiset pronominit JOKU ja JOKIN ilmaisevat epämääräisyyttä eli eräänlaista epävarmuutta. Esimerkiksi ilmaisussa ”*ehkä kuitenkin jossain välissä – –*” tuodaan esille, että ”väli” eli ajankohta, jossa puheena oleva asia tapahtuu, on toistaiseksi epävarma. Konjunktiolla JOS ja MIKÄLI puolestaan ilmaistaan konditionaalista suhdetta. Konditionaalinen suhde on sellainen lauseiden välinen suhde, jossa toisella lauseella esitetään ehto, jonka pitäessä paikkansa toisella lauseella ilmaistava asiointila toteutuu tai on mahdollinen (VISK määritelmät s.v. konditionaalinen suhde). Myös MIKSEI-sanana olen luokitellut 'epävarmuus'-kategoriaan, sillä seuraavassa lauseessa sana implikoi nähdäkseni jonkinlaista epävarmuutta sen suhteen, mitä sanotaan tyttö- ja poikahiihtäjistä: ”*Tässä vinkki tyttöhihtäjille, tai miksei pojillekin – –*”. Tosin osuvampi semanttinen

katteoria sanalle olisi 'myönnytyt', mutta sellaista katteoriaa luokittelussani ei ole. Niinpä olen päätyntyt sijoittamaan sanan 'epävarmuuden' luokkaan. Myös heikkoon mahdollisuuteen (KS s.v. *oljenkorsi*) viittaava substantiivi OLJENKORSI kuuluu luokittelussani 'epävarmuuden' katteoriaan, sillä käytettävissä olevista merkitysluokista se on nähdäkseni sille osuvin.

'Additiivisuus'-ryhmä pohjautuu käsitteeseen "additiivinen", joka on Ison suomen kieliopin mukaan sellainen "lauseiden, lausekkeiden, sanojen tai sananosien välinen suhde, jossa ilmaistaan, että toisiinsa kytketyillä elementeillä on esitysyhteydessään jotain yhteistä ja että ne pitävät yhtä lailla paikkansa" (VISK määritelmät s.v. additiivinen). Tähän luokkaan kuuluu useita kertoja aineistossa esiintyvä konjunktio JA, mutta myös muita additiivisuutta ilmaisevia taipumattomia adverbeja ja partikkeleita: YHDESSÄ, KANSSA, SEKÄ, LISÄKSI, MUUTEN, TAAS, MYÖS ja NIIN KUIN rakenteessa "niin X kuin Y:kin". Lisäksi olen luokitellut tähän luokkaan nominit YHTEINEN, MUU sekä TOINEN ("toisena vaihtoehtona"). Voidaan ajatella, että myös konjunktio TAI ilmentää eräänlaista additiivista suhdetta, mutta se kuitenkin kuuluu jaottelussani 'epävarmuus'-luokkaan, sillä nähdäkseni epävarmuuden merkitys on sanalla vahvempi. Luokat 'määrä, intensiteetti' ja 'additiivisuus' voidaan nähdä merkitykseltään osin päällekkäisinä, sillä osa additiivista suhdetta ilmentävistä sanoista, esimerkiksi konjunktio JA, ilmaisevat samalla myös eräänlaista entiteettien lisäystä ja sikäli määrää. Sanojen ryhmittelyssä on kuitenkin ajateltava luokkien sisäistä yhtenäisyyttä: luokka 'määrä, intensiteetti' voisi kasvaa merkitykseltään liian laajaksi, jos myös additiivista suhdetta ilmentävät sanat lisättäisiin siihen, sillä jo nyt siihen kuuluu keskenään hyvin erilaisia sanoja. Niinpä 'additiivisuus' ja 'määrä, intensiteetti' ovat ryhmittelyssä erillisiä luokkia.

'Kausaalisuus'-luokka perustuu 'additiivisuus'-luokan tavoin Ison suomen kieliopin jaotteluun lauseiden välisistä semanttisista suhteista (VISK § 1121); luokan taustalla on kausaalisen suhteen käsite. Kausaalinen suhde on lauseiden välinen suhde, jossa esitetään, että kyseessä olevat asiointilat ovat keskenään syyn ja seurauksen suhteessa (VISK määritelmät s.v. kausaalinen suhde). Täten luokan jäsenet ovat lähtökohtaisesti konjunktioita ja muita taipumattomia sanoja; sanat JOTTA, JOTEN, KOSKA, VARTEN, MYÖTÄ, KUN, TIIMOILTA ja SIKSI kuuluvat tähän luokkaan. Lisäksi olen luokitellut substantiivit SYY, TAUSTA (*taustalla*) ja AIKA (*saavat – – aikaan*) sekä verbit AIHEUTUA, AIHEUTTAA, VAIKUTTAA ja TUODA kausaalisuuden luokkaan. Kukin näistä sanaesiintymistä ilmentää kontekstissaan syy–seuraus-suhdetta, esimerkiksi verbi AIHEUTTAA seuraavassa ilmaisussa: "sähkökiuas, jonka epäillään mahdollisesti **aiheuttaneen** saunassa kuivumassa olleiden vaatteiden syttymisen".

Niin ikään 'vastakkaisuus'-luokka perustuu Ison suomen kieliopin (VISK) esittämään ryhmittelyyn lauseiden välisistä merkityssuhteista. Luokkaan kuuluu kontrastiivista ja konsessiivista suhdetta ilmentäviä sanoja. Kontrastiivinen suhde on lauseiden tai lausekkeiden välinen suhde, jossa

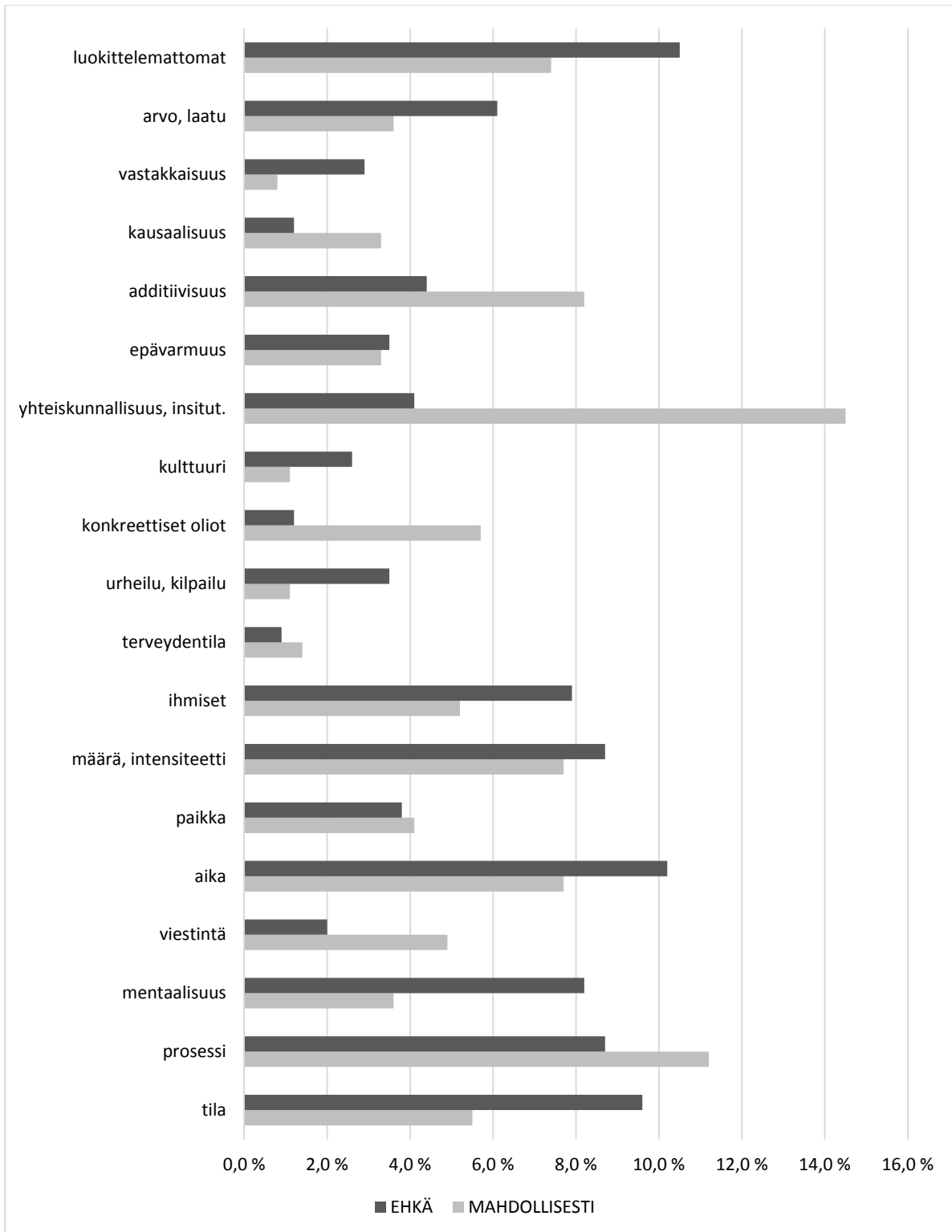
elementit esitetään toisilleen jollakin tavalla vastakkaisina; kaksi asiaintilaa voidaan asettaa vastakkain esimerkiksi siten, että ensin kokonaan tai osin kielletään jokin asiaintila ja sitten tarjotaan tilalle pätevämpi vaihtoehto (VISK määritelmät s.v. kontrastiivinen suhde, § 1105). Esimerkiksi MUTTA-konjunktion avulla esitetään seuraavassa lauseessa tilalle ensin mainittua pätevämpi vaihtoehto: ”Suomesta ei sopivia kouluja löydy, **mutta** mahdollisesti joskus hänen kannattaa hakea oppia ulkomailta – –”. Konjunktio MUTTA voi toisaalta ilmentää myös konsessiivista eli odotuksenvastaisuutta ilmaisevaa suhdetta, samoin kuin partikkelit SITTEINKIN, SITTENKÄÄN, KUITENKIN ja KUITENKAAN (VISK § 1121, 1139, 1142). Esimerkiksi lauseessa ”Ehkä on **sittenkin** parempi mennä Turkuun” toteutuu konsessiivinen suhde. Siinä implikoidaan, että Turkuun menemättä jättäminen olisi odotuksenmukaista, ja SITTEINKIN-partikkelilla ilmaistaan, että Turkuun meneminen on odotuksenvastaista.

Luokkaan ’arvo, laatu’ kuuluu arvoon eli asioiden hyvyyteen tai huonouteen liittyviä sanoja, kuten HYVÄ, KANNATTAA, MIELUUMMIN ja HAITTA, mutta myös erilaisia muita asioiden laadullisiin ominaisuuksiin viittaavia sanoja, kuten MUSTA, REPÄISEVÄ ja TEKNINEN. Olisi mahdollista muodostaa hyvyyteen ja huonouteen liittyvistä sanoista erillinen semanttinen luokka ’arvo’, joka onkin Jantusen tutkimuksessa (2004) semanttisen preferenssin yhtenä luokkana, ja muista laatuun viittaavista ominaisuuksista puolestaan toinen luokka, ’laatu’. Semanttisessa luokittelussa on kuitenkin syytä kiinnittää huomiota myös merkitysluokkien kokonaismäärään: jotta analyysi olisi paremmin hallittavissa, olen sijoittanut keskenään riittävässä määrin samankaltaiset ’arvo’- ja ’laatu’-ryhmät samaan luokkaan, etenkin kun ’laatu’-luokkaan yksinään ei kuuluisi montaakaan sanaa. Tämän luokan sanat ovat pääasiassa adjektiiveja, mutta myös esimerkiksi VOIDA-verbin yksi esiintymä kuuluu tähän luokkaan. Lauseessa ”Ehkä enemmän pelejä **vois olla**” VOIDA-verbi ei nähdäkseni nimittäin ilmennä epistemistä modaalisuutta eikä siten ole luokiteltavissa ’epävarmuus’-luokkaan, vaan se ilmaisee pikemminkin deonttista modaalisuutta ja puhujan toivetta, joka on eräänlainen arvon ilmaisu. Myös verbi KANNATTAA kuuluu ’arvo, laatu’ -luokkaan, sillä sen voidaan ajatella ilmaisevan hyvyyttä, hyvää vaihtoehtoa ja siten arvoa. Toisaalta on huomattava, että monet muukin aineiston sanoista kuin ’arvo, laatu’ -luokkaan ryhmitellyt sanat ilmaisevat arvoa tai laatua. Esimerkiksi ’terveydentila’-luokkaan kuuluvat sanat SAIRAS ja TAUTI ilmentävät ei-arvostettavaa, negatiivista asiaa, monet ’mentaalisuus’-luokan sanoista ilmentävät tunteita (SURULLINEN, VIEHÄTYS) ja siten arvoa, ja edelleen esimerkiksi TERÄVÄ ilmauksessa ”*terävimmän vauhdin*” ilmentää positiiviseksi miellettyä asiaa. Voidaan itse asiassa ajatella, että nämä merkityspiirteet liittyvät pikemminkin semanttiseen prosodiaan kuin semanttiseen preferenssiin. Joka tapauksessa, kuten jo edellä mainittiin, tässä tutkielmassa ’arvo, laatu’ -luokkaan kuuluu vain sellaisia arvoa ja laatua ilmaisevia sanoja, joita ei voida sijoittaa muihin merkitysluokkiin.

Tarkastelualueen sanojen luokittelu ei ole aukotonta, joten osa sanoista jää luokittelematta varsinaisiin merkitysluokkiin ja kuuluu siten 'luokittelemattomat'-ryhmään. Tällaisia ovat esimerkiksi sanat, jotka Pajusen (2001) verbikategorioinnissa ovat sekundaariverbejä. Sekundaariverbejä tämän työn aineistossa ovat perfektin ja pluskvamperfektin apuverbinä käytetty OLLA-verbi (OLLA-lekseemin muiden esiintymien ryhmittelystä ks. luokan 'tila' kuvailu s. 74), kieltoverbi EI sekä seuraavat apuverbit tai niiden kaltaiset verbit: *alkavat mellastaa*, *tulee käytettyä*, *saa lykkäämään*, *voidaan estää* ja *tulisi laatia*. Näistä VOIDA-verbi ei tulkintani mukaan edusta tässä ensisijaisesti episteemistä, vaan dynaamista modaalisuutta, 'pystymistä' (*Jos huumeisiin puututaan nyt heti, tilanteen ryöstäytyminen käsistä voidaan ehkä estää*). Luokittelemattomien ryhmään kuuluu lisäksi muita sanoja, jotka eivät nähdäkseni merkitykseltään sovi mihinkään edellä luetelluista ryhmistä, esimerkiksi pronomineja (JOKA, SE) ja kontekstissaan abstrakteja substantiiveja (ASIA, TAPAUS).

5.2 Semanttisen preferenssin analyysin tulokset

Kuviossa 1 on kuvattu modaalisanon semanttisen preferenssin analyysin tulokset. Kuvioista nähdään, kuinka suuri osuus kullakin semanttisella ryhmällä on kaikista modaalisanon kontekstissa tarkastelualueella esiintyvistä sanoista. Siitä havaitaan ensinnäkin, että modaalisanon kollokaatit jakautuvat suhteellisen tasaisesti eri semanttisiin luokkiin sikäli, että mikään semanttinen ryhmä ei saa esimerkiksi kymmenien prosenttien osuutta kummankaan sanan lähikontekstissa. Suurin yksittäinen prosentuaalinen osuus on MAHDOLLISESTI-sanon 'yhteiskunnallisuus, institutionaalisuus' -luokan kollokaattien osuus sanan kaikista kollokaateista; luokan osuus on 14,5 %. Tämä on pienempi kuin suurimman yksittäisen semanttisen luokan osuus esimerkiksi Jantusen tutkimuksessa, jossa OIKEIN-määritteellä on yksi erittäin voimakas semanttinen preferenssi muihin astemääritteisiin nähden: merkitysluokan 'arvo' osuus kontekstin koko merkityskirjosta on jopa 43,8 % (Jantunen 2004: 138–139). Vaikka modaalisanoilla ei siis ole erityisen vahvoja semanttisia preferenssejä, kuvioista kuitenkin nähdään, että kummallakin sanalla toiset merkitysluokat ovat suurempia kuin toiset. Tarkastelen seuraavaksi, mitkä EHKÄ-sanon merkitysluokista ovat suurimpia eli mihin semanttisiin ryhmiin EHKÄ-sanon kollokaatit osoittavat ainakin jonkinlaista preferenssiä, ja sitten tarkastelen vastaavasti samaa MAHDOLLISESTI-sanasta.



Kuvio 1. Semanttisten luokkien osuudet EHKÄ- ja MAHDOLLISESTI-sanojen kontekstissa.

EHKÄ-partikkelilla on kontekstissaan paljon sanoja, jotka jäävät luokittelematta ('luokittelemattomat'-ryhmästä ks. s. 85), mutta sen kontekstissa suurin varsinainen semanttinen ryhmä on 'aika' (10,2 %, n=35). Täten jo tilastollisesti merkitsevien kollokaattien analyysissä tehty havainto, että sana suosii kontekstissaan aikaa merkitseviä sanoja, saa vahvistuksen. Muun muassa sanat ENSI, JO ja

JOSKUS ovat EHKÄ-sanan aikaa merkitseviä kollokaatteja. Toiseksi suurin merkitysryhmä EHKÄ-sanan kontekstissa on 'tila' (9,6 %, n=33), ja kolmanneksi suurimmat ovat 'prosessi' ja 'määrä, intensiteetti' (8,7%, n=30). 'Tila'- ja 'prosessi'-ryhmissä suurin osa sanoista on verbejä, joten erilaiset tilaa ja prosessia ilmaisevat verbit ovat siis tyypillisiä EHKÄ-sanan kollokaatteja; verbikollokaattianalyysissäkin havaittiin, että verbikollokaattien määrä sanalla on varsin suuri. Näistä ryhmistä luokkaa 'tila' voidaan myös pitää EHKÄ-sanalle tyypillisenä kontekstipiirteenä synonyymiinsa MAHDOLLISESTI nähden, sillä MAHDOLLISESTI-sanan kontekstissa tilaan viittaavia sanoja on vain 5,5 % koko merkityskirjosta (n=20).

Yhdessä 'prosessi'-luokan kanssa kolmanneksi suurin ryhmä on siis 'määrä, intensiteetti' (8,7%, n=30). Jo kollokaattianalyysissä adverbit (mm. HIEMAN, VÄHÄN, LIIAN), substantiivi AAVISTUS (*aavistuksen*) sekä numeraalit viittasivat tähän preferenssiin. Seuraavaksi eniten EHKÄ-sanan kontekstissa esiintyy 'mentaalisuuteen', 'ihmisiin' ja 'arvoon, laatuun' viittaavia sanoja, ja nämä kaikki preferenssit olivat havaittavissa jollakin tavalla jo kollokaattianalyysissä. 'Mentaalisuutta' edustavat sanat kattavat 8,2 % EHKÄ-sanan kontekstin koko merkityskirjosta (n=28), ja merkitysryhmä nousi esiin kollokaattianalyysissä siten, että sanan verbikollokaateista suuri osa osoittautui mentaaliseksi verbeiksi. Merkitysryhmä 'ihmiset' puolestaan kattaa 7,9 % kaikista merkityksistä (n=27). Se ei ollut varsinaisesti oletuksena semanttiseksi preferenssiksi kollokaattianalyysin pohjalta, mutta pronominkollokaatit ME, MINÄ ja HE kuitenkin kiinnittivät tilastollisesti merkitsevinä kollokaatteina huomion. Koska nämä persoonapronominit viittaavat ihmiseen, kollokaattianalyysin havainto näiden pronomien tilastollisesta merkitsevyydestä tukee tätä semanttisen preferenssin analyysissä havaittua 'ihmiset'-preferenssiä. Myös 'arvo, laatu' -merkitysluokan sanoja esiintyy melko usein EHKÄ-sanan kontekstissa. Luokka kattaa koko merkityskirjosta 6,1 % (n=21). Preferenssi oli havaittavissa kollokaattianalyysissä: adjektiiveja esiintyi EHKÄ-sanan kontekstissa runsaasti erityisesti sen synonyymiin verrattuna, ja lisäksi adverbianalyysissä esiin nousseen preferenssin 'tapa' voidaan katsoa osittain merkitykseltään kuuluvan 'arvo, laatu'-ryhmään.

Kaiken kaikkiaan kollokaattianalyysissä tehdyt hypoteesit siis saavat pääosin vahvistusta ja osoittautuvat tosiksi EHKÄ-sanan semanttisen preferenssin analyysissä. Kuitenkin yksi kollokaattianalyysissä esiin nousseista merkitysryhmistä ei osoittaudu semanttisessa analyysissä sanan kontekstin tärkeäksi merkitysryhmäksi: vaikka EHKÄ-sanan kontekstissa esiintyy useita tilastollisesti merkitseviä, vastakkaisuutta ilmaisevia partikkelikollokaatteja, (MUTTA, SITTEENKIN, KUITENKIN, LUKUUNOTTAMATTA, SENTÄÄN), semanttisista merkitysluokista 'vastakkaisuus' kattaa vain 2,9 % koko merkityskirjosta (n=10). Aivan pienin luokka se ei kuitenkaan ole. Kaikkein pienimmät semanttiset ryhmät EHKÄ-sanan kontekstissa ovat 'terveydentila' (0,9%, n=3), 'kausaalisuus' (1,2%, n=4) ja

'konkreettiset oliot' (1,2%, n=4), joten näihin luokkiin kuuluvat sanat eivät ole yleisiä sanan lähi-kontekstissa.

EHKÄ-sanalla on kotekstissaan myös paljon sanoja, joita ei ole luokiteltu mihinkään varsinaiseen semanttiseen ryhmään. Itse asiassa 'luokittelemattomat'-ryhmä osoittautuu suurimmaksi ryhmäksi noin 10,5 %:n (n=36) osuudella koko merkityskirjosta. Tätä voidaan pitää osin hieman ongelmallisena tuloksena, sillä olisi ideaalista, että analyysissä kaikki sanat pystyttäisiin sijoittamaan varsinaisiin semanttisiin ryhmiin eikä 'luokittelemattomat'-luokkaan jäisi yhtään sanaa. Sanoja ei kuitenkaan ole voitu väkisin sijoittaa merkitysluokkiin, joihin ne eivät luontevasti merkityksensä puolesta kuulu. 'Luokittelemattomat'-ryhmään kuuluu merkitykseltään lähinnä hyvin yleisiä ja erikoistumattomia sanoja, kuten kieliopillisia sanoja, esimerkiksi kieltosana EI ja pronominit SE ja JO-KA.

MAHDOLLISESTI-sana osoittaa melko selkeästi preferenssiä 'yhteiskunnallisuuteen, institutionaalisuuteen': 14,5 % (n=53) sen merkityskirjosta kuuluu tähän ryhmään. Tässä ryhmässä myös sanan ero synonyymiinsa on suurehko: EHKÄ-sanalla on vain 4,1 % sanoista viittaa yhteiskunnallisuuteen tai institutionaalisuuteen. Substantiivikollokaattien analyysissä nousi esiin, että 'rikokset' olisi tyypillinen merkitysryhmä sanan kotekstissa. Luokka 'yhteiskunnallisuus, institutionaalisuus' muodostuukin muun muassa rikoksiin ja oikeuslaitoksiin viittaavista sanoista – mutta kuitenkin myös muista, esimerkiksi talouselämän ja politiikan sanoista. Toiseksi eniten MAHDOLLISESTI-sanalla on 'prosessiin' viittaavia kollokaatteja (11,2 %, n=41). Nämä kollokaatit ovat siis lähinnä primääri-B-verbejä eli konkreettisia verbejä. Kollokaattianalyysissä havaittiin, että primääri-B-verbit ovat tyypillisiä sanan kotekstissa, ja arveltiin, että suurin osa kollokaateista olisi prosessia ilmaisevia verbejä ja että tilaa ilmaisevia olisi vähemmän. Täten semanttisen preferenssin analyysi osoittaa tämän oletuksen paikkansapitäväksi: semanttisessa luokassa 'prosessi' MAHDOLLISESTI-sanalla on jonkin verran enemmän kollokaatteja kuin 'tila'-luokassa (5,5 %, n=20). Muun muassa TULLA, ANTAA, HAKEA ja LAATIA kuuluvat 'prosessi'-luokkaan.

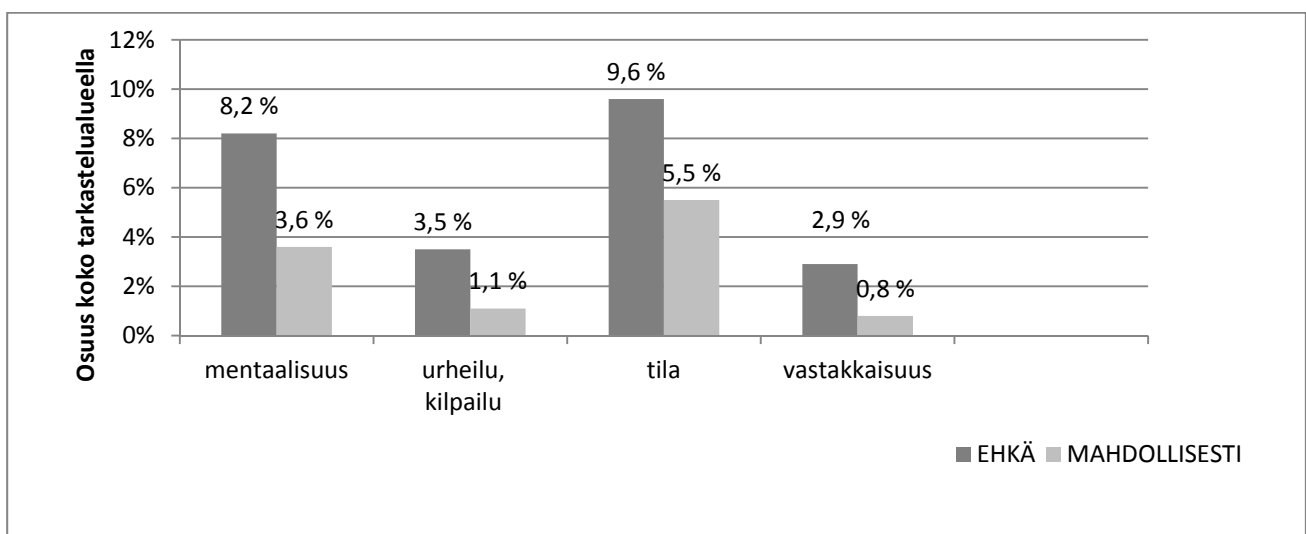
Seuraavaksi suurin ryhmä MAHDOLLISESTI-sanalla on 'additiivisuus' (8,2 %, n=30). Merkitysluokassa on erityisesti JA-konjunktion esiintymiä, mutta myös esimerkiksi sanoja LISÄKSI ja MYÖS. Additiivisuutta ilmaisevat sanat ovat jonkin verran yleisempiä MAHDOLLISESTI-sanalla kuin EHKÄ-sanalla kotekstissa: jälkimmäisellä niitä on 4,4 % (n=15). Jo kollokaattianalyysissä, partikkeleita analysoitaessa, tuli ilmi, että additiivisuutta ilmaisevat sanat olisivat hieman yleisempiä MAHDOLLISESTI-sanalla kuin sen synonyymien kotekstissa, ja havainto saa siis vahvistusta näistä semanttisen preferenssin analyysissä ilmenneistä osuuksista.

Myös 'aikaan' ja 'määrään, intensiteettiin' viittaavia sanoja esiintyy jonkin verran MAHDOLLISESTI-sanalla kotekstissa. Molemmat luokat kattavat 7,7 % (n=28) merkityskirjosta. Näistä 'aika'

osoittautui merkittäväksi ryhmäksi jo kollokaattianalyysissä, sillä niin substantiivi-, numeraali- kuin adverbikollokaateissakin oli 'aikaan' viittaavia sanoja (mm. TULEVAISUUS, 2000, JO). Myös t-testin kymmenen kärkikollokaatin joukossa korostuivat aikaan viittaavat sanat. Näistä havainnoista huolimatta modaalisanoista EHKÄ-sanalla on kuitenkin suurempi prosentuaalinen osuus aikaan viittaavia sanoja kotekstissaan (10,2%); kollokaattianalyysissäkin EHKÄ osoitti sekin kollokoivansa aikaan viittaavien sanojen kanssa. MAHDOLLESTI-sanasta havaitaan myös, että se osoittaa vähiten preferenssiä 'vastakkaisuuteen' (0,8%, n=3), 'kulttuuriin' (1,1%, n=4) ja 'urheiluun, kilpailuun' (1,1%, n=4).

MAHDOLLESTI-sanallakin ryhmässä 'luokittelemattomat' on paljon sanoja (7,4 %, n = 27). Se on hieman ongelmallinen ryhmä, aivan kuten EHKÄ-sanankin tapauksessa, sillä MAHDOLLESTI-sanallakin luokittelemattomien ryhmä on suurehko. Ryhmän sanat eivät kuitenkaan luontevasti sovi muihin merkitysryhmiin; muun muassa konjunktio ETTÄ on jätetty luokittelemattomien ryhmään.

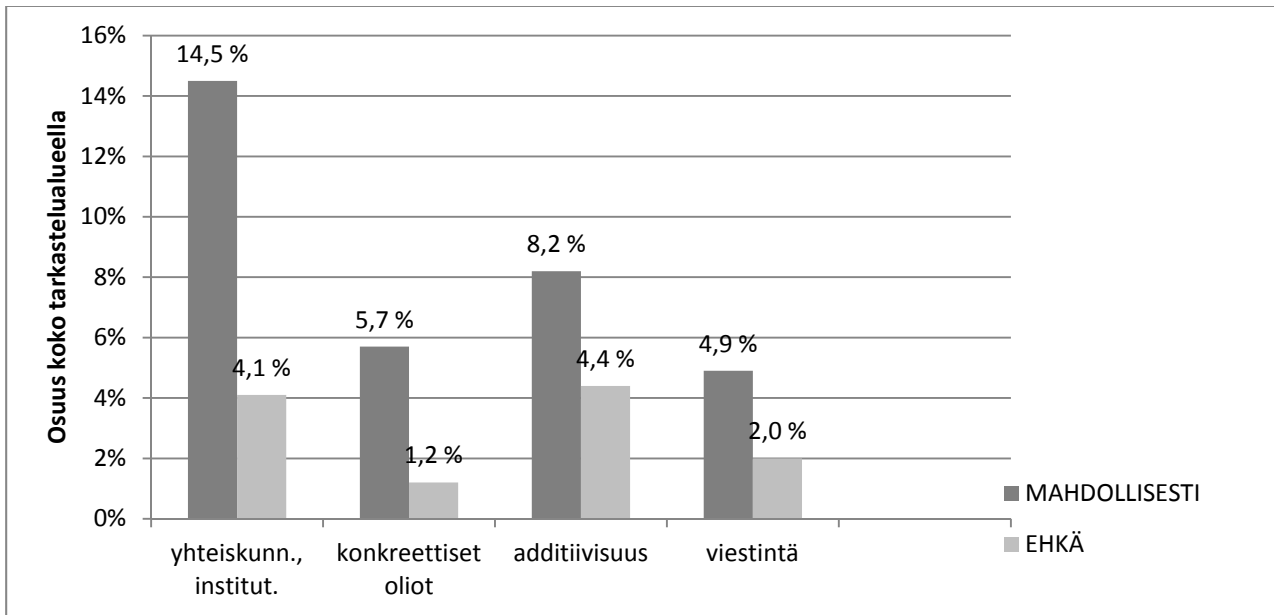
Kollokaatit jakautuvat siis kummallakin modaalisanalla melko tasaisesti eri merkitysryhmiin. Seuraavaksi tarkastelen, mitkä semanttiset ryhmät puolestaan erottavat modaalisanoja toisistaan eniten. Edellä on jo tullut esille, että MAHDOLLESTI-sanalla on synonyymiinsa nähden selvä preferenssi luokkaan 'yhteiskunnallisuus, institutionaalisuus' ja hienoinen preferenssi luokkaan 'additiivisuus' ja että EHKÄ puolestaan osoittaa synonyymiaan suurempaa preferenssiä ainakin luokkaan 'tila'. Muitakin erottavia ryhmiä kuitenkin on. Kaiken kaikkiaan modaalisanojen tärkeimmät erot semanttisen preferenssin osalta on kuvattu kuvioissa 2 ja 3. Kuvioihin on otettu merkitysryhmistä mukaan ne, joissa modaalisanojen ero on tilastollisesti melkein merkitsevä, merkitsevä tai erittäin merkitsevä.



Kuvio 2. EHKÄ-sanan tilastollisesti merkitsevimmät semanttiset preferenssit MAHDOLLESTI-sanaan nähden.

Kuviossa 2 on esitetty merkitysluokat, joissa EHKÄ-sanan kollokaattien osuus on suurempi kuin MAHDOLLIESTI-sanan ja joissa modaalisanojen välinen ero on tilastollisesti vähintään melkein merkitsevä ($p < 0,05$). Kuviossa kuvataan, kuinka suuri prosenttiosuus kummankin modaalisanan kotekstista edustaa kutakin merkitysluokkaa. Merkitysryhmät on esitetty erojen mukaisessa suuruusjärjestyksessä siten, että ensimmäisenä vasemmalla on merkitysryhmä, jossa modaalisanojen välinen ero on tilastollisesti merkitsevin. Prosenttiyksiköiden erot ovat varsin pieniä, esimerkiksi 'urheilu, kilpailu' -luokassa vain 2,4 %, mutta prosenttiyksiköiden erojen suuruudet sinänsä eivät kerro tilastollisesta merkitsevyydestä. EHKÄ-sanalla voidaan kuvion perusteella sanoa olevan synonyymiinsa nähden preferenssi merkitysryhmiin 'mentaalisuus' ($z=2,63$, $p < 0,01$), 'urheilu, kilpailu' ($z=2,16$, $p < 0,05$), 'tila' ($z=2,10$, $p < 0,05$) ja 'vastakkaisuus' ($z=2,08$, $p < 0,05$). Kuvion neljästä merkitysluokasta 'mentaalisuus'-luokassa modaalisanojen välinen prosenttiosuuskien ero ylittää varsinaiseen tilastolliseen merkitsevyyteen. Merkitysryhmän osalta on siis 1 prosentin tai pienempi todennäköisyys, että kyse modaalisanojen erossa on puhtaasta sattumasta. Muissa kolmessa luokassa ero on tilastollisesti melkein merkitsevä, eli on 5 prosentin tai pienempi todennäköisyys, että kyse on sattumasta. 'Urheilu, kilpailu'- ja 'vastakkaisuus'-luokissa absoluuttiset frekvenssit ovat hyvin pieniä, joten sikäli 'mentaalisuus' nousee luokista kiinnostavimmaksi modaalisanoja erottavaksi EHKÄ-sanana preferenssiksi.

Kuviossa 2 on esitetty merkitysryhmät, joissa puolestaan MAHDOLLIESTI-sana osoittaa suurinta preferenssiä EHKÄ-synonyymiinsa nähden. Nämä merkitysryhmät ovat 'yhteiskunnallisuus, institutionaalisuus', 'konkreettiset oliot', 'additiivisuus' ja 'viestintä'. Näistä 'yhteiskunnallisuutta, institutionaalisuutta' voidaan pitää modaalisanoja toisistaan hyvin selvästi erottavana tekijänä, sillä ero modaalisanojen välillä on tilastollisesti erittäin merkitsevä ($z=4,73$, $p < 0,001$). Ero näkyy myös prosenttiosuuskien erossa: luokan osuus MAHDOLLIESTI-sanana kotekstissa on siis 14,5% ja EHKÄ-sanana kotekstissa vain 4,1 %. Samoin merkitysluokassa 'konkreettiset oliot' sanojen välinen ero on tilastollisesti erittäin merkitsevä ($z=3,30$, $p < 0,001$). 'Additiivisuus'-luokassa ero puolestaan on tilastollisesti melkein merkitsevä ($z=2,09$, $p < 0,05$), samoin 'viestintä'-luokassa ($z=2,08$, $p < 0,05$).



Kuvio 3. MAHDOLLISESTI-sanana tilastollisesti merkitsevimmät semanttiset preferenssit EHKÄ-sanana nähden.

Kaiken kaikkiaan semanttisen preferenssin analyysi siis osoittaa, että kollokaatit jakautuvat melko tasaisesti eri semanttisiin ryhmiin. Jantunen toteaa tutkimuksessaan (2004: 140), että astemääritteillä abstrakteina, kieliopillisina sanoina ei ole yhtä tai kahta selvää semanttista preferenssiä ja että taajimmin esiintyviä astemääritteitä voidaankin pitää eräänlaisina yleismääritteinä, joiden kontekstissa saattaa esiintyä hyvin monenlaisia semanttisia piirteitä. Samoin modaalisanat ovat varsin abstrakteja merkitykseltään ja myös kieliopillisia sanoja, ja se lienee merkittävä syy siihen, miksi niiden kontekstissa esiintyy laaja kirjo erilaisia semanttisia piirteitä.

Kummallakin sanalla kuitenkin on joitain semanttisia preferenssejä toisiinsa nähden, kuten luvut osoittavat. Niistä voidaan suurimpina ja tärkeimpinä eroina pitää niitä kolmea merkitysluokkaa, joissa modaalisanojen prosentiosuuksien erot ovat tilastollisesti merkitseviä tai erittäin merkitseviä. MAHDOLLISESTI osoittaa siis selvää preferenssiä ensinnäkin yhteiskunnallisuuteen ja institutionaalisuuteen viittaaviin sanoihin. Täten MAHDOLLISESTI siis esiintyy EHKÄ-synonyymiaan tyypillisemmin sellaisissa konteksteissa, joissa on puhe esimerkiksi politiikasta, taloudesta tai oikeudenkäynnistä. Lisäksi tärkeä ero on, että EHKÄ-sanalla on synonyymiaan suurempi taipumus kollokoida 'mentaalisuus'-luokkaan kuuluvien sanojen kanssa ja MAHDOLLISESTI-sanalla puolestaan konkreettisiin olioihin viittaavien sanojen kanssa. Siten erottavaksi piirteeksi nousee mentaalisuus-konkreettisuus-jako. Toisaalta myös EHKÄ-sanana yhteydessä esiintyy konkreettisuuteen viittaavia sanoja, sillä konkreettisuuteen viittaavat myös esimerkiksi monet 'prosessi'-luokan sanat, eivät ainakaan konkreettiset oliot. 'Prosessi' onkin EHKÄ-sanana merkitysluokista yksi suurimpia, neljän-

neksi suurin. Siitä huolimatta 'mentaalisuus'- ja 'konkreettiset oliot' -luokkien erojen tilastollinen merkitsevyys osoittaa, että modaalisanojen yhtenä erona voidaan pitää sitä, että EHKÄ-sanana kontekstissa painottuvat merkitykseltään mentaaliset kollokaatit ja MAHDOLLISESTI-sanana kontekstissa puolestaan merkitykseltään konkreettiset kollokaatit.

6 PÄÄTÄNTÖ

6.1 Modaalisanoja yhdistävät ja erottavat kollokaatit ja semanttiset preferenssit

Tutkielmassa olen tarkastellut lähisynonymien EHKÄ ja MAHDOLLISESTI kotekstuaalisia piirteitä kollokaattien ja semanttisen preferenssin osalta. Tarkoituksena on ollut selvittää, miten kollokaatit ja semanttinen preferenssi erottavat synonyymisia modaalisanoja toisistaan ja toisaalta yhdistävät niitä. Aineistona on ollut Suomen kielen tekstikokoelman sanomalehtiaineisto, joten tuloksia ei voida yleistää kaikkeen kielenkäyttöön, mutta tutkielma tarjoaa tietoa sanomalehtikielen piirteistä.

Modaalisanoja yhdistäviksi kotekstuaalisiksi tekijöiksi on osoittautunut ensinnäkin kymmenen kollokaattia: molempien modaalisanojen tilastollisesti merkitseviä kollokaatteja ovat ENSI, JOKIN, JOKU, JOPA, MYÖHEMMIN, TAI, TULEVAISUUS, TULLA, VIELÄ ja VOIDA. Näistä 'epävarmuuden' merkitystä ilmentävät JOKIN, JOKU, VOIDA ja TAI. Kuitenkaan semanttisen preferenssin analyysissä 'epävarmuus' ei noussut kummankaan sanan vahvaksi semanttiseksi preferenssiksi: EHKÄ-sanankotekstissa merkitysluokka kattaa 3,5 % koko merkityskirjosta ja MAHDOLLISESTI-sanankotekstissa puolestaan 3,3 %. Synonyymeja yhdistävä kotekstuaalinen tekijä on myös verbikollokaattien suuri määrä, sillä EHKÄ-sanalla on 18 verbikollokaattia ja MAHDOLLISESTI-sanalla 16. Tätä voidaan pitää myös kolligaatiopiirteenä. EHKÄ-sanankotekstissa verbikollokaateista suuri osa on nimenomaan mentaalisiä verbejä, mutta toisaalta synonyymeja yhdistää se, että molemmilla esiintyy kotekstissaan primääri-A-verbeiksi luokiteltavia verbejä eli sellaisia verbejä, jotka ilmentävät konkreettista tilaa tai prosessia. Etenkin konkreettista prosessia ilmentävät sanat näyttävät korostuvan kummankin modaalisanankotekstissa. Tämä ilmenee myös semanttisten luokkien analyysissä, jossa 'prosessi' on molempien sanan vahvimpia semanttisia preferenssejä: EHKÄ-sanankotekstin merkitysluokista se on kolmanneksi suurin ja kattaa 8,7 % aineistosta, ja MAHDOLLISESTI-sanalla se on puolestaan toiseksi suurin merkitysluokka 11,2 %:n osuudella. Kaiken kaikkiaan prosessia ilmentävät sanat, jotka yleensä ovat siis verbejä, ovat tyypillisiä modaalisanojen kollokaatteja. Toisaalta myös paljon verbejä sisältävä merkitysluokka 'tila' on 9,6 %:n osuudellaan EHKÄ-sanankotekstin suurimpia merkitysluokkia, toiseksi suurin.

Toisaalta modaalisanoja yhdistää myös se, että niiden kollokaatit jakautuvat melko tasaisesti eri semanttisiin ryhmiin; kummallakaan sanalla ei ole yhtä erittäin vahvaa semanttista preferenssiä. Molempien kotekstissa kuitenkin esiintyy 'prosessiin' viittaavien sanojen lisäksi paljon 'aikaan' viittaavia sanoja. Tämä ilmeni jo tilastollisesti merkitsevistä kollokaateista, sillä TULEVAISUUS, ENSI,

MYÖHEMMIN ja tietyissä konteksteissa myös VIELÄ ja TULLA ilmentävät jollain tavalla aikaa. 'Aika' on EHKÄ-sanalla hieman vahvempi semanttinen preferenssi kuin synonyymillaan MAHDOLLISESTI: EHKÄ-sanalla se on kaikkein suurin merkitysluokka 10,2 %:n osuudella kontekstin sanoista ja MAHDOLLISESTI-sanalla puolestaan viidenneksi suurin 7,7 %:n osuudella kontekstin sanoista. Ero ei kuitenkaan ole tilastollisesti merkitsevä ($z=1,19$, $p > 0,05$), ja voidaankin sanoa, että molempien modaalisanojen kontekstissa esiintyy tyypillisesti 'aikaan' viittaavia sanoja. Myös 'määrä, intensiteetti' on tyypillinen merkitysluokka molempien sanojen kontekstissa. EHKÄ-sanalla kontekstin merkityskirjosta se kattaa 8,7 % ja on kolmanneksi suurin merkitysluokka yhdessä samankokoisen 'prosessi'-luokan kanssa. MAHDOLLISESTI-sanalla kontekstissa puolestaan 7,7 % sanoista ilmentää 'määrä, intensiteettiä', ja luokka on viidenneksi suurin. Lisäksi 'kausaalisuuteen' viittaavat sanat nousivat jossain määrin esille kollokaatioanalyysissä, mutta semanttisen preferenssin analyysin mukaan 'kausaalisuus' ei ole kummallakaan modaalisanalla suuri merkitysluokka: EHKÄ-sanalla luokkaa edustaa 1,2 % kontekstin sanoista ja MAHDOLLISESTI-sanalla 3,3 % kontekstin sanoista.

Kollokaatioanalyysi osoitti joitakin modaalisanoja erottavia piirteitä. Ensinnäkin kummallakin modaalisanalla on lukuisia sellaisia tilastollisesti merkitseviä kollokaatteja, joita vastaavasti toisella ei ole. Lisäksi tuli ilmi, että EHKÄ-sanalla on kollokaatteina huomattavasti enemmän adverbeja – etenkin määrän suuruuteen ja tavan intensiteettiin viittaavia – sekä adjektiiveja kuin synonyymillaan MAHDOLLISESTI. Adjektiiveista monet esiintyvät nimenomaan superlatiivimuodossa. Lisäksi sanan verbikollokaateissa korostuvat mentaaliset verbit ja pronominikollokaateissa synonyymiinsa nähden persoonapronominit. MAHDOLLISESTI-sanalla ei ole missään sanaluokkaryhmässä enemmän tilastollisesti merkitseviä kollokaatteja kuin synonyymillaan. Verbikollokaateissa sillä kuitenkin korostui verbien infiniittimuotoisuus sekä konkreettisten verbien suurempi määrä. Partikkeli- ja adpositiokollokaateissa sillä hienoisesti korostuivat additiivisuutta jollain tavalla ilmentävät kollokaatit. Monet kollokaatioanalyysissä esiin tulleista piirteistä ovat samalla itse asiassa myös kolligatiivipiirteitä; niitä nousi tässä tutkielmassa esiin kollokaatioanalyysin ohessa. Kolligatioiden tutkiminen kvantitatiivisesti ja systemaattisesti antaisi modaalisanojen synonyymiasta vielä täydellisemmän kuvan.

Modaalisanoja erottavia, tilastollisesti vähintään melkein merkitseviä merkitysluokkia on kahdeksan. EHKÄ-sana osoittaa synonyymiinsa nähden preferenssiä luokkiin 'mentaalisuus', 'urheilu, kilpailu', 'tila' ja 'vastakkaisuus'. MAHDOLLISESTI-sanalla kontekstissa puolestaan esiintyy synonyyminsa kontekstia useammin 'yhteiskunnallisuuteen, institutionaalisuuteen', 'konkreettisiin olioihin', 'additiivisuuteen' ja 'viestintään' viittaavia sanoja. Voidaan ajatella, että erityisesti se, että MAHDOLLISESTI esiintyy tyypillisesti kontekstissa, jossa on puhe virallisista instituutioista ja yhteiskunnallisista asioista, antaa yhden vastauksen siihen, miksi MAHDOLLISESTI vaikuttaa tyyliin vi-

rallisemmalta kuin synonyyminsa. Samoin verbikollokaattien infiniittivimuotoisuus saattaa yhtenä tekijänä selittää, miksi MAHDOLLISESTI tuntuu tyyliiltään virallisemmalta ja ehkä kirjallisemmalta kuin EHKÄ. Infiniittimuodot liittyvät nimittäin usein kirjakielelle tyypillisiin rakenteisiin (Saukkonen 1970). EHKÄ-sanana puhekielisyttä taas ilmentää puhekielelle tyypillinen kollokaatti NO.

Havaintujen erojen perusteella voidaan pohtia, onko modaalisanon yksi ero myös siinä, millaisesta episteemisyden lajista on kyse. Aiemmassa tutkimuksessa on keskusteltu etenkin episteemisten modaali-ilmausten subjektiivisuudesta ja objektiivisuudesta, mutta eri tutkijoilla on asiasta erilaisia näkemyksiä (ks. mm. Lyons 1977: 797–800, Palmer 1986:16, Traugott 1989: 36, Lehtinen 1997: 233–235). Voidaan joka tapauksessa ajatella, että molemmilla modaalisanoilla on jonkinlainen varmuuden asteeseen ja totuuden arviointiin liittyvä peruserkitys (ks. Hakulinen ja Sorjonen 1989: 97, Lehtinen 1997: 246) ja että eri konteksteissa tarkat tulkinnat saattavat poiketa toisistaan hieman sen suhteen, korostuuko niissä esimerkiksi subjektiivinen tai objektiivinen modaalisuus. Virkettä (105) voidaan tarkastella esimerkkinä, sillä se on tutkimuksen tulosten perusteella tyypillinen EHKÄ-sanana koteksti, koska virkkeessä esiintyvät sanan tilastollisesti merkitseviin kollokaatteihin kuuluvat HIEMAN ja *turhan* sekä adjektiivi (PÄÄLLEKÄYVÄ), jotka niin ikään ovat tyypillisiä sanan kotekstissa. Tutkimuksen tulosten perusteella voidaan esittää oletus, että sana EHKÄ edustaa usein subjektiivista modaalisuutta ja yksilön tietoa tai mielipidettä. Esimerkiksi virke (105) lienee syytä tulkita niin, että puhuja esittää omana mielipiteenään arvion kappaleen alun päällekkäyvydestä sen sijaan, että kyse olisi objektiivisesta faktasta.

105) Kappaleen alku oli **ehkä** hieman turhan päällekkäyvä.

106) Poliisi pyytää tietoja murtoon ja pahoinpitelyyn **mahdollisesti** liittyvistä havainnoista.

Virke (106) puolestaan on yksi MAHDOLLISESTI-sanana tyypillisistä käyttökonteksteista, sillä siinä esiintyvät sen tilastollisesti merkitseviin kollokaatteihin kuuluvat LIITTYÄ ja HAVAINTO sekä lisäksi rikoksiin liittyviä sanoja (MURTO ja PAHOINPITELY). MURTO ja PAHOINPITELY eivät nousseet kollokaattianalyyssissa tilastollisesti merkitseviksi kollokaateiksi, mutta ovat kuitenkin rikoksiin ja siten 'yhteiskunnallisuuteen, institutionaalisuuteen' viittaavina sanoina MAHDOLLISESTI-sanana kotekstissa tyypillisiä. Sana MAHDOLLISESTI ei näytä useinkaan edustavan subjektiivista modaalisuutta ja yksilön henkilökohtaista mielipidettä, vaan pikemmin objektiivisempaa episteemisyttä. Virkkeessä (106) modaalisanalla ilmaistu epävarmuus kohdistuu siihen, liittyvätkö havainnot todella kyseessä olevaan murtoon ja pahoinpitelyyn vai eivät. Voidaan eräältä kannalta ajatella, että kyseessä on objektiivinen fakta: havainnot joko liittyvät kyseiseen rikokseen tai sitten eivät, ja se, miten asia on, ei ole kenenkään subjektiivinen näkemys. Täten virkkeen modaali-ilmaus MAHDOLLISESTI voidaan

tulkita objektiiviseen modaalisuuteen kuuluvaksi. Toisaalta objektiivisen modaalisuuden käsitettä on kritisoitu (ks. mm. Lehtinen 1997: 234, Kangasniemi 1992: 150), ja tilalla voidaan pikemmin käyttää käsitettä intersubjektiivisuus, jonka ottaa esiin Kangasniemi (mp.).

Ero voidaan toisaalta kuvata myös käsitteillä yksilön tieto ja yhteisön tieto, joita Jaakola (2011: 523) käyttää kuvatessaan sanojen TIETTÄVÄSTI ja TIETÄÄKSENI eroa. Sanaa TIETTÄVÄSTI voidaan hänen mukaansa pitää yhteisön tietoa kuvaavana ja sille lähimerkityksistä sanaa TIETÄÄKSENI puolestaan yksilön tietoa kuvaavana. Kyse on episteemisyiden alalajista, evidentialisuudesta, jolla osoitetaan, millä tavalla puhuja on tietonsa saanut (VISK § 1557). Modaalisanon osalta kyse näyttäisi ainakin osin olevan siitä, että EHKÄ saa usein tulkinnan, joka voidaan evidentialisuuden jatkumolla sijoittaa pikemmin yksilön tiedoksi kuin yhteisön tiedoksi, ja MAHDOLLESTI puolestaan tulkinnan, joka on pikemmin yhteisön tietoa. Tätä ajatusta tukee ensinnäkin se, että EHKÄ-sanon kotekstissa esiintyy sen synonyymiin nähden huomattavasti useammin adjektiiveja ja adverbeja; adjektiivit ja adverbit ilmaisevat esimerkiksi asioiden laatua, tapaa ja arvoa, jotka ovat usein suhteellisia, yksilön subjektiivisesta tulkinnasta riippuvia asioita. Myös semanttinen preferenssi 'mentaalisuus' tukee esitettyä ajatusta, sillä luokkaan kuuluu emootioihin viittaavia sanoja ja emootiot ovat nimenomaan yksilön subjektiivisia kokemuksia. Lisäksi voidaan ajatella, että pronominkollokaateista tehty havaintokin viittaa tähän: EHKÄ-sanalla on synonyymistaan poiketen tilastollisesti merkitsevien kollokaattien joukossa persoonapronominit MINÄ ja ME, joilla puhuja viittaa nimenomaan itseensä tai joukkoon, johon itse kuuluu. EHKÄ myös esiintyy aineistossa huomattavasti synonyymiaan tyypillisemmin erilaisissa kirja-, elokuva- ja muissa arvosteluissa eli teksteissä, joissa on kyse nimenomaan puhujan subjektiivisista arvioista, ei objektiivisesta tekstistä, kuten sanomalehdissä pääosin. Toiseksi MAHDOLLESTI-sanon semanttisia preferenssejä synonyymiinsa nähden ovat 'yhteiskunnallisuus, institutionaalisuus' sekä 'konkreettiset oliot'. Yhteiskunnalliset ja institutionaaliset asiat koskevat koko yhteisöä, ja konkreettisiin olioihin viittaavilla sanoilla ilmaistaan asioita, jotka ovat kaikkien konkreettisesti havaittavissa, eivät yksilöllisiä tulkintoja.

Täten tutkimus tuo esiin oletuksen, jonka mukaan sanaa EHKÄ voidaan pitää erityisesti subjektiivisena, puhujan henkilökohtaisia mielipiteitä ja yksilön tietoa esiin tuovana modaali-ilmauksena ja sanaa MAHDOLLESTI objektiivisena, erityisesti faktoja korostavana, yhteisön tietoa ilmentävänä modaali-ilmauksena. Toisaalta täytyy huomata, että kaikki sanaesiintymät eivät välttämättä noudata tätä jakoa. Onkin korostettava, että subjektiivisuuteen, objektiivisuuteen, intersubjektiivisuuteen sekä yksilön ja yhteisön tietoon liittyvät seikat vaatisivat vielä tarkempaa tutkimusta niin sanojen EHKÄ ja MAHDOLLESTI kuin muidenkin modaali-ilmausten osalta.

Esitettyä tulkintaa ei voida pitää ainoana olennaisena sanoja erottavana tekijänä myöskään sikäli, että modaalisanon yksi ero voi piillä myös siinä, missä sanat tyypillisesti sijaitsevat epis-

teemisten ilmausten jatkumolla. Jatkumoa voidaan karkeasti kuvata niin, että siinä ovat asteet mahdoton - epätodennäköinen - epävarma - mahdollinen - todennäköinen - varma; kategoriat eivät kuitenkaan ole käytännössä selvärajaisia (VISK § 1556). Tutkimuksen – ja myös kielellisen intuition – perusteella vaikuttaa siltä, että molemmat modaalisanat sijoittuvat jatkumolla 'epävarman' kategoriaan. Eri esiintymät voivat kuitenkin poiketa toisistaan, ja etenkin on huomattava, että sana EHKÄ ei aina ilmennä sitä, että puhuja olisi varsinaisesti epävarma. Kyse voi olla myös eräänlaisesta pehmentelystä, mikä tuli esille esimerkiksi luvussa 2.4.1 virkkeessä ”*Ehkä meillä on puutetta siitä yhteydestä.*” Voidaan ajatella, että tämä ilmenee myös esimerkiksi EHKÄ-sanana adjektiivikollokaattien superlatiivimuotoisuudesta ja tilastollisesti merkitsevästä kollokaatista KAIKKEIN, sillä nämä piirteet vaativat usein eräänlaista pehmentelyä; puhuja ei kenties halua esittää voimakasta väitettä ilman lieventävää modaalisanaa. MAHDOLLISESTI-sana puolestaan vaikuttaa esiintyvän aina nimenomaan epävarmuutta ilmentämässä, eikä sillä näytä olevan samanlaista pehmentelyn merkitystä kuin synonyymillaan EHKÄ. Täten modaalisanojen erot voivat siis liittyä sekä evidentialiseen jatkumoon ”yhteisön tieto – yksilön tieto” että toisaalta episteemiseen jatkumoon, jossa on kyse tiedon varmuusasteesta. Jatkotutkimuksessa voisikin olla hedelmällistä perehtyä tarkemmin synonyymisten modaalisanojen eroon evidentialisen ja episteemisen jatkumon kannalta.

6.2 Tutkimuksen arviointia

Tässä tutkimuksessa erityisesti semanttisen preferenssin analyysi vaati vahvasti tutkijan oman kielellisen intuition käyttöä. Kielen sanojen luokittelu semanttisiin luokkiin ei ole yksitulkintaisista. Ensinnäkin olisi mahdollista ottaa mukaan erilaiset 19 luokkaa kuin tässä tutkielmassa tehtiin tai vaihtoehtoisesti eri määrä merkitysluokkia. Toisekseen voidaan sanoa, että mitkä merkitysluokat jaottelun pohjaksi valitaankin, sanat on joka tapauksessa mahdollista luokitella sanat käytettyihin ryhmiin monella eri tavalla, sillä monet sanat eivät edusta ainoastaan yhtä merkitysryhmää, vaan useita. Tällaisissa tapauksissa luokittelussa oli usein valittava se merkitysryhmä, jota sana tuntuu kontekstissään vahvemmin edustavan. Kyse on siis väistämättä siitä, että tutkija käyttää omaa intuitiotaan luokittelua tehdessään. Voidaan kuitenkin ajatella, että intuitiiviset näkemykset kielestä ovat läsnä ylipäättään kaikessa kielentutkimuksessa. Monitulkintaisten sanojen luokittelussa käytettiin myös periaatetta, jonka mukaan luokat 'prosessi' ja 'arvo, laatu' ovat toissijaisia muihin luokkiin nähden. On mahdollista, että esimerkiksi luokka 'arvo, laatu' olisi osoittautunut huomattavasti suuremmaksi luokaksi etenkin EHKÄ-sanalla, jos kaikki arvoon tai laatuun viittaavat sanat olisi todella luokiteltu 'arvo, laatu' -luokkaan eikä merkitykseltään kapea-alaisimpiin ryhmiin. Kollokaatti-

analyysi nimittäin osoitti, että EHKÄ-sanalla on paljon adjektiivikollokaatteja, mutta adjektiivien eli laatuun viittaavien sanojen suuri määrä ei kuitenkaan näkynyt voimakkaana semanttisena preferenssinä 'arvo, laatu'.

Semanttisen preferenssin analyysissa tuli esiin toinenkin ongelma. Kun nimittäin tutkittavien sanojen kollokaatteja luokitellaan merkitysryhmiin, herää kysymys, onko sanoja oikeastaan mahdollista luokitella merkitysryhmiin pelkkinä leksemeinä ja olla ottamatta merkityksessä huomioon niiden kulloistakin kieliopillista muotoa. Esimerkiksi tämän tutkielman 'prosessi'-merkitysluokassa on monta partisiippimuotoista verbiä, kuten *jäävänsä*, *käytetyistä*, *leivottuja* ja *liikuteltava*. Partisiippimuotoisuus nähdäkseni heikentää verbien 'prosessiin' liittyvää merkitystä. Jos samat verbit olisivat esimerkiksi muodoissa *jään*, *käyttävät*, *leipovat* ja *liikuttelemme*, 'prosessin' merkitys tulisi sanoista kenties paremmin esiin. Samoin esimerkiksi superlatiivimuotoisuus voi tuoda adjektiiviin 'määrän' merkitystä (*surullisinta*), liitepartikkeli *-ko* verbiin 'epävarmuuden' merkitystä (*oliko*) ja paikallissija adjektiiviin 'paikan' merkitystä (*korkeammallakin*). Työssä pyrin tarkastelemaan sanoja leksemeinä ja jättämään merkitystä pohtiessani huomiotta kunkin saneen kieliopillisen muodon, mutta voidaan kysyä, onko semanttisen preferenssin analyysi syytä todella perustaa ainoastaan sanojen merkityksen tutkimiseen. Esimerkiksi edellä luetelluista esimerkeistä huomataan, että lekseemin merkityksen lisäksi kulloinenkin kieliopillinen muoto vaikuttaa sanan merkitykseen. Kyse on itse asiassa kontekstuaalisessa semantiikassa tähdenntystä leksikon ja kieliopin erottamattomuudesta.

Voidaankin ajatella, että lekseemien merkitykset yksinään eivät ratkaise synonyymien valintaa, vaan prosessi on huomattavasti monimutkaisempi. Siihen, mikä modaalisana valitaan lauseeseen syntagmaattiseen suhteeseen, vaikuttaa semanttisten piirteiden lisäksi myös muun muassa sanojen kieliopilliset muodot, kuten se, käytetäänkö lauseessa esimerkiksi partisiippimuotoista verbiä. Semanttinen preferenssi kontekstuaalisena valintatekijänä ei olekaan erillään muista kontekstuaalisista piirteistä. Kaiken kaikkiaan koska kielenkäyttöön ja synonyymien keskinäiseen valintaan vaikuttavat useat eri muuttujat, synonyymien kontekstuaalisesta valinnasta saataisiin perusteellisempi kuva hyödyntämällä monimuuttujamenetelmiä, sillä niiden avulla voidaan tarkastella usean eri tekijän samanaikaista vaikutusta synonyymien valintaan. Suomen kielen tutkimuksessa menetelmää on toistaiseksi käyttänyt Arppe (2008).

Semanttisen preferenssin analyysissa ongelmallisena voidaan pitää niin ikään aineiston pienekköä kokoa. Voidaan ajatella, että mitä laajempi aineisto on, sitä luotettavampia tutkimustulokset ovat. Tässä tutkielmassa ei ollut tutkielman rajallisen laajuuden vuoksi mahdollista tarkastella erityisen laajaa aineistoa semanttisen preferenssin osalta. EHKÄ-sanank tarkastelualueella oli 343 sanetta ja MAHDOLLISESTI-sanank tarkastelualueella 366 sanetta. Ongelmallista on etenkin, että pienimmät

merkitysluokat koostuvat vain muutamasta sanasta; tällöin suhteellisten osuuksien testauksen tuloksia voidaan pitää etenkin pienimpien merkitysluokkien osalta lähinnä suuntaa antavana. Etukäteen ei voitu toisaalta tietää, kuinka paljon sanoja mihinkin luokkaan tulisi, vaan luokittelu tehtiin sen mukaan, millaisia ryhmiä ja piirteitä aineistosta nousi esiin. Lisäksi esimerkiksi Jantusen (2004: 85, 133) vastaavan analyysin aineisto on kooltaan suunnilleen samansuuruinen ja kahden tutkittavan sanan osalta jopa pienempi kuin tässä tutkimuksessa: Jantusen semanttisen preferenssin analyysissä HYVIN-astemääritteen aineisto on 352 sanetta, KOVIN-astemääritteen 176 sanetta ja OIKEIN-astemääritteen 121 sanetta. Ei ole siis tavatonta, että semanttista preferenssiä tarkastellaan suhteellisen pienistä aineistoista. Analyysin luonne selittääkin aineiston pienuutta: jokainen sanaesiintymä on käytävä yksitellen läpi ja luokiteltava merkityksen mukaiseen luokkaan, joten aineisto ei voi koostua esimerkiksi miljoonista sanoista, toisin kuin kollokaatioanalyysissä. Toisaalta on huomattava, että analyysin luotettavuutta lisää se, että kollokaatioanalyysissä esiin nousseet hypoteesit modaalisanon semanttisista preferensseistä ovat hyvin samansuuntaisia kuin itse semanttisen preferenssin analyysissä tehdyt huomiot. Kollokaatioanalyysi siis tukee semanttisen preferenssin analyysia.

Tässä tutkielmassa viitataan kollokaatiolla sanojen esiintymiseen teksteissä toistensa läheisyydessä. Kyseessä on siis kollokaation tekstuaalinen määrittely, jota muiden kollokaation määrittelmien tavoin käsiteltiin luvussa 2.2.2. Kollokaatioanalyysissä kuitenkin lisäksi korostuu kollokaation tilastollinen määrittelmä, sillä analysoitava aineisto on muodostettu siinä käyttämällä tilastollisia testejä. On ajateltavissa, että tilastollisuus antaa kollokaation määrittelmään eräänlaisen lisäulottuvuuden: kaikki kollokaatiot eivät välttämättä ole tilastollisesti merkitseviä, mutta osa on, ja tätä piirrettä voidaan tarkastella tilastollisin testein.

Voidaan ajatella, että kenties erityisesti tilastollinen merkitsevyys liittyy myös kollokaation kolmanteen määrittelytapaan, assosiatiiiviseen määrittelyyn, eli kertoo myös siitä, että sanat assosioituvat toisiinsa kielenkäyttäjän tajunnassa. Tilastolliset menetelmät ovat toki ensisijaisesti tilastollista tarkastelua varten, mutta voidaan kuitenkin ajatella, että esimerkiksi tilastollisesti merkitsevä kollokaatio EHKÄ <HIEMAN> ilmentää paitsi sitä, että kollokaatio on tilastollisesti merkitsevä, myös jollain tavalla sitä, että kielenkäyttäjän kognitiossa nämä sanat yhdistyvät voimakkaasti toisiinsa. Sille, että jokin myötäesiintymä nousee tilastollisesti merkitseväksi, on nimittäin oltava jokin syy, ja yhtenä syynä voidaan pitää sitä, että sanat assosioituvat toisiinsa kognitiossa, muodostavat siellä valmiin rakenteen ja sen myötä realisoituvat kielenkäytössä taajaan – ja synnyttävät tilastollisesti merkitseviä myötäesiintymiä. Täten kieli ei ole sattumanvaraisesti rakentunut, vaan perustuu siihen, miten kieltä on aina aiemmissa kielenkäyttötilanteissa tyypillisesti käytetty. Kun EHKÄ ja HIEMAN assosioituvat toisiinsa kielenkäyttäjän kognitiossa, kielenkäyttäjän on helppo toistaa tämä val-

mis rakenne valitessaan leksikaalisia yksiköitä syntagmaattisiin suhteisiin. Kyse on siis Hoeyn (2007) esiin ottamasta primingista (ks. luku 2.2.2), siitä, että aikaisemmat kielenkäyttötilanteet viritävät kielenkäyttäjän niin, että tämä käyttää sanaa usein sellaisissa konteksteissa, joissa sitä aiemminkin on käytetty. Myös synonyymien eroja voidaan tarkastella kollokaation tilastollisen ja assosiativisen määrittelyn sekä primingin käsitteen kautta: kollokaatio MAHDOLLISESTI <HIEMAN> ei ainakaan tilastollisin menetelmin nouse merkitseväksi kollokaatioksi, joten voidaan ajatella, että sanat eivät myöskään ole yhdistyneet toisiinsa kielenkäyttäjän kognitiossa yhtä vahvasti kuin EHKÄ ja HIEMAN. Toisaalta tilastollisin menetelmin ei toki suoraan voida tutkia sanojen assosioitumista toisiinsa tai muita mentaalisia ilmiöitä, mutta voidaan silti ajatella, että kollokaation vahva tilastollinen merkitsevyys kertoo vahvemmassa assosiaatiosta myös kognition tasolla verrattuna kollokaatioon, joka on esimerkiksi vain heikosti tai ei lainkaan merkitsevä.

6.3 Tutkimustulosten hyödyntäminen

Tutkimuksessa on syvennyt synonyymien EHKÄ ja MAHDOLLISESTI kontekstuaalisista piirteistä kollokaatteihin sekä semanttiseen preferenssiin. Tavoitteena on ollut tarkastella, millaisia kollokaatteja ja semanttisia preferenssejä näillä sanoilla on ja miten nämä piirteet yhdistävät synonyymeja toisiinsa ja erottavat niitä. Saatuja tuloksia tarkasteltiin kokoavasti luvussa 6.1. Suomen kielen tutkimuksessa etenkin semanttisen preferenssin tutkiminen on ollut toistaiseksi kaiken kaikkiaan hyvin vähäistä, joten tutkielma tarjoaa siten tuoretta näkökulmaa alalle. Semanttista preferenssiä voidaan pitää tärkeänä osana lekseemin ominaisuuksia, sillä se määrittää, millaisissa semanttisissa yhteyksissä sanaa käytetään. Sanan käyttöympäristöt taas ovat olennaisia koko kielenkäytön kannalta, sillä voidaan ajatella koko kielenkäytön perustuvan siihen, että sanoja valitaan paradigmaattisista suhteista syntagmaattisiin suhteisiin. Kielen fraseologiset piirteet ovat ikään kuin sääntöjä sille, millaisiin syntagmaattisiin suhteisiin kielen yksiköitä valitaan kielenkäytössä. Semanttinen preferenssi, kuten myös kollokaatiot ja fraseologinen tutkimus ylipäätään, kertovat siis olennaisia asioita kielestä.

Modaalisanon kollokaatteja ja semanttisia preferenssejä koskevia tutkimustuloksia, kuten fraseologisten yksiköiden tutkimusta laajemminkin, voidaan hyödyntää esimerkiksi suomi toisena kielenä -opetuksessa. Kielenoppijan kannalta kollokaatiot ovat yksi vaikeimmista osa-alueista sanaston oppimisessa, sillä ne on arbitraarisuutensa vuoksi yleensä opeteltava ulkoa (Jantunen 2009b: 361). Edistyneenkin tason kielenoppija tuottaa usein epätyypillisiä kollokaatioita. Esimerkiksi oppijan äidinkieli vaikuttaa suuresti kollokaatioiden muodostamiseen. (Nesselhauf 2003: 235–238; Jan-

tunen 2009b: 361.) Vastaavasti opettajien mielestä kollokaatiot ovat yksi hankalimmista asioista kielenopetuksessa: Tsui (2005: 335–339) toteaa tutkimuksessaan, että englantia vieraana kielenä opettavilla yksi eniten kysymyksiä herättävä asia ovat juuri kollokaatiot. Mauranen (2004: 95–96) puolestaan toteaa, että oppijoille saatetaan opettaa sellaisia fraseologisia ilmauksia, joita ei todellisuudessa esiinny kielessä tai jotka eivät ole lainkaan frekventtejä. Kollokaatiot ja ylipäättään kielen fraseologinen luonne onkin otettava huomioon kielenopetuksessa (Nesselhauf 2003: 237–238; Jantunen 2009b: 373–374). Kollokaatioiden lisäksi myös synonyymien erot kuuluvat kielenopetuksessa hankalimmiksi koettuihin piirteisiin (Tsui 2005: 338–339). Tämä tutkimus siis vastaa kahden modaalisanan osalta sekä kollokaatioihin että synonyymien eroihin liittyvään tiedon tarpeeseen. Vastaavasti mitä useammista leksikaalisista yksiköistä tehdään jatkossa tutkimusta fraseologisesta näkökulmasta, sitä paremmin fraseologista tietoa sanoista voidaan hyödyntää kielenopetuksessa.

Kuten Jantunen (2009b: 374–375) toteaa, kielenopetuksen avuksi olisi saatava nykyistä enemmän materiaalia, joka perustuu todellisen kielenkäytön tutkimukseen, esimerkiksi sanakirjoja. Jantunen mukaan suomen kielen sanakirjat, etenkin pienet opiskelijasanakirjat, eivät yleensä tarjoa kovin paljon tietoa sanojen fraseologisista piirteistä, vaan selittävät ilmauksia merkityksenselityksen, synonyymien ja esimerkkilauseiden avulla. Nähdäkseni fraseologista tutkimusta voitaisiin hyödyntää leksikografiassa tarjoamalla sana-artikkeleissa eksplisiittisesti tietoa esimerkiksi kollokaatioista ja myös valitsemalla esimerkkilauseiksi sellaisia lauseita, jotka kuvaavat sanan tyypillisiä käyttötapoja, esimerkiksi kollokaatioita ja semanttisia preferenssejä. Fraseologia voi olla jopa aivan keskeisessä asemassa sanakirjassa, kuten se on suomenoppijoille tarkoitetussa ConLexis-verkkosanakirjassa (sanakirjasta ks. Jantunen, Kumpulainen, Tammimies & Tokola 2013). Tällainen sanakirja tarjoaa kielenkäyttäjälle mahdollisuuden saada helposti tietoa sanojen tyypillisimmistä käyttötapoista.

Lehtosalo ja Tyysteri (2013: 14) toteavat puhuessaan sanakirjoista ylipäättään, eivät vain oppijansanakirjoista, että sana-artikkeleihin olisi valittava esimerkkejä, jotka kuvaavat, miten sanoja käytetään tyypillisesti. Juuri fraseologinen, korpuspohjainen tutkimus, jota tämäkin tutkielma siis edustaa, tarjoaa tietoa sanojen tyypillisistä käyttötapoista. Grönros (2013: 47) toteaa, että korpuukset tarjoavat leksikografeille tietoa nykyään enemmän kuin aiemmin on ollut saatavissa. Toisaalta kuten Grönros (mp.) mainitsee, kattavimmatkin korpuukset ovat aina rajallisia, ja jo niitä koostettaessa on tehty valintoja. Tämän tutkielman aineisto on rajallinen esimerkiksi sillä tavoin, että se perustuu vain sanomalehtikieleen. Modaalisanoilla EHKÄ ja MAHDOLLISESTI voi olla muita tyypillisiä käyttötapoja muissa konteksteissa, kuten puhekielessä. Tutkimus kuitenkin osaltaan vastaa esimerkiksi Sinclairin (1991a: 36–40) esittämään vaatimukseen, jonka mukaan sanakirjaa laadittaessa on käytettävä hyödyksi erityisesti todelliseen, autenttiseen kielenkäyttöön pohjautuvaa tietoa. Kvan-

titatiivisena fraseologiaa tarkastelevana korpustutkimuksena tämä tutkimus tarjoaa niin leksikografian kuin kielenopetuksenkin tueksi sellaista modaalisanoja koskevaa tietoa, jota esimerkiksi pelkästään kielellisen intuition perusteella ei ole mahdollista saavuttaa.

LÄHTEET

- Arppe, Antti 2001: Corpus-based observations on the use and occurrence of inflected word forms in Finnish – the case of synonyms. – Jussi Niemi & Janne Heikkinen (toim.), *Nordic and Baltic morphology: Papers from a NorFA course. Tartu, June 2000* s. 1–14. Studies in Languages 36. Joensuu: University of Joensuu.
- 2002: The usage patterns and selectional preferences of synonyms in a morphologically rich language. – Annie Morin & Pascale Sébillot (toim.), *JADT 2002. 6th International conference on textual data statistical analysis. Vol. 1* s. 21–32. Rennes: Inria.
- 2008: *Univariate, bivariate and multivariate methods in corpus-based lexicography – a study of synonymy*. Publications of the Department of General Linguistics, University of Helsinki 44. Helsinki: University of Helsinki.
- 2009: Monta tapaa ”ajatella” – tilastollisten menetelmien hyödyntäminen aineistolähtöisessä sanastontutkimuksessa. – *Virittäjä* 113 (2) s. 1–8, verkkoliite. – <http://ojs.tsv.fi/index.php/virittaja/article/view/4185/3902> – 29.12.2015.
- Baker, Paul, Hardie, Andrew & McEnery, Tony 2006: *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.
- Barnbrook, Geoff 1996: *Language and computers. A practical introduction to the computer analysis of language*. Edinburgh Textbooks in Empirical Linguistics. Edinburgh: Edinburgh University Press.
- Bednarek, Monika 2008: Semantic preference and semantic prosody re-examined. – *Corpus Linguistics and Linguistic Theory* 4 (2) s. 119–139.
- Begagić, Mirna 2013: Semantic preference and semantic prosody of the collocation *make sense*. – *Jezikoslovlje* 14 (2–3) s. 403–416.
- Biber, Douglas & Conrad, Susan 1999: Lexical bundles in conversation and academic prose. – Hilde Hasselgård & Signe Oksefjell (toim.), *Out of corpora. Studies in honour of Stig Johansson* s. 181–190. Amsterdam: Rodopi.
- Biber, Douglas, Conrad, Susan & Reppen, Randi 2006 [1998]: *Corpus linguistics. Investigating language structure and use*. Cambridge Approaches to Linguistics. Cambridge: Cambridge University Press.
- Clear, Jeremy 1993: From Firth principles. Computational tools for the study of collocation. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and technology. In honour of John Sinclair* s. 271–292. Philadelphia: John Benjamins.
- Cruse, D. Alan 1991: *Lexical semantics*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- 2000: *Meaning in language. An introduction to semantics and pragmatics*. Oxford: Oxford University Press.
- CSCa: *Finnish text collection. Description*. – <https://research.csc.fi/-/finnish-text-collection> 2.1.2016.
- CSCb: *Suomen sanomalehtikielen taajuussanasto*. Kielipankki, CSC – Tieteen tietotekniikan keskus Oy. – <https://korp.csc.fi/suomen-sanomalehtikielen-taajuussanasto-B9996.txt> 2.1.2016.
- CSCc: *WWW-Lemmie 2.0: User’s manual*. Saatavilla CSC:n käyttäjätunnuksilla Tutkijan käyttöliittymässä Lemmie-ohjelmassa osoitteen <https://sui.csc.fi/web/guest> kautta. 2.1.2016.
- Dixon, Robert M. W. 2004: Adjective classes in typological perspective. – Robert M. W. Dixon & Alexandra Y. Aikhenvald (toim.), *Adjective classes. A cross-linguistic typology* s. 1–49. Explorations in Linguistic Typology 1. Oxford: Oxford University Press.
- Firth, John Rupert 1957: *Papers in linguistics 1934–1951*. London: Oxford University Press.
- 1968: A synopsis of linguistic theory, 1930–55. – F. R. Palmer (toim.), *Selected papers of J. R. Firth 1952–59* s. 168–205. London: Longmans.

- Fraser, Bruce 1976: *The verb-particle combination in English*. Taishukan Studies in Modern Linguistics. New York: Academic Press.
- Granger, Sylviane & Meunier, Fanny 2008: Introduction. The many faces of phraseology. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology. An interdisciplinary perspective* s. XIX–XXVIII. Amsterdam: John Benjamins.
- Granger, Sylviane & Paquot, Magali 2008: Disentangling the phraseological web. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology. An interdisciplinary perspective* s. 27–49. Amsterdam: John Benjamins.
- Gries, Stefan Th. 2008: Phraseology and linguistic theory. A brief survey. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology. An interdisciplinary perspective* s. 3–26. Amsterdam: John Benjamins.
- Grönros, Eija-Riitta 2013: Sanakirjojen normatiivisuus ja deskriptiivisyys. – Laura Tyysteri & Kaisa Lehtosalo (toim.), *Hyvä sanakirja. Tieteellisiä, käytännöllisiä ja eettisiä näkökulmia leksikografiaan* s. 35–52. Turun yliopiston suomen kielen ja suomalais-ugrilaisen kielentutkimuksen oppiaineen julkaisuja 3. Turku: Turun yliopisto.
- Hakulinen, Auli 1996: Keskustelunanalyysin profiilista ja tilasta. – Auli Hakulinen (toim.), *Suomalaisen keskustelun keinoja II* s. 9–22. Helsinki: Helsingin yliopiston suomen kielen laitos.
- & Sorjonen, Marja-Leena 1989: Modaaliverbit spontaanissa keskustelussa. – Auli Hakulinen (toim.), *Suomalaisen keskustelun keinoja I* s. 73–97. Helsinki: Helsingin yliopiston suomen kielen laitos.
- Heikkilä, Tarja 2014a: *Tilastollinen tutkimus*. Helsinki: Edita.
- 2014b: *Tutkimustuki. Kvantitatiivinen tutkimus*. Tilastollinen tutkimus -kirjaan liittyvä verkkomateriaali. – *tilastollinentutkimus.fi* 8.12.2015.
- Hoey, Michael 2007: Lexical priming and literary creativity. – Michael Hoey, Michaela Mahlberg, Michael Stubbs & Wolfgang Teubert (toim.), *Text, discourse and corpora. Theory and analysis* s. 7–29. London: Continuum.
- Hoffmann, Sebastian, Fischer-Starcke, Bettina & Sand, Andrea 2013: Introduction. – *International Journal of Corpus Linguistics* 18 (1) s. 1–5.
- Isokääntä, Tarja 2014: *Kollokaatiot ja sanasto Erkki Laatikaisen kolumneissa*. Maisterintutkielma. Kielten laitos, Jyväskylän yliopisto. – <http://urn.fi/URN:NBN:fi:juu-201403051310> 29.12.2015
- Jaakola, Minna 2011: Epäspesifi evidentialiaali: nykysuomen *tiettävästi*. – *Virittäjä* 115 (4) s. 515–541.
- Jantunen, Jarmo Harri 2001: “Tärkeä seikka” ja “keskeinen kysymys”: Mitä korpuslingvistinen analyysi paljastaa lähisyronyymeista? – *Virittäjä* 105 (2) s. 170–192.
- 2004: *Synonymia ja käännössuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin*. Joensuun yliopiston humanistisia julkaisuja 35. Joensuu: Joensuun yliopisto.
- 2009a: Ei pelkästään mielikuvituksen puutteen vuoksi – Kieliaineistojen systemaattinen käyttö kielentutkimuksessa. – *Virittäjä* 113 (1) s. 101–113.
- 2009b: ”Minulla on aivan paljon rahaa” – fraseologiset yksiköt suomen kielen opetuksessa. – *Virittäjä* 113 (3) s. 356–381.
- & Brunn, Sisko 2012: Morfolooginen priming ja fraseologia vieraan kielen oppimisessa: korpustutkimus oppijansuomesta. – Annekatrin Kaivapalu, Pirkko Muikku-Werner, Jarmo Harri Jantunen & Maria-Maren Sepper (toim.), *Lähivõrdlusi. Lähivertailuja* 22 s. 71–100. Tallinn: Eesti Rakenduslingvistika Ühing.
- , Kumpulainen, Marjo, Tammimies, Tanja & Tokola, Teemu 2013: Korpuspohjaista oppijansanakirjaa tekemässä: esimerkkinä ConLexis. – Pirkko Muikku-Werner, Johanna Laakso, Katre Öim & Maria-Maren Sepper (toim.), *Lähivõrdlusi. Lähivertailuja* 23 s. 89–120. Tallinn: Eesti Rakenduslingvistika Ühing.
- Kangasniemi, Heikki 1992: *Modal expressions in Finnish*. Studia Fennica Linguistica 2. Helsinki: Suomalaisen Kirjallisuuden Seura.

- 1997: *Sana, merkitys, maailma. Katsaus leksikaalisen semantiikan perusteisiin*. Helsinki: Finn Lectura.
- Kanji, Gopal K. 1993: *100 statistical tests*. London: Sage Publications.
- Karlsson, Fred 2006: *Yleinen kielitiede*. Helsinki: Yliopistopaino.
- Kenny, Dorothy 2001: *Lexis and creativity in translation. A corpus-based study*. Manchester: St. Jerome Publishing.
- Kielitoimiston ohjepankki 2015a: *Pilkku*. – www.kielitoimistonohjepankki.fi/ohje/86 12.4.2015.
- 2015b: *Pilkku ja puhutteluilmaus, tarkennus tai muu lisäys*. – www.kielitoimistonohjepankki.fi/ohje/105 – 12.4.2015.
- Kivimies, Yrjö 1974: *Synonymisanasto*. Helsinki: Otava.
- Koivusalo, Esko 1992: Sanomalehti ja äidinkieli. – *Kielikello* (2) – <http://www.kielikello.fi/index.php?mid=2&pid=11&aid=846> – 7.1.2016.
- KS = *Kielitoimiston sanakirja*. 2014. Kotimaisten kielten keskuksen verkkojulkaisuja 35. Helsinki: Kotimaisten kielten keskus. URN:NBN:fi:kotus-201433, ISSN 2323-3370. Verkkojulkaisu HTML. Päivitettävä julkaisu. Päivitetty 11.11.2014. – 5.1.2016.
- Kuiri, Kaija 2012: *Johdatus semantiikkaan*. Helsinki: Finn Lectura.
- Laitinen, Lea 2004: Kieltosana ja Kieletär: Yhden kielikiistan kulku ja ideologiat. – Katja Huumo, Lea Laitinen & Outi Paloposki (toim.), *Yhteistä kieltä tekemässä. Näkökulmia suomen kirjakielen kehitykseen 1800-luvulla* s. 177–222. Suomalaisen Kirjallisuuden Seuran toimituksia 979. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Leech, Geoffrey 1990: *Semantics. The study of meaning*. Harmondsworth: Penguin Books.
- Lehtinen, Esa 1997: Episteemisyys, yksilö ja yhteisö. Eräiden modaali-ilmausten pragmatiikkaa. – *Virittäjä* 101 (2) s. 233–250.
- Lehtosalo, Kaisa & Tyysteri, Laura 2013: Sanakirjatyön valinnat ja leksikografian ammattietiikka. – Laura Tyysteri & Kaisa Lehtosalo (toim.), *Hyvä sanakirja. Tieteellisiä, käytännöllisiä ja eettisiä näkökulmia leksikografiaan* s. 9–20. Turun yliopiston suomen kielen ja suomalais-ugrilaisen kielentutkimuksen oppiaineen julkaisuja 3. Turku: Turun yliopisto.
- Leino, Antti, Hyvönen, Saara & Salmenkivi, Marko 2006: Mitä murteita Suomessa onkaan? Murren sanaston levikin kvantitatiivista analyysistä. – *Virittäjä* 110 (1) s. 26–45.
- Leino, Antti & Leino, Pirkko 1996: *Synonymisanasto*. Helsinki: Otava.
- Lindquist, Hans 2009: *Corpus linguistics and the description of English*. Edinburgh Textbooks on the English Language – Advanced. Edinburgh: Edinburgh University Press.
- Liu, Zhanyi, Wang, Haifeng, Wu, Hua & Li, Sheng 2009: Collocation extraction using monolingual word alignment method. – *Proceedings of the 2009 conference on empirical methods in natural language processing* s. 487–495. – <http://www.aclweb.org/anthology/D09-1051> 8.1.2016.
- Louw, Bill 1993: Irony in the text or insincerity in the writer? – The diagnostic potential of semantic prosodies. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and technology. In honour of John Sinclair* s. 157–176. Philadelphia: John Benjamins.
- Lyons, John 1968: *Introduction to theoretical linguistics*. London: Cambridge University Press.
- 1977: *Semantics: 2*. Cambridge: Cambridge University Press.
- 1996: *Linguistic semantics. An introduction*. Cambridge: Cambridge University Press.

- Makkonen-Craig, Henna 2005: *Toimittajan läsnäolo sanomalehtitekstissä. Näkökulmia suomen kielen dialogisiin passiivilauseisiin*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1026. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Martin, Willy, Al, Bernard & van Sterkenburg, Piet 1983: On the processing of a text corpus: from textual data to lexicographical information. – R. R. K. Hartmann (toim.), *Lexicography: principles and practice* s. 77–87. London: Academic Press.
- Mauranen, Anna 2004: Spoken corpus for an ordinary learner. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 89–105. Studies in Corpus Linguistics 12. Amsterdam: John Benjamins.
- Metin, Senem Kumova & Karaođlan, Bahar 2011: Measuring collocation tendency of words. – *Journal of Quantitative Linguistics* 18 (2) s. 174–187.
- Miettinen, Jorma 1984: *Toimitustyö. Journalistiksi suunnistautuvan oppikirja*. Helsinki: Gaudeamus.
- Muikku-Werner, Pirkko, Jantunen, Jarmo Harri & Kokko, Ossi 2008: *Suurella sydämellä ihan sikana. Suomen kielen kuvaileva fraasisanakirja*. Helsinki: Gummerus.
- Murphy, M. Lynne 2003: *Semantic relations and the lexicon. Antonymy, synonymy, and other paradigms*. Cambridge: Cambridge University Press.
- 2013: What we talk about when we talk about synonyms (and what it can tell us about thesauruses). – *International Journal of Lexicography* 26 (3) s. 279–304.
- Mäkinen, Kirsti 1992: Muuttuuko lehtikieli? Muuttaako lehtikieli? – Valma Yli-Vakkuri, Maija Länsimäki & Aarre Nyman (toim.), *Yhteiskunta muuttuu - Kieli muuttuu. Nykysuomen seuran 10-vuotisjuhlakirja* s. 211–235. Porvoo: WSOY.
- Nesselhauf, Nadja 2003: The use of collocations by advanced learners of English and some implications for teaching. – *Applied Linguistics* 24 (2) s. 223–242.
- Nissi, Riikka 2000: *Episteemisten, deonttisten ja dynaamisten modaaliainesten käyttö raamattupiirikeskusteluissa*. Pro gradu -tutkielma. Jyväskylän yliopiston suomen kielen laitos. – <http://urn.fi/URN:NBN:fi:jyu-2000859469> 29.12.2015
- Orpana, Terttu 1988: *Kuvaus vai kommentti. Tutkimus suomen kielen adjektiiviadverbien semanttisesta tulkinnasta*. Opera Fennistica & Linguistica 2. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
- Pajunen, Anneli 2001: *Argumenttirakenne. Asiaintilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Suomi 187. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Palmer, F. R. 1981: *Semantics*. Cambridge: Cambridge University Press.
- 1986: *Mood and modality*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Partington, Alan 2004: “Utterly content in each other’s company”. Semantic prosody and semantic preference. – *International Journal of Corpus Linguistics* 9 (1) s. 131–156.
- Päiviö, Pia 2007: *Suomen kielen asti ja saakka. Terminatiivisten partikkelien synonymia, merkitys, käyttö ja kehitys sekä asema kielioipissa*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 75. Turku: Turun yliopisto.
- Saukkonen, Pauli 1970: Puhekielen luonteesta. – *Kielikello* (3). – <http://www.kielikello.fi/index.php?mid=2&pid=11&aid=57> 7.1.2016.
- Sinclair, John 1991a: *Corpus, concordance, collocation*. Describing English Language. Oxford: Oxford University Press.

- 1991b: Shared knowledge. – James E. Alatis (toim.), *Linguistics and language pedagogy: The state of the art* s. 489–500. Georgetown University Round Table on Language and Linguistics. Washington D.C.: Georgetown University Press.
- 1998: The lexical item. – Edda Weigand (toim.), *Contrastive lexical semantics* s. 1–24. Amsterdam: John Benjamins.
- 2004: The search for units of meaning. – John Sinclair & Ronald Carter (toim.), *Trust the text. Language, corpus and discourse* s. 24–48. London: Routledge.
- 2008: Preface. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology. An interdisciplinary perspective* s. XV–XVIII. Amsterdam: John Benjamins.
- Stanojević, Maja 2009: Cognitive synonymy: a general overview. – *Facta Universitatis. Linguistics and Literature* 7 (2) s. 193–200. – <http://facta.junis.ni.ac.rs/lal/lal200902/lal200902-05.pdf> 7.1.2016.
- Stubbs, Michael 1995a: Collocations and semantic profiles: on the cause of the trouble with quantitative studies. – *Functions of Language* 2 (1) s. 23–55. – Uudelleen julkaistu osoitteessa <https://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf> s. 1–24. 12.12.2015.
- 1995b: Corpus evidence for norms of lexical collocation. – *Principle & practice in applied linguistics: studies in honour of H. G. Widdowson* s. 245–256. Oxford: Oxford University Press.
- 1996: *Text and corpus analysis. Computer-assisted studies of language and culture*. Language in Society 23. Oxford: Blackwell.
- 2001a: On inference theories and code theories: Corpus evidence for semantic schemas. – *Text* 21 (3) s. 437–465.
- 2001b: *Words and phrases. Corpus studies of lexical semantics*. Malden: Blackwell.
- 2002: On text and corpus analysis: A reply to Borsley and Ingham. – *Lingua* 112 (1) s. 7–11.
- Suoraniemi, Henriikka 2011: *Vuoksi ja takia. Korpuspohjainen tutkimus lähisynonyymeistä*. Pro gradu -tutkielma. Suomen kielen oppiaine, Oulun yliopisto.
- Tognini-Bonelli, Elena 2001: *Corpus linguistics at work*. Studies in Corpus Linguistics 6. Amsterdam: John Benjamins.
- Traugott, Elizabeth Closs 1989: On the rise of epistemic meanings in English: An example of subjectification in semantic change. – *Language* 65 (1) s. 31–55.
- Tsui, Amy B. M. 2005: ESL teachers' questions and corpus evidence. – *International Journal of Corpus Linguistics* 10 (3) s. 335–356.
- Tuuri, Emilia 2014: *Adpositiot kausaalisten suhteiden ilmaisussa. Lähimerkityksisyyden korpuspohjaista analyysia*. Pro gradu -tutkielma. Tampereen yliopisto. Kieli-, käännös- ja kirjallisuustieteiden yksikkö, suomen kieli.
- Vanhatalo, Ulla 2003: Kyselytestit vs. korpuslingvistiikka lähisynonyymien semanttisten sisältöjen arvioinnissa – mitä vielä keskeisestä ja tärkeästä? – *Virittäjä* 107 (3) s. 351–369.
- 2005: *Kyselytestit synonymin selvittämisessä. Sanastotietoutta kielenpuhujilta sähköiseen sanakirjaan*. Helsinki: Helsingin yliopisto, suomalais-ugrilainen laitos.
- Viinikainen, Taru 2010: *Taipuuko "akrobaatti Aleksandra"? Nimikekonstruktio ja nimikkeen taipuminen lehtikielessä 1900-luvulta 2000-luvulle*. Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology 1. Joensuu: Itä-Suomen yliopisto.
- VISK = Auli Hakulinen, Maria Vilkkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. – <http://scripta.kotus.fi/visk> 5.1.2016

Weinreich, Uriel 1980: *On semantics*. Toim. William Labov & Beatrice S. Weinreich. Philadelphia: University of Pennsylvania Press.

Zgusta, Ladislav 1971: *Manual of lexicography*. *Janua linguarum. Series maior* 39. Prague: Academia.

Aineistolähteet

Aamulehti 1995. Sähköinen 3,8 miljoonan saneen laajuinen suomen kielen tekstikokoelma. Koostajat: Kotimaisten kielten tutkimuskeskus; CSC – Tieteellinen laskenta Oy. Käytettävissä Kielipankin kautta. URN: <http://urn.fi/urn:nbn:fi:lb-201403268>.

Karjalainen 1998. Sähköinen 6 miljoonan saneen laajuinen suomen kielen tekstikokoelma. Koostajat: Yleisen kielitieteen laitos, Helsingin yliopisto; Yleinen kielitiede, Joensuun yliopisto; CSC – Tieteellinen laskenta Oy. Käytettävissä Kielipankin kautta. URN: <http://urn.fi/urn:nbn:fi:lb-201403268>.

LIITE 1

T-testin mukaiset tilastollisesti merkitsevät kollokaatit (lihavoitu ne, jotka kummallekin sanalle yhteisiä).

kollokaatti (EHKÄ)	t-testin arvo	kollokaatin lukumäärä (vasemmalla/oikealla puolella noodista)	kollokaatti (MAHDOLLISESTI)	t-testin arvo	kollokaatin lukumäärä (vasemmalla/oikealla puolella noodista)
mutta	9,18	242 (113/129)	jopa	5,57	37 (3/34)
se	8,63	424 (103/321)	jo	5,31	53 (4/49)
ei	8,07	434 (321/113)	tai	4,64	40 (19/21)
hieman	7,18	65 (4/61)	ensi	4,62	32 (5/27)
jokin	7,09	85 (16/69)	tulla	3,66	43 (20/23)
vähän	7,05	77 (8/69)	liittyä	3,56	17 (2/15)
liian	6,58	61 (1/60)	henkilö	3,14	13 (2/11)
paljon	6,45	120 (21/99)	voida	3,04	38 (21/17)
siksi	5,94	47 (9/38)	aiheuttaa	2,93	12 (1/11)
vielä	5,90	110 (18/92)	vielä	2,92	25 (6/19)
kaikkein	5,68	36 (0/36)	ja	2,88	191 (160/31)
hyvä	4,78	112 (10/102)	muu	2,82	24 (4/20)
tai	4,77	96 (50/46)	syksy	2,71	12 (5/7)
joskus	4,77	33 (2/31)	aiheutua	2,70	8 (1/7)
joku	4,65	37 (9/28)	seurata	2,60	13 (2/11)

jopa	4,58	45 (3/42)	myydä	2,28	9 (2/7)
syy	4,35	42 (9/33)	jokin	2,24	12 (3/9)
liikaa	4,31	24 (0/24)	perustaa	2,21	8 (1/7)
johtua	4,29	26 (9/17)	sekä	2,18	19 (15/4)
yllättää	4,27	26 (2/24)	ne	2,17	24 (13/11)
hiukan	4,25	22 (2/20)	myöhemmin	2,14	7 (4/3)
[pilkku]	4,23	1482 (697 / 785)	havainto	2,10	5 (1/4)
hyvin	4,20	68 (6/62)	siirtyä	2,08	8 (4/4)
turhan	4,00	18 (0/18)	2000	2,08	6 (1/5)
sittenkin	3,88	17 (4/13)	uusi	2,01	24 (9/15)
sitten	3,81	62 (9/53)	tapahtua	2,00	8 (2/6)
koskaan	3,76	24 (0 / 24)			
me	3,74	60 (23/37)			
no	3,47	15 (15/0)			
ymmärtää	3,41	23 (2/21)			
tuntea	3,18	30 (9/21)			
kuin	3,06	92 (34/58)			
minä	2,99	37 (13/24)			
niin	2,96	74 (21/53)			
viisas	2,96	11 (3/8)			
tulevaisuus	2,91	23 (11/12)			
voida	2,90	121 (47/74)			

kuluttua	2,88	14 (1/13)
tulla	2,77	118 (46/72)
suuri	2,77	72 (25/47)
kuitenkin	2,73	56 (12/44)
odottaa	2,67	29 (10/19)
myöhemmin	2,63	17 (9/8)
silloin	2,62	22 (14/8)
tärkeä	2,60	32 (4/28)
kova	2,55	24 (4/20)
paha	2,54	18 (1/17)
tässä	2,53	15 (0/15)
kymmenen	2,49	26 (4/22)
he	2,44	53 (28/25)
parikymmen- tä	2,43	10 (0/10)
lukuunotta- matta	2,39	9 (0/9)
aavistus	2,35	6 (0/6)
sata	2,35	18 (1/17)
vaikuttaa	2,28	21 (4/17)
osittain	2,27	9 (1/8)
ajatella	2,24	15 (5/10)
pian	2,18	11 (1/10)

muutama	2,18	25 (3/22)
luulla	2,17	9 (0/9)
hanakasti	2,16	5 (0/5)
ensi	2,15	41 (8/33)
koko	2,12	40 (7/33)
kaivata	2,11	11 (4/7)
tämä	2,09	119 (51/68)
täällä	2,09	16 (4/12)
huomata	2,08	13 (2/11)
osata	2,08	16 (1/15)
tarpeeksi	2,02	8 (1/7)

LIITE 2

MI-testin mukaiset tilastollisesti merkitsevät kollokaatit (lihavoitu ne, jotka kummallekin sanalle yhteisiä).

kollokaatti (EHKÄ)	MI-testin arvo	kollokaatin lukumäärä (vasemmalla/oikealla puolella noodista)	kollokaatti (MAHDOLLISESTI)	MI-testin arvo	kollokaatin lukumäärä (vasemmalla/oikealla puolella noodista)
hanakasti	4,87	5 (0/5)	aiheutua	4,47	8 (1/7)
aavistus	4,60	6 (0/6)	havainto	4,08	5 (1/4)
kaikkein	4,21	36 (0/36)	jopa	3,58	37 (3/34)
turhan	4,12	18 (0/18)	henkilö	2,96	13 (2/11)
sittenkin	4,11	17 (4/13)	liittyä	2,87	17 (2/15)
hiukan	3,40	22 (2/20)	2000	2,74	6 (1/5)
no	3,27	15 (15/0)	aiheuttaa	2,69	12 (1/11)
hylky	3,22	5 (5/0)	ensi	2,44	32 (5/27)
viisas	3,21	11 (3/8)	myöhemmin	2,40	7 (4/3)
hieman	3,19	65 (4/61)	vai	2,38	6 (5/1)
liikaa	3,05	24 (0/24)	perustaa	2,20	8 (1/7)
siksi	2,90	47 (9/38)	syksy	2,19	12 (5/7)
paras mahdollinen	2,84	5 (0/5)	pyytää	2,14	5 (1/4)
sen takia	2,69	5 (0/5)	etsiä	2,09	6 (3/3)
liian	2,66	61 (1/60)	myydä	2,07	9 (2/7)
johtua	2,66	26 (9/17)	siirtyä	1,93	8 (4/4)

sopeutua	2,65	5 (1/4)	tai	1,91	40 (19/21)
yllättää	2,62	26 (2/24)	joku	1,89	7 (3/4)
oleellinen	2,57	5 (0/5)	jo	1,88	53 (4/49)
joskus	2,56	33 (2/31)	tulo	1,88	7 (2/5)
useasti	2,48	5 (1/4)	tulevaisuus	1,84	7 (0/7)
vähän	2,34	77 (8/69)	seurata	1,83	13 (2/11)
toisin	2,33	5 (0/5)	käsitellä	1,83	5 (1/4)
lukuunot- tamatta	2,29	9 (0/9)	tapahtua	1,78	8 (2/6)
hymyillä	2,17	5 (2/3)	osallistua	1,75	7 (2/5)
kuluttua	2,12	14 (1/13)	vaihe	1,64	5 (2/3)
jokin	2,11	85 (16/69)	lisätä	1,63	8 (3/5)
parikym- mentä	2,10	10 (0/10)	johtaa	1,62	8 (2/6)
koskaan	2,10	24 (0/24)			
joku	2,08	37 (9/28)			
jotenkin	2,08	5 (0/5)			
kuuluisa	2,06	5 (0/5)			
osittain	2,03	9 (1/8)			
kuulija	2,02	5 (2/3)			
osaltaan	2,01	6 (1/5)			
hahmo	2,00	5 (3/2)			
mökki	1,94	5 (5/0)			
osin	1,93	5 (1/4)			
sentään	1,87	7 (0/7)			
luulla	1,86	9 (0/9)			
tarpeeksi	1,80	8 (1/7)			
ymmärtää	1,79	23 (2/21)			
tahto	1,78	5 (2/3)			
400	1,74	7 (2/5)			

siirto	1,74	5 (4/1)			
odotella	1,67	5 (3/2)			
jopa	1,65	45 (3/42)			
mielen- kiintoinen	1,63	8 (2/6)			
kuvitella	1,61	6 (0/6)			
syy	1,60	42 (9/33)			
jonkin verran	1,59	6 (0/6)			

LIITE 3

Sanojen luokittelu semanttisiin ryhmiin semanttisen preferenssin analyysia varten.

	EHKÄ	MAHDOLLISESTI
'tila'	oli, eloon, jääneet, olisi, oli, omimmillaan, ovat, esiintyisi, olla, on, on, puuttuvat, on, olla, ovat, onkin, on, omankin, olisi, on, olisi, on, ollut, ollut, on, olivat, oli, piilee, omat, olikin, oleskelleet, sopinut, on	olevien, lomailla, oleilevasta, olevalla, ollut, on, ovat, ole, oliko, olleet, löydy, olevat, säilyneestä, oleviin, olleiden, onkin, jäljelläolevia, olisi, on, ollaan
'prosessi'	opetus, siirtyisi, rakennetta- viin, saanut, saa, kääntää, kor- jattu, liikkeellä, saatu, saa, käydään, saan, katoamiseen, toilailu, tekee, käytettyäkin, toimineille, menettää, teke- mään, painua, liikuteltava, muuttaisin, mennä, menestys, perhostutkimuksen, ryöstäy- tyminen, estää, tekeviin, teh- dä, korjaan	varmistaneet, tulevalla, teroit- taa, antamisesta, menossa, muodostettaisiin, tulivat, tulla, saa, jäävänsä, ostetaan, nur- metetaan, myyneet, toimin- nasta, höystettynä, ohjelmalla, hakea, järjestyy, etsimään, saavat, käytävissä, palaavan- sa, antaa, vietäväksi, kytke- tään, selvittämään, saastuneet, leivottuja, vapautuksen, toteu- tuneille, vahvistettavaan, kui- vumassa, käytetyistä, tuleeko, perääntyessä, pelastava, pakottanut, kartoittamaan, käytön, laatia, tulossa
'mentaalisuus'	kuvan, katsoo, pyrkimyksiä, tyytymättömiä, johtopäätök- sen, nähtyä, valmiudet, so-	kiinnostuneiden, hätäillä, ha- luavansa, masentamille, näh- dään, kantansa, katsotaan,

	peudu, yksitoikkoiseenkin, kannalta, hyväksy, harmita, tunnetaan, henkisesti, ymmärtävät, stressaantuneita, samaistuneet, surullisinta, havaita, viehätys, yleistetään, luutumien, haluaa, päätelmänsä, harmittaa, lukutaidotonta, luullakseni, arvellut	haluaisimme, opiskelevista, kiinnostunut, järki, väärinymmärrykset, epäillään
'viestintä'	ilmaistessaan, sanoi, TV-uutiset, media, mellastaa, sanoma, Metsäradiossa	vastasi, sanoen, sitaatti, sanoi, ilmaista, julkaisi, todetaan, osapuolet, kertoi, ilmoituksen, velvoitettu, anomuksesta, määrätystä, lisäpyyntöjä, ohjeet, vaatimukset, rohkaiseva, kaiut
'aika'	2003, nyt, jatkuvalla, huhtitoukokuun, vaihteessa, ensi, kevään, nyt, silloin, pitkäjänteiseen, silloin, valmis, jo, 1988, välissä, koskaan, antiikin, vaiheessa, joudamme, vielä, loppujen, lopuksi, lykäämään, ajankohtaa, tuonemmaksi, seuraavaksi, syyskuussa, marraskuussa, joulukuussa, vanhojen, päiväkausia, joskus, sen_hetkiseen [senhetkiseen], aina, pieniä	mennessä, uusi, jo, jo, syksyllä, kevään, lauantai-iltana, ensimmäisellä, entinen, kesäkuun, vuoden, päästä, uuden, ennen, tuleva, joskus, nyt, piankin, vielä, vuoden, aikana, jo, jo, tulevillekin, sairastamisajasta, tiistaihin, vielä, jälkeen
'paikka'	Wärtsilän, lähetyville, Sveitsistä, Helsingin, alueelle, Sak-	ilmassa, Tampereella, Suomeen, Lieksan, Saarijärvellä,

	sassa, Suomessa, täällä, Eurooppaan, kalmisto, Suomen, Moskovan, asunnossa	pois, Geneven, matkalla, Moskovaan, kotiin, tehdasalueella, alueet, edellä, saunassa, paikallisella
'määrä, intensiteetti'	enemmän, erityisen, 30, ylimääräistä, enemmän, terävimmän, 1200, metrissä, osittain, vielä, korkeammallakin, liikakasvu, kaikesta, vähän, ahkerammin, kaikkein, jarrutusmatka, lyhenekään, helpommin, ainoa, pelkästä, kaikkein, syvin, hyvinkin, juuri, vähän, suuri, osa, kaksi, viidesosaa	määrä, lähes, puolet, 6–7, kooltaan, pieniä, jopa, 785, muutaman, jopa, 100 miljoonan, riittävästi, kilometriä, 40, kilometriä, 15, jopa, puolet, erityisesti, jopa, pieniin, erittäin, suhteellisen, runsaasti, lisätä, sahausmääriä, vieläkin, lisää
'ihmiset'	lähiomaiset, työ- [työkaverit], koulukaverit, Riikan, Kaupin, säveltäjän, vastustajat, esittäjätkin, Jari, suomalaisista, henkilön, ohjaaja, hän, nuorelta, meilläkin, Jere Lehtinen, Juha Lind, Esa Tikkanenkin, järjestäjien, aluepäällikkö, Leena Nurmela, me, pojillekin, Topeliusta, lapsia, hänet, miehiin	henkilöllä, Ilkka, Vasala, runoilija, kaksikko, seuraa, psykologin, miehellä, Wood, jäsenet, ihmisen, he, hänen, oppilaasta, avustajia, henkilöitä, hän, oppilasta, Clintonin
'terveydentila'	kilpirauhassyöpätapausta, vammaansa, sairaita	rabies, tauteja, taudista, lääkkeitä, terve
'urheilu, kilpailu'	siirto, ylisuojausta, MM-kisat, juoksee, vauhdin, sinkauttaa, maajoukkuevalintatavalla,	Crystal P, Barnsley, putoamisensa, olympiakisojen

	leijonien, tasapeli, sauvarinteeseen, turnaus, Izvestija	
'konkreettiset oliot'	kuorma-autoja, moottorialus, tsasouna, hirveen	sieni-itiöiden, idänkäestä, lintupiippariin, koptereita, kuoretta, maistiaisiksi, karhulla, lineaarikiihdytin, kuoretta, maistiaiseksi, Hornetien, koneita, linja, pääverkostoon, hävittäjälentäjien, kuvat, koneiden, pienoismallit, SB2 pommituskoneesta, sähkökiuas, tukilla
'kulttuuri'	kuorojen, pelejä, mayakulttuurin, sadun, museaalista, ortodoksinen, elokuvassa, kulttuurikaupunki, Linna	romaaneeja, rooleihin, Ilmailumuseo, perinnemaisemia
'yhteiskunnallisuus, institutionaalisuus'	rahaa, työstä, hankkeen, työhön, väestön, rintamatehtäväsä, sopimusvaihtoehdoista, vaalien, Keski-Karjalan, työtä, turvallisuusneuvosto, lakon, äänestäjistä, yhteistyöllä	osuuskunnan, jäsenyydestä, perheenyhdistämisen, oikeustalohanke, valtion, kiinteistölaitoksen, syytettyjen, lahjoman, pörssiin, suuryksikköä, korotuksia, palvelut, vuokralla, kierroksella, keskustalainen, pääministeri, Kiinalle, Pohjoismaiden, Metsähallituksen, tulosalueita, yhtiöittää, KOK:n, ääniään, hakijakaupungeille, maahantuloaikeeni, lailliset, EU-maiden, valuutan, välikohtauksia, kouluja, maiden, rauhansopimuksen, rau-

		hanneuvotteluissa, ilmataistelun, puolustuspolitiikan, puolustuksen, kiristäminen, leikkaa, inflaatio-odotuksia, kiinteistöveroista, suojeleohjelmiin, Naturaan, asianosaisilta, poliisin, valvontaa, säästöjä, kaupungin, henkilösuojan, ympäristöalan, yhteistyösopimus, neuvotteluista, jutun, penkille
'epävarmuus'	jostain, mikäli, jonkun, jossain, saattoi, jossakin, voisin, saatatte, miksei, jos, tai (2)	selviäisi, lähinnä, miten, voi, jos, vaihtoehtona, ainakin, voidaan, oljenkorsi, tai (3)
'additiivisuus'	taas, niin kuin [niin X kuin Y:kin], muuten, ja (12)	muistakin, toisena, yhdessä, muiden, kanssa, lisäksi, yhteisen, muita, muidenkin, yhteisen, myös (3), sekä (2), ja (15)
'kausaalisuus'	kun, siksin, siksi, syynä	jotta, joten, myötä, vaikuttaa, koska, taustalla, aiheutuvista, aikaan, aiheuttaneen, toisi, varten, tiimoilta
'vastakkaisuus'	kuitenkin, sittenkin, sittenkään, mutta (7)	kuitenkaan, mutta (2)
'arvo, laatu'	mustan, vakavaa, merkityksestä, parhaimman, tyyli, kannattaa, vois, paras, perustavat, tärkein, kunnianteko, raskaimmassa, kannaisi ['kannattaisi'], mieluummin, arvoa,	kannata, haitoista, vahingoista, repäisevällä, kannattaa, yhtenäisen, parhaiten, erilliset, tekniset, hankalampaa, tiukimmin, vieras, pääsee

	huonoiten, parempi, edullisin, kannattaisi, ensiluokan, puutteen	
luokittelemattomat	josta, olisi, ei, tämä, siitä, on, olisi, kokoon, on, että, eivät, alkavat, ei, tällainen, sitä, tulee, ei, kuin, niiden, parissa, jotka, se, asia, joka, ei, kyse, saa, siinä, varassa, käsistä, voidaan, ei, että, sitä, olivat, ei	siihen, ei, sitä, ei, voi, on, ne, tapauksen, ei, itse, ovat, eivätkä, piirissä, jossa, että, kuin, olisivat, tämän, niihin, niille, että, tutkimussaavutuksilleen, jonka, on, tulisi, sen, niitä