

**HAVAITA- JA HUOMATA-KVASIRAKENTEIDEN KOLLO-  
KAATTI JA LÄHISYNONYMIA SUOMI24-KESKUSTE-  
LUISSA**

Aaro Kumpulainen  
Maisterintutkielma  
Suomen kieli  
Kieli- ja viestintätieteiden  
laitos  
Jyväskylän yliopisto  
Kevät 2021



# JYVÄSKYLÄN YLIOPISTO

Tiedekunta Humanistis-yhteiskuntatieteellinen	Laitos Kieli- ja viestintätieteiden laitos
Tekijä Aaro Kumpulainen	
Työn nimi HAVAITA- ja HUOMATA-kvasirakenteiden kollokaatit ja lähisynonymia Suomi24-keskusteluissa	
Oppiaine Suomen kieli	Työn laji Maisterintutkielma
Aika Toukokuu 2021	Sivumäärä 50 + liitteet
Tiivistelmä	
<p>Tutkin maisterintutkielmassani HAVAITA- ja HUOMATA-kvasirakenteiden (esim. <i>on havaitsevinaan, olen huomaavinani</i> jne.) merkitseviä kollokaatteja ja lähisynonymiaa korpuslingvistiikan ja tilastollisten menetelmien avulla. Käyttämäni Suomi24-korpuksen laajuus on noin 3,5 miljardia sanetta. Tutkimukseni tavoitteena oli selvittää, millaisia merkitseviä kollokaatteja HAVAITA- ja HUOMATA-kvasirakenteet saavat suomenkielisissä verkkokeskusteluissa sekä mitä näiden kollokaattien avulla voidaan päätellä kyseisten kvasirakenteiden välisestä synonymiasuhteesta. Tutkimukseni teoreettisen viitekehyksen muodostavat kontekstuaalinen semantiikka, fraseologia ja synonymia. Tarkastelin kollokaatioanalyysissäni ensin molempien kvasirakenteiden yhteisiä merkitsevimpiä kollokaatteja ja sen jälkeen sellaisia kollokaatteja, jotka ovat merkitseviä vain jommankumman kvasirakenteen kontekstissa. Hyödynsin kollokaatioanalyysissäni rinnakkain määrällisiä ja laadullisia menetelmiä: Määrällinen analyysini perustuu t- ja MI-testien merkitsevyysarvoihin, joiden avulla tarkastelin kollokaatteja vertailevalla otteella. Laadullisessa analyysissäni käsitelin sekä kollokaattien kieliopillisia että niiden semanttisia seikkoja.</p> <p>Kollokaatioanalyysini tulosten mukaan HAVAITA- ja HUOMATA-kvasirakenteet kollokoivat vahvasti abstraktien substantiivien kanssa. Erityisen merkitseviä molempien kvasirakenteiden kollokaatteja ovat ivalisuuteen viittaavat jaolliset substantiivit IRONIA ja SARKASMI. Muita erityisen merkitseviä molempien kvasirakenteiden kollokaatteja olivat OLLA, ETTÄ, JOKIN ja HIENOINEN. Näitä kollokaatteja voidaan siis pitää HAVAITA- ja HUOMATA-kvasirakenteiden välistä synonymiaa yhdistävinä piirteinä. Vastaavasti kyseisten kvasirakenteiden välistä synonymiaa rikkovina tekijöinä kollokaatioanalyysissäni korostui se, että HUOMATA-kvasirakenne kollokoi selvästi voimakkaammin verbien kanssa ja ihmisiin viittaavien pronomien ja substantiivien kanssa, kuten HÄN, MINÄ, NAINEN ja MIES. Erityisen vahvasti synonymiaa rikkoo kieltoverbi EI, joka kollokoi hyvin voimakkaasti HUOMATA-, mutta ei HAVAITA-kvasirakenteen kanssa. Lisäksi itse verkkokeskustelemisesta keskustelemiseen viittaavat kollokaatit, kuten KIRJOITUS, PALSTA ja KOMMENTTI kollokoivat voimakkaammin HAVAITA- kuin HUOMATA-kvasirakenteen kanssa. Tutkimukseni tuloksia voidaan hyödyntää esimerkiksi kielenopetuksessa, sanakirjatyössä tai kielenkäytön kuvauksissa, kun keskitytään kieliopillisiin verbirakenteisiin.</p>	
Asiasanat kvasirakenne, fraseologia, kollokaatio, synonymia, korpuslingvistiikka	
Säilytyspaikka Jyväskylän yliopisto	
Muita tietoja	



## KUVIOT

KUVIO 1	Kielitoimiston sanakirjan (KS) selitteet ja esimerkkilauseet hakusanalle <i>havaita</i> . ....	14
KUVIO 2	Kielitoimiston sanakirjan (KS) selite ja esimerkkilauseet hakusanalle <i>huomata</i> . ....	14

## TAULUKOT

TAULUKKO 1	HAVAITA- ja HUOMATA-kvasirakenteiden 40 merkitsevintä kollokaattia MI-arvon mukaan (molemmille kvasirakenteille yhteiset kollokaatit lihavoitu). ....	25
TAULUKKO 2	HAVAITA- ja HUOMATA-kvasirakenteiden 40 merkitsevintä kollokaattia MI-arvon mukaan (molemmille kvasirakenteille yhteiset kollokaatit lihavoitu). ....	27
TAULUKKO 3	HAVAITA-kvasirakenteiden 40 merkitsevintä kollokaattia MI-arvon mukaan (HAVAITA-kvasirakenteelle [HUOMATA-kvasirakenteeseen verrattuna] paljon tyypillisemmät kollokaatit lihavoitu; vain HAVAITA-kvasirakenteelle tyypilliset kollokaatit merkitty tähdellä *). ....	30
TAULUKKO 4	HAVAITA-kvasirakenteiden 40 merkitsevintä kollokaattia t-arvon mukaan (HAVAITA-kvasirakenteelle [HUOMATA-kvasirakenteeseen verrattuna] paljon tyypillisemmät kollokaatit lihavoitu; vain HAVAITA-kvasirakenteelle tyypilliset kollokaatit merkitty tähdellä *). ....	33
TAULUKKO 5	HUOMATA-kvasirakenteiden 40 merkitsevintä kollokaattia MI-arvon mukaan (HUOMATA-kvasirakenteelle [HAVAITA-kvasirakenteeseen verrattuna] paljon tyypillisemmät kollokaatit lihavoitu). ....	36
TAULUKKO 6	HUOMATA-kvasirakenteiden 40 merkitsevintä kollokaattia t-arvon mukaan (HUOMATA-kvasirakenteelle [HAVAITA-kvasirakenteeseen verrattuna] paljon tyypillisemmät kollokaatit lihavoitu; vain HUOMATA-kvasirakenteelle tyypilliset kollokaatit merkitty tähdellä *). ....	39

# SISÄLLYS

1	JOHDANTO .....	2
2	TEORIATAUSTA.....	4
2.1	Kvasirakenne .....	4
2.2	Kontekstuaalinen semantiikka ja fraseologia.....	7
2.2.1	Kontekstuaalinen semantiikka .....	7
2.2.2	Fraseologia.....	7
2.2.3	Fraseologinen yksikkö .....	8
2.2.4	Kollokaatio .....	10
2.3	Synonymia.....	11
2.3.1	Synonymian lajeja.....	11
2.3.2	Verbien HAVAITA ja HUOMATA synonymia.....	13
2.3.3	Aiempiä fraseologisia korpustutkimuksia suomenkielisten sanojen synonymiasta.....	15
3	AINEISTO JA METODIT.....	16
3.1	Aineisto.....	16
3.2	Korpuslingvistiikka .....	17
3.3	Kollokaatioanalyysi ja tilastolliset testit.....	18
3.3.1	t-testi.....	21
3.3.2	MI-testi .....	21
3.3.3	Esimerkki t- ja MI-testien laskemisesta .....	22
3.3.4	Tilastollisten testien tulosten kokoaminen ja kollokaattien laadullinen analyysi .....	22
4	TULOKSET .....	24
4.1	HAVAITA- ja HUOMATA-kvasirakenteiden yhteiset merkitsevät kollokaatit .....	24
4.1.1	MI-testi .....	24
4.1.2	t-testi.....	26
4.2	HAVAITA-kvasirakenteen merkitsevät kollokaatit.....	30
4.2.1	MI-testi .....	30
4.2.2	t-testi.....	33
4.3	HUOMATA-kvasirakenteen merkitsevät kollokaatit.....	35
4.3.1	MI-testi .....	35
4.3.2	t-testi.....	39
5	PÄÄTÄNTÖ.....	44
5.1	Yhteenvedo.....	44

5.2 Tutkimuksen arviointi ja pohdinta .....	45
LÄHTEET .....	48
LIITTEET	

# 1 JOHDANTO

Tutkin maisterintutkielmassani Suomi24-verkkokeskusteluissa esiintyviä HAVAITA- ja HUOMATA-kvasirakenteita<sup>1</sup> vertailevalla otteella. Suomen kielen kvasirakennetta (*olla olevinaan/tekevinään jne.*) on toistaiseksi tutkittu suhteellisen vähän. Kattavimmin sitä on tutkinut Salminen (2000), joka keskittyy väitöskirjassaan kvasirakenteen merkitykseen, käyttöön ja kehitykseen ja jonka näkökulmana on erityisesti moniäänisyyden käsite ja teoreettisena kehyksenä kieliopillistuminen.

Verbit HAVAITA ja HUOMATA ovat toistensa lähisyronyymejä (ks. KS s.v. *havaita*, s.v. *huomata*). Tutkielmani tavoitteena on tarkastella näistä kahdesta verbistä muodostettujen kvasirakenteiden välisiä eroja ja yhtäläisyyksiä korpusaineiston avulla. Kyseiset verbit ovat Salmisen (2000: 45–46) ja Kumpulaisen (2019) mukaan kvasirakenteissa poikkeuksellisen tyypillisiä, ja niiden esiintymisfrekvenssi kvasirakenteen vartaloverbinä on Suomi24-verkkokeskusteluissa hyvin korkea. Siksi nimenomaan HAVAITA- ja HUOMATA-kvasirakenteet ovat soveltuva ja kiinnostava valinta tutkimukseni kohteiksi, ja niiden tutkiminen voi paljastaa suhteellisen paljon kvasirakenteesta ylipäättään. Kvasirakenteiden runsaamman ja monipuolisemman tutkimisen avulla lingvistinen ymmärrys suomen kielestä, sen verbirakenteista sekä synonymiasta syvenee ja täsmentyy.

Korpuslingvistiikka soveltuu erityisen hyvin synonymian tutkimiseen, sillä korpukset antavat kattavasti tietoa todellisesta kielenkäytöstä (Biber, Conrad & Reppen 2006: 43). Suomenkielisten sanojen välisiä synonymiasuhteita onkin tutkittu korpusten ja fraseologisen viitekehyksen kautta jo jonkin verran (ks. esim. Jantunen 2001; 2004; 2009a), mutta verbien HAVAITA ja HUOMATA synonymiaa tai eri vartaloverbeistä muodostettujen kvasirakenteiden välisiä synonymiasuhteita ei tiettävästi ole tutkittu aikaisemmin. Synonymiaa käsittelevät tutkimukset harvoin keskittyvät tarkastelemaan sanojen käyttöä ja semantiikkaa vain tietyn kieliopillisen rakenteen kontekstissa, joten tämä tutkimukseni erottuu muista tutkimuksista paitsi

---

<sup>1</sup> Käytän tässä tutkielmassani korpustutkimuksessa vakiintunutta merkintätapaa, jossa sanojen lemmamuodot kirjoitetaan kapiteelein.



tarkasteltavien lekseemien valinnan myös spesifisen kieliopillisen, kvasirakenteisiin keskittyvän rajauksensa osalta.

Työni pohjautuu kontekstilähtöiseen kielentutkimukseen, ja sen teoreettisen viitekehyksen muodostavat kontekstuaalinen semantiikka, fraseologia ja synonymia. Hyödynnän työssäni paitsi korpuslingvistisiä menetelmiä myös merkitsevimpien kollokaattien laadullista luokittelua kieliopillisesti ja semanttisesti sekä tilastollisia menetelmiä.

Tutkimuskysymykseni ovat:

1. Millaisia merkitseviä kollokaatteja HAVAITA- ja HUOMATA-kvasirakenteilla on?
2. Mitä kollokaatit kertovat HAVAITA- ja HUOMATA-kvasirakenteiden synonyymisyydestä?

Ensimmäinen tutkimuskysymykseni keskittyy siis kahden eri kvasirakenteen tekstiympäristöihin eli koteksteihin ja niille tyypillisiin kielellisiin valintoihin. Toinen tutkimuskysymykseni puolestaan pohjautuu kahteen ensimmäiseen tutkimuskysymykseeni ja hakee johtopäätöksiä, joita näiden pohjalta voidaan tehdä. Tutkimukseni tulosten jäsentely etenee tilastollisesta näkökulmasta laadullisiin tulkintoihin ja johtopäätöksiin. Ensimmäisessä tutkimuskysymyksessä korostuu enemmän määrällisen tutkimuksen näkökulma, kun taas toisessa tutkimuskysymyksessä esillä ovat lähinnä laadulliset päätelmät.

Tutkimukseni hypoteesina on, että kollokaatioanalyysissa nousee esiin sekä HAVAITA- ja HUOMATA-kvasirakenteiden synonymiaa vahvistavia että sitä rikkovia kollokaatteja. Hypoteesini perustuu aiempiin fraseologisiin, korpuslingvistisiin synonymiatutkimuksiin, joiden tulokset osoittavat, että synonyymeinä pidettävien sanojen fraseologisuus on osittain erilaista keskenään. Tutkimuksissa on havaittu lähisynonyymien välillä eroja esimerkiksi myöntö- tai kieltihakaisuudessa (Jantunen 2004: 221–222) ja useiden kollokaattien semanttisissa piirteissä (Jantunen 2009a: 373) sekä vastaavasti lähisynonyymien välisinä yhdistävinä tekijöinä esimerkiksi kollokaattien muodostamia yhteisiä merkitysryhmiä (Pirkola 2016; Laakkonen 2017). Kerron aiempien tutkimusten havainnoista lisää kohdassa 2.3.3.

Tutkielmani etenee siten, että luvussa 2 esittelen tutkimukseni teoreettisen taustan, joka sisältää tietoa kvasirakenteesta, kontekstuaalisesta semantiikasta, fraseologiasta ja siihen liittyvistä käsitteistä. Teorialuvussa esittelen myös tiiviisti joitakin aiempia tutkimuksia, jotka käsittelevät samoja teemoja kuin tämä tutkimukseni. Luvussa 3 kerron tutkimusaineistostani ja kuvailen käyttämiäni määrällisiä ja laadullisia menetelmiä, joilla käsittelin ja analysoin aineistoani. Tutkielman luku 4 on analyysiluku. Siinä analysoin HAVAITA- ja HUOMATA-kvasirakenteiden merkitseviä kollokaatteja monipuolisesti MI- ja t-testien avulla sekä laadullisten havaintojen ja tulkintojen kautta. Lopuksi luvussa 5 teen tuloksistani koontia ja johtopäätöksiä, pohdin tulosten yleistettävyyttä ja rajoituksia sekä suhteutan tutkimukseni laajempaan tutkimuskenttään.

## 2 TEORIATAUSTA

Tutkielmani teoriaosuudessa esittelen ensin yleisesti kvasirakennetta ja sen käyttöä ja merkityksiä. Seuraavaksi kerron, mitä fraseologia teoreettisena viitekehyksenä merkitsee tämän tutkimuksen kannalta, ja määrittelen siihen liittyviä käsitteitä. Sen jälkeen selitän kontekstuaalisen semantiikan käsitteen sekä sen suhteen tähän tutkimukseen. Teoriaosuuden lopuksi esittelen yleisellä tasolla sanojen merkitysten välisiä suhteita *synonymian* käsitteen kautta.

### 2.1 Kvasirakenne

Kvasirakenne eli rakennetyyppi *olla tekevinään* on verbirakenne, jolla voidaan ilmaista esimerkiksi teeskenneltyä, kuviteltua, uneksittua, leikittyä tai todenperäisyydeltään kyseenalaista tekemistä (KS s.v. *olla*; VISK § 453; Setälä 1926: 119). Se vastaa ikään kuin (= lat. *quasi*) -komparatiivisivulausetta, joka on predikaatiltaan konditionaalinen (Hakulinen 1979: 572; Ikola, Palomäki & Koitto 1989: 440). Rakenne ei ole kovin yleinen suomen yleiskielessä eikä puhekielessä (Salminen 2000: 43, 45). Esimerkiksi Salmisen (mts. 45) tutkimusaineistoon kuuluvissa Suomen Kuvalehden kahdessa vuosikerrassa oli yhteensä 61 kvasirakenne-esiintymää.

Kvasirakenne muodostetaan suomen yleiskielessä finiittiverbillä OLLA, jota voi kutsua apuverbiksi, sekä nominaalimuotoisella verbillä, joka on possessiivisuffiksillinen I partisiipin essiivi ja jota voi kutsua pääverbiksi (mts. 11, 13; Hakulinen 1979: 572; Setälä 1926: 119). Kvasirakennetta voidaan pitää verbiliittona, sillä sen pääverbillä on kiteytynyt nipputunnus, *-vinA-+POS*, ja apu- ja pääverbi muodostavat yhdessä kokonaisuuden, jonka merkitystä ei voi osista päätellä (VISK § 451). Kvasirakenteen lisäksi muita suomenkielisiä verbirakenteita, jotka muodostetaan finiittiverbin OLLA ja sen yhteydessä esiintyvän monimorfeemisen nominaalimuotoisen verbin avulla, ovat esimerkiksi niin sanottu

mahdollisuusrakenne *on tehtävissä* sekä niin sanottu propinkvatiivi *on tekemäisillään* (Salminen 2000: 13; ks. VISK § 1584, § 1521).

Kvasirakenteen possessiivisuffiksi ei aina ole persoonaviitteinen, vaan sitä saatetaan käyttää vaihtelemattomana, esimerkiksi *olin olevinaan* (VISK § 123). Vanhan kirjasuomen teksteissä kvasirakenteen possessiivisuffiksi (esim. *olevanans/tekevänäns*) on Forsman-Svenssonin (2000: 45) mukaan jopa useammin 3. persoonassa silloinkin, kun rakenteen subjektina on 1. tai 2. persoona. Kvasirakenteen partisiippi on kaikissa hänen tarkastelemissaan vanhan kirjasuomen tekstinäytteissä yksikössä (*olla tekevänäns*). Kvasirakenteen partisiipin monikkomuotoisuus, jota käytetään nykyisessä yleiskielessä, on lähtöisin itämurteista – länsimurteissa on käytetty sen sijaan yksikkömuotoista partisiippiä (Hakulinen 1979: 573). Kvasirakenteet on vanhassa kirjasuomessakin muodostettu usein verbistä OLLA – tällaisia on lähes puolet kvasirakenne-esiintymistä. Myös mentaaliverbistä TIETÄÄ muodostetut kvasirakenteet ovat vanhassa kirjasuomessa suhteellisen yleisiä. Mentaaliverbeistä muodostetut kvasirakenteet ovat nykyisessä yleiskielessä kuitenkin yleisempiä kuin vanhassa kirjasuomessa. (Mts. 45, 52–53.)

Kvasirakenteen vartaloverbiksi<sup>2</sup> käy mikä tahansa verbi, eli kyseessä on produktiivinen rakenne (Salminen 2000: 14). Kvasirakenteen subjekti voi olla elollinen tai eloton, ja kvasirakennetta käytetään paitsi myönteisessä myös kielteisessä muodossa. Kvasirakenteen konteksti, vartaloverbi, subjekti sekä myönteisyys (tai kielteisyys) vaikuttavat siihen, millaisia affektiivisia tai todellisuuteen suhtautuvia merkityksiä rakenteen merkitykseen liitetään. Kvasirakenteiden tarkkarajainen ryhmittely merkityksen perusteella on kuitenkin ongelmallista, sillä kvasirakenteisiin liitetyt merkityspiirteet riippuvat paljolti kuulijan tai lukijan tulkinnoista. (Mts. 12–13.)

Todellisuuden näkökulmasta tarkasteltuna kvasirakenne voi ilmaista joko nonfaktuaalista tilaa, jolloin puhuja ei ota kantaa merkityssisällön totuuteen, tai kontrafaktuaalista tilaa, jolloin puhuja pitää merkityssisältöä epätotena. Non- ja kontrafaktuaalisuuden välinen ero ei kuitenkaan ole ehdoton ja tarkkarajainen, eikä niitä ole syytä pitää toisilleen vastakkaisina. Tulkintaan kvasirakenteen nonfaktuaalisuudesta tai kontrafaktuaalisuudesta vaikuttavat muiden muassa rakenteen subjekti pääverbin semantiikka ja määritteet. Esimerkiksi mentaaliverbien 1. persoonasta muodostettujen kvasirakenteiden, kuten *olen ollut huomaavinani*, tulkitaan usein viittaavan puhujan tuntemukseen, kuvitelmaan tai luuloon. Mentaaliverbien 3. persoonasta muodostetut kvasirakenteet, kuten *ei ollut tietävinäänkään*, sen sijaan liitetään monesti rakenteen subjektin teeskentelyyn tai esittämiseen tiedon tai havainnon puutteesta. (Salminen 2000: 57–58.)

Moniäänisyyden näkökulma puolestaan kuvaa sitä, että kvasirakenteen avulla keskusteluun tuodaan mukaan toinen ääni. Kvasirakenteen käyttö viittaa vähintään kahteen tilanteeseen samanaikaisesti: puhehetken lisäksi ilmauksessa kuuluu kaiku

---

<sup>2</sup> Kvasirakenteen vartaloverbillä tarkoitan verbiä, josta kvasirakenteen pääverbi muodostetaan nipputunnuksen *-vinA-*+POS avulla.

jostakin toisesta tilanteesta, joka voi olla kielellinen tai ei-kielellinen. Nämä kvasirakenteen esiin tuomat näkökulmat ovat jollakin tavalla toisilleen vastakohtaisia. Niitä voivat olla esimerkiksi kuvitelma ja todellisuus tai puhujan ja jonkun toisen henkilön mielipide. (Salminen 2000: 17–18, 57.) Konkreettisimmillaan moniäänisyys ilmenee Salmisen (mts. 16–17) mukaan esimerkiksi puheen tai ajatuksen referoimisessa, kuten hänen Suomen Kuvalehdestä poimimassaan esimerkissä, jossa on kaksi eksplisiittisesti erotettavissa olevaa ääntä:

Pentti Peurasuo on *haistavinaan* uudenlaisia vaatimuksia puistojen ja muidenkin julkisten ulkotilojen suhteen. "Ilmeisesti Suomi on nyt ehtinyt urbanisoitua siinä määrin, että kaupungilta odotetaan viihtyisyyttä tällaisten ulkoilmaolohuoneiden puitteissa."

Tyypillisiä kvasirakenteen käyttökonteksteja ovat muiden muassa toisen henkilön puheen tai mielipiteen referointi, fiktiivisen tilanteen kuvaaminen, kerronnassa henkilön tajuntaan viittaava sisäinen fokalisaatio sekä affektiivisuuden ilmaiseminen. Esimerkiksi puheen referoinnissa kvasirakenteen luomaa tulkintaa vahvistetaan usein jollakin muulla referointiin viittaavalla aineksella, kuten puhetta ilmaisevalla verbillä, konditionaalilla tai *muka*-partikkelilla. Lisäksi persoonamuoto korreloi monesti kontekstin ja merkityksen kanssa. Esimerkiksi 1. persoonassa puhuja voi kvasirakenteen avulla muun muassa kuvata omaa aikaisempaa ajatustaan, joka on myöhemmin osoittautunut vääräksi, tai havaintoaan tai epämääräistä tai epävarmaa mentaalista tilaansa. (Salminen 2000: 131–132.) Myöhemmin vääräksi osoittautunutta ajatusta tai havaintoa kuvaavasta kvasirakenteesta Salminen (mts. 110) esittää Muotoopin arkistosta poimimansa murre-esimerkin:

kerram minä *olin kuulevinani* mutta sekin oli lehmä ja, juu.

Tutkijat ovat hahmottaneet kvasirakenteen osittain eri tavalla keskenään ja käyttäneet siksi kvasirakenteen rakenneosista keskenään erilaisia nimityksiä: Osa tutkijoista korostaa rakenteen kiteytyneisyyttä ja fraasimaisuutta kutsumalla rakenteen nominaaliosaa predikatiiviadverbiaaliksi ja OLLA-verbiä kopulaksi (ks. Hakulinen 1979: 572; Ikola ym. 1989: 440; Setälä 1926: 119). Tätä näkemystä tukee se kvasirakenteen johdosmainen ominaisuus, että tietynlaiset verbit, kuten mentaaliverbit, ovat rakenteen vartaloverbeinä poikkeuksellisen yleisiä. Toiset taas rinnastavat kvasirakenteen liittomuotoihin (ks. Vilkuna 2003: 257), jolloin partisiippimuotoa voidaan pitää taivutusmuotoisena pääverbinä ja OLLA-verbiä apuverbinä. Tätä näkemystä voidaan puolestaan perustella sillä, että kvasirakenne on produktiivinen eli muodostettavissa mistä tahansa verbistä, mikä on taivuttamiselle ominainen piirre. (Salminen 2000: 13–14.) Tässä tutkielmassani kutsun Salmisen (2000: 15) tavoin kvasirakenteen monimorfeemista nominaalimuotoa pääverbiksi ja OLLA-verbiä apuverbiksi korostaakseni kvasirakenteen säännönmukaista produktiivisuutta. Tutkimuksessani korostuukin nimenomaan eri verbeistä muodostettujen kvasirakenteiden välinen vertailu.

## 2.2 Kontekstuaalinen semantiikka ja fraseologia

### 2.2.1 Kontekstuaalinen semantiikka

Semantiikka eli merkitysoppi tarkastelee kielellisten ilmausten merkityksiä (Kangasniemi 1997: 22). Käsittelen tutkimuksessani semantiikkaa kontekstuaalisesta näkökulmasta. Kontekstuaalisessa semantiikassa sanojen, lauseiden ym. merkityksiä pyritään ymmärtämään analysoimalla niiden tekstiyhteyttä tai sosiaalista tilannetta (Jantunen 2004: 7–9). Se poikkeaa esimerkiksi konseptuaalisesta semantiikasta siten, että kontekstuaalisessa semantiikassa sanan merkitystä ei pidetä abstraktina käsitteenä, mielikuvana tai ideana (vrt. Kangasniemi 1997: 10), vaan merkitys riippuu käyttöyhteydestä ja tilannekontekstista. Konseptuaalisella merkitysanalyysilla voidaan siis kartoittaa kielenkäyttäjille yhteisiä käsitemaailmoja, kun taas kontekstuaalisessa näkökulmassa keskitytään todellisen kielenkäytön kuvaamiseen (Stubbs 1996: 22–23).

Kontekstuaalinen semantiikka soveltuu hyvin tämän tutkimuksen viitekehyyksiksi, koska tutkin kvasirakenteita korpusten avulla, ja korpukset edustavat todellista kielenkäyttöä todellisine tekstiyhteyksineen. Tarkastelen niitä siis nimenomaan käytön ja tekstiyhteyden näkökulmasta – enkä esimerkiksi introspektion avulla mentaalisina prosesseina, kuten konseptuaalisessa semantiikassa.

### 2.2.2 Fraseologia

Fraseologia on deskriptiiviseen kielentutkimukseen kuuluva ala, joka toimii ideaalisena yhtymäkohtana tekstikorpusten ja kielen kuvaamisen välissä. Ainakin länsimaisessa tutkimusperinteessä fraseologia on kuitenkin pitkään sivuutettu ja jätetty vähälle huomiolle, kunnes nykyaikaiset tekstikorpukset ovat nousseet pääasialliseksi tiedonlähteeksi kielen analyysissa. Tämä johtuu osittain siitä, että fraseologia tarkastelee kieltä kokonaisvaltaisesti eikä tee selvää eroa kieliopin ja sanaston välille, minkä vuoksi fraseologialle ei ole ollut tilaa kielentutkimuksen perinteisessä järjestelmässä. Lisäksi fraseologiassa keskitytään erityisesti syntagmaattisiin rakenteisiin ja kielenyksiköiden sekventiaalsiin suhteisiin, kun taas suurin osa kieliopista on paradigmaattisia eli keskittyy eri muotojen ja lekseemien vaihtoehtoisuussuhteisiin kielen asettamien muottien sisällä. Koska fraseologialla ei ole ollut vahvaa tutkimusperinnettä, siihen liittyviä käsitteitä on käytetty sekavasti ja terminologian määrittely on jakanut paljon mielipiteitä. Samalla fraseologian tutkimuskentän rajat ovat sekoittuneet muiden kielentutkimuksen suuntausten, kuten morfologian ja semantiikan kanssa. (Sinclair 2008: XV–XVI; Granger & Paquot 2008: 27.)

Koska fraseologialle ja siihen liittyville termeille on esitetty useita keskenään ristiriitaisiakin määritelmiä ja niitä on käytetty useissa eri merkityksissä, fraseologiaa

hyödyntävissä tutkimuksissa on hyvin tärkeää määritellä, mitä fraseologialla varsinaisesti tarkoitetaan. Tässä tutkimuksessa fraseologia ei rajoitu tarkastelemaan vain esimerkiksi idiomeja tai kiinteitä fraaseja (ks. Granger & Paquot 2008: 28). Sen sijaan käsittelen fraseologiaa laajemmassa mielessä: fraseologiassa tarkastellaan myös muita monisanaisia yksiköitä sekä yksittäisten sanojen kontekstuaalisia ja syntagmaattisia suhteita, kuten läheiseen tekstiympäristöön liittyviä valintarajoituksia ja -preferenssejä. (Ks. Granger & Meunier 2008a: XIX–XX; Jantunen 2009a: 358.) Fraseologia yhdistää toisiinsa sanoja, kielioppia, semantiikkaa ja sosiaalista käyttöä (Granger & Meunier 2008b: 5).

### 2.2.3 Fraseologinen yksikkö

Fraseologinen yksikkö tarkoittaa lekseemin ja sen ko(n)tekstuaalisten valintojen välisiä suhteita. Näitä ovat lekseemien keskinäiset suhteet eli kollokaatiot ja klusterit, lekseemien ja kieliopillisten kategorioiden väliset suhteet eli kolligaatiot sekä lekseemien ja semanttisten ominaisuuksien väliset suhteet, joita ovat semanttinen preferenssi ja semanttinen prosodia. (Jantunen 2012: 205; Jantunen 2009a: 358–359; Stubbs 2001: 87–88.)

Griesin (2008: 4–6) mukaan fraseologisia myötäesiintymisen ilmiöitä määriteltäessä on otettava huomioon ainakin seuraavat kuusi parametria (ks. myös Jantunen 2004: 16–23):

1. fraseologisen yksikön elementtien luonne
2. fraseologisen yksikön elementtien lukumäärä
3. ilmauksen esiintymien vähimmäislukumäärä voidakseen olla fraseologinen yksikkö
4. fraseologisen yksikön elementtien välinen sallittu etäisyys
5. fraseologisen yksikön elementtien leksikaalinen ja syntaktinen joustavuus
6. fraseologisen yksikön elementtien semanttinen yhtenäisyys ja semanttinen epäkompositionaalisuus.

Ensiksi tulee siis määritellä fraseologisen yksikön luonne. Tässä tutkimuksessani keskityn pääasiassa kahden sanan myötäesiintymiin eli kollokaatioihin. Lisäksi nostan esiin joitakin havaintoja sanojen ja tiettyjen semanttisten piirteiden myötäesiintymistä eli semanttisista preferensseistä sekä joitakin kolligaatioon liittyviä havaintoja esimerkiksi kollokaattien sanaluokista tai kieliopillisista muodoista.

Toiseksi tulee määrittää fraseologiseen yksikköön sisältyvien elementtien lukumäärä, joka on tässä tutkimuksessani kaksi. Toisin sanoen en keskity analyysissäni tarkastelemaan kolmesta tai useammasta sanasta muodostuvia sanaketjuja eli klustereita (ks. Jantunen 2004: 19), vaan analysoin sitä, millaisissa kahden sanan myötäesiintymissä HAVAITA- ja HUOMATA-kvasirakenteet esiintyvät.

Kolmanneksi on syytä asettaa vähimmäisraja-arvo ilmauksen esiintymille. Tässä työssä asetan fraseologisen yksikön elementtinä olevan lekseemin kollokaattiesiintymien vähimmäisraja-arvoksi viisi. Jätän siis tarkastelun ulkopuolelle ne lekseemit, jotka esiintyvät alle 5 kertaa HAVAITA- tai HUOMATA-kvasirakenteen kollokaattina. Tämän absoluuttisen frekvenssin raja-arvon lisäksi tilastollisten testien tulokset rajaavat sitä, mihin kollokaatioihin keskityn analyysissäni eniten. Tutkimukseni kannalta lähtökohtaisesti kiinnostavimpia ovat luonnollisesti merkitsevimmät kollokaatit.

Neljäntenä parametrina on fraseologisen yksikön elementtien välinen etäisyys. Määritän elementtien välin siten, että saneiden ei välttämättä tarvitse esiintyä aivan peräkkäin ollakseen toistensa kollokaatteja, vaan niiden välissä voi olla rajattu määrä muita sanoja. Tämä tarkastelualue on tässä tutkimuksessa 4:4, ja kerron siitä tarkemmin tilastollisten testien yhteydessä alaluvussa 3.3.

Viidenneksi tulee rajata, minkä verran leksikaalista ja syntaktista joustavuutta sallitaan fraseologisen yksikön elementeille. Tässä tutkimuksessa leksikaalista joustavuutta ei ole paljon, sillä fraseologisen yksikön elementeistä vähintään jommankumman on oltava HAVAITA- tai HUOMATA-kvasirakenteen pääverbi. Syntaktista joustavuutta puolestaan rajoitin siten, että kollokaatteja laskiessani otin huomioon vain ne saneet, jotka esiintyivät samassa virkkeessä kuin noodi eli tarkastelun kohteena oleva sana. Virkkeiden sisäisiin lauserajoihin sen sijaan en kiinnittänyt huomiota laskiessani kollokaatteja.

Kuudentena ja viimeisenä määriteltävänä parametrina on fraseologisen yksikön semanttisen yhtenäisyyden (engl. *semantic unity*) ja semanttisen epäkompositionaalisuuden (engl. *semantic non-compositionality* t. *semantic non-predictability*) merkitys tutkimuksen kannalta. Semanttisessa yhtenäisyydessä on kyse siitä, muodostavatko fraseologisen yksikön elementit yhdessä uuden semanttisen yksikön. Tätä on pidetty monissa fraseologisissa tutkimuksissa jopa kaikkein tärkeimpänä fraseologisen yksikön kriteerinä. Semanttinen epäkompositionaalisuus taas on sitä, että fraseologisen yksikön elementtien yhdistämisestä muodostuva merkitys on eri kuin elementtien merkitys. (Gries 2008: 6; Granger & Paquot 2008: 31; ks. myös Heinonen 2013: 45–48) Esimerkiksi idiomi *repiä pelihousunsa* on siis semanttisesti sekä yhtenäinen että epäkompositionaalinen, sillä lekseemit REPIÄ ja PELIHOUSUT muodostavat uuden semanttisen yksikön, jonka merkitys on 'suuttua', 'raivostua' tai 'menettää malttinsa', ja ilmauksen merkitys on eri kuin lekseemien merkitys (ks. KS s.v. *repiä*). Tässä tutkimuksessa semanttinen yhtenäisyys ja semanttinen epäkompositionaalisuus eivät ole fraseologisen yksikön määrittelyn kriteereinä, mutta idiomaattisia useamman sanan lausumia havaitessani nostan ne analyysissäni esiin sivuhuomioina.

## 2.2.4 Kollokaatio

Kollokaatio on eräs fraseologisen tutkimuksen keskeisistä käsitteistä, joita keskenään eri näkökulmiin keskittyvät tutkijat ovat määritelleet eri tavoin. Lyhyesti sanottuna kollokaatio viittaa sanojen myötäesiintymiseen (Stubbs 2001: 29). Kaikilla kielen lekseemeillä on omat yksilölliset käyttäytymismallinsa. Toistuvien ja tarkoituksellisten myötäesiintymisten perusteella kollokaatioissa voidaan havaita kaavamaisuutta ja tunnistaa toistuvia malleja. Näiden tietojen perusteella voidaan kuvata esimerkiksi lekseemien semanttisia ja syntaktisia ominaisuuksia. (Partington 1998: 16, 27; Karlsson 2004: 232–233.) Kohdesanan tekstiyhteydessä eli kotekstissa esiintyviä sanoja kutsutaan *kollokaateiksi* (Biber ym. 2006: 35). Puhtaasti kollokaatiota tarkasteltaessa sanojen välisiin syntaktisiin suhteisiin ei kiinnitetä lainkaan huomiota, toisin kuin esimerkiksi kolligaatioiden analysoimisessa (Stubbs 2001: 64; ks. Sinclair 1991: 170). Esittelen tässä alaluvun alaluvussa kolme näkökulmaa kollokaation määrittelyyn. Näitä ovat tekstuaalinen, assosiatiivinen ja tilastollinen näkökulma. (Ks. Partington 1998: 15–16.)

Tekstuaalisesta näkökulmasta määriteltynä kollokaatioissa on kyse sanojen taajasta myötäesiintymisestä eli esiintymisestä tekstissä toistensa läheisyydessä (Sinclair & Carter 2004: 28; Sinclair 1991: 170). Toisin sanoen kollokaatio on kahden tai useamman sanan välinen leksikaalinen suhde, jossa sanoilla on taipumus esiintyä toistensa kotekstissa. Kollokaatti ei siis tekstuaalisesta näkökulmasta tarkasteltuna esiinny välttämättä aivan noodin vieressä, vaan se voi olla muutaman sanan päässä siitä – käytännössä tutkimuksissa usein korkeintaan neljän sanan päässä (Stubbs 2001: 24; Sinclair 1991: 170). Sanavalinnat vaikuttavat toisiin sanavalintoihin asettamalla näille ehtoja, ja kielenaineokset ovat erottamattomalla tavalla sidoksissa ympäristöönsä (Sinclair & Carter 2004: 19).

Leech (1990: 17) puolestaan määrittelee ilmiötä psykologisesta, assosiatiivisesta näkökulmasta siten, että kollokaatiivinen merkitys koostuu assosiatiivisista merkityksistä, jotka sanaan liitetään sen ympäristössä esiintyvien sanojen vuoksi. Semanttisesti lähellä tosiaan olevilla sanoilla on taipumus esiintyä toistensa lähellä myös tekstiympäristössä. Elinikäisen kielelle altistumisen myötä kielenkäyttäjälle syntyy Partingtonin (1998: 16) mukaan odotuksia siitä, mitkä kielenaineokset esiintyvät teksteissä usein yhdessä toisten kanssa. Hoey (2005) käyttää tästä ilmiöstä nimitystä *leksikaalinen priming* (engl. *lexical priming*). Tuntemattoman sanan kohdatessaan kielenoppija etsii kotekstista vihjeitä päätelläkseen tuntemattoman sanan merkityksen niiden perusteella (Partington 1998: 16).

Vastaavasti Hoeyn (1991: 7) mukaan kollokaatio tarkoittaa lekseemin suhdetta kielenaineuksiin, jotka esiintyvät sen tekstiympäristössä sattumanvaraista suuremmalla todennäköisyydellä. Noodin voidaan siis katsoa ennustavan, että tietyt sanat esiintyvät ennen sitä tai sen jälkeen (Stubbs 2001: 30). Tämä tilastollisesta näkökulmasta esitetty määritelmä on sovellettavissa erityisesti korpuslingvistiikkaan,



jossa analysoidaan suuria tekstiaineistoja tietokoneiden avulla. Siten voidaan empiirisesti tutkia kollokaatioita tarkastelemalla konkordansseja sekä frekventimpien kollokaattien listoja, jolloin saadaan tietoa sanan uusuksesta eli siitä, miten ja missä yhteyksissä sitä tyypillisesti käytetään (ks. Karlsson 2004: 232–233). Korpustutkimuksen kautta empiirisesti hankittu tieto on tarpeellista, koska introspektion eli itsehavainnoinnin avulla saatava tieto sanojen merkityksistä on usein epäluotettavaa tai vähintäänkin vaillinaista (ks. Stubbs 2001: 89).

Tässä korpuslingvistiksessä tutkimuksessani korostuu eniten tilastollinen näkökulma kollokaatioon. Havaintoni, tulokseni ja johtopäätökseni kuitenkin luonnollisesti liittyvät myös kollokaation tekstuaaliseen ja assosiatiiviseen näkökulmaan.

## 2.3 Synonymia

### 2.3.1 Synonymian lajeja

Synonymia eli samanmerkityksisyys on eräs kielen keskeisistä paradigmaattisista merkityssuhteista, joita ovat lisäksi polysemia, homonymia, hyponymia, vastakohtaisuus, osa-kokonaisuussuhde sekä attribuutio. Synonyymeiksi kutsutaan lekseemejä, jotka poikkeavat toisistaan muodoltaan mutta joilla on identtinen denotaatio eli potentiaali viitata johonkin todelliseen tai kuviteltuun tarkoitteeseen. Kyse on siis sanojen tai ilmausten merkitysten samuudesta. Absoluuttisessa eli täydellisessä synonymiassa sanat ovat identtisiä kaikissa konteksteissa ja kaikissa merkityksissään – paitsi denotaatioltaan myös konnotaatioiltaan, tyyliltään ja affektiivisuudeltaan. Jos jokin näistä ei toteudu, kyse on siis epäabsoluuttisesta eli epätäydellisestä synonymiasta. Absoluuttista synonymiaa on luonnollisissa kielissä hyvin vähän – kenties ei lainkaan. Käytännössä kaikki synonyymit ovat siten epäabsoluuttisia, ja absoluuttisen synonyymien käsitteestä on hyötyä lähinnä silloin, kun sitä halutaan käyttää synonyymisuuden asteikon ääripäähän viittaamiseen tai täydellisen vastaavuuden puutteen korostamiseen. Vaikka kieleen syntyisi kahden sanan välille absoluuttinen synonymia, ajan saatossa sanojen merkitykset muuttuvat, jolloin myös synonyymien merkitykset erkaantuvat toisistaan vähintäänkin hieman, mikä riittää absoluuttisen synonymian hajoamiseen. (Kangasniemi 1997: 41–42; Karlsson 2004: 207, 219–220; Lyons 1995: 60–61; Lyons 1968: 447; Cruse 2000: 157–158; Jantunen 2004: 55.) Absoluuttinen synonymian idea ei kielentutkimuksen näkökulmasta olisi muutenkaan yhtä kiinnostavaa ja moninaista kuin lähisynonymia eli plesionymia, joka herättää tutkimuksellisesti mielenkiintoisia kysymyksiä esimerkiksi siitä, mitkä tekijät rikkovat tai eivät riko intuitiota sanojen merkitysten samuudesta (Cruse 2000: 156).

Lähisynonyymeiksi kutsutaan semanttisesti tai syntaktisesti jossain määrin samankaltaisia mutta epäidenttisiä lekseemejä (Lyons 1995: 60; Karlsson 2004: 220).

Cruse (2000: 159–160) esittää, että lähisynonyymien välisten semanttisten erojen, jotka eivät riko synonymiaa, on oltava joko vähämerkityksisiä tai taka-alalla – tai molempia. Tällaisia eroja ovat hänen mukaansa vierekkäisyys asteikolla (esim. BIG – HUGE [ISO – VALTAVA]), tietyt verbien adverbiaaliset erityisalajat (esim. CHUCKLE – GIGGLE [HIHITTÄÄ – KIKATTA]), näkökulmaerot (esim. STATE – DISPOSITION [TILA – MIELENLAATU]) sekä prototyyppikeskuksen erot (esim. BRAVE – COURAGEOUS [URHEA; prototyyppisesti fyysinen – ROHKEA; prototyyppisesti intellektuaalinen ja moraalinen]). Esimerkkinä merkityserosta, joka on merkittävä mutta taka-alalla, Cruse esittää englanninkieliset lähisynonyymiset sanat PRETTY, joka viittaa tyypillisemmin naispuolisiin, ja HANDSOME, joka viittaa tyypillisemmin miespuolisiin (suomen kielessä vastaavasti KAUNIS ja KOMEA: ks. Jantunen 2004: 9). Näissä sanoissa etualalla on siis ulkoisten ominaisuuksien miellyttävyyden, joka on molemmille sanoille yhteinen, kun taas erottava tekijä, sukupuoli, on taka-alalla. Sen sijaan sanojen WOMAN (NAINEN) ja MAN (MIES) välinen sukupuoli koskeva merkitysero sekä merkittävä että etualalla, ja siksi nämä sanat eivät ole lähisynonyymeja. (Cruse 2000: 160.) Lähisynonyymien PRETTY ja HANDSOME välisiä eroja kollokatiivisen merkityksen näkökulmasta esittelee Leech (1990: 17), joka luettelee kyseisille adjektiiveille tyypillisiä substantiivikollokaatteja: PRETTY-sanan todennäköisiä substantiivikollokaatteja ovat esimerkiksi GIRL (TYTTÖ), BOY (POIKA), WOMAN, FLOWER (KUKKA), GARDEN (PUUTARHA), COLOUR (VÄRI) ja VILLAGE (KYLÄ). HANDSOME puolestaan kollokoi vahvasti muun muassa substantiivien BOY, MAN, CAR (AUTO), VESSEL (ALUS), OVERCOAT (PÄÄLLYSTAKKI), AIRLINER (MATKUSTAJALENTOKONE) ja TYPEWRITER (KIRJOITUSKONE) kanssa.

Cruse (2000: 158) esittää jaottelun lähisynonymian ja propositionaalisen synonymian välille. Propositionaalinen synonymia voidaan määritellä lauseen paikkansapitävyyden kautta. Propositionaaliset synonyymit ovat vaihdettavissa toistensa paikalle ilman, että lauseen totuusarvo muuttuu. Toisin sanoen sanan vaihtaminen sen propositionaaliseen synonyymiin ei muuta lauseen paikkansapitävyyttä. Propositionaalisten synonyymien välisiin merkityseroihin liittyy aina eroavaisuutta vähintään yhdessä ei-propositionaalisisessa merkityksessä, esimerkiksi ero ekspressiivisessä merkityksessä, tyyllinen ero (asteikolla arkisuus–muodollisuus) tai ero diskurssikentässä. Yhtenä esimerkkinä propositionaalisesta synonymiasta Cruse (2000: 158) esittää lauseen *Silloin he olivat olleet yhdynnässä / rakastelleet / panneet ensimmäistä kertaa* (*This was the first time they had had intercourse / made love / fucked*), jossa minkä tahansa verbivaihtoehdon valinta säilyttää lauseen totuusehtoja koskevat ominaisuudet samoina. Erona kyseisten verbivaihtoehtojen välillä on erityisesti se, missä tilannekontekstissa kutakin niistä todennäköisimmin käytettäisiin, sekä se, millaista tyyliä ne edustavat. Lyons (1995: 44, 63–65) käyttää propositionaalisesta synonymiasta termiä deskriptiivinen synonymia ja toteaa, että monet semanttiset teoriat pitävät deskriptiivistä synonymiaa käytännössä ainoana synonymian muotona. Hän määrittelee ilmiön niin, että kaksi ilmausta on deskriptiivisesti synonyymisiä, jos ja vain jos propositionit (eli jotakin asiantilaa ilmaisevat lauseet), joihin sisältyy yksi ilmauksista, implikoivat

muutoin identtisiä propositioita, joihin sisältyy toinen ilmauksista, ja päinvastoin. Lyons (1995: 63) esittää deskriptiivisestä synonymiasta klassisena esimerkkinä sanat *poikamies (bachelor)* ja *naimaton mies (unmarried man)*. Deskriptiivistä synonyymisyyttä voi testata esimerkiksi tässä tapauksessa selvittämällä, voiko ketä tahansa poikamieheksi kuvailtua henkilöä kutsua naimattomaksi mieheksi – ja päinvastoin.

Yhtenä synonymiaa rikkovana tekijänä Lyons (1968: 451) pitää sitä, jos sanat eivät ole keskenään vaihtokelpoisia normaalissa kielenkäytössä. Esimerkiksi ilmaukset *cow* ('lehmä') ja *mature female bovine animal* ('täysikasvuinen naaraspuolinen nautaeläin') eivät tämän Lyonsin esittämän kriteerin mukaan ole synonyymisiä. Tämä johtuu siitä, että englannin kielen natiivi kielenpuhujat ei käyttäisi jälkimmäistä ilmausta tavanomaisessa kielenkäyttötilanteessa sanan *cow* vaihtoehtona, vaikka se on sekä kielipiillisesti että semanttisesti hyvin muotoiltu. Ilmaus *mature female bovine animal* voisi esiintyä lähinnä määritelmänä sanakirjassa.

Lyons (1968: 452) huomauttaa, että synonymia on hyvin kontekstisidonnaista eikä kyse ole itsessään rakenteellisesta suhteesta. Periaatteessa kaikki sanojen väliset merkityssuhteet ovat kontekstista riippuvaisia, mutta erityisesti synonymiassa kontekstin merkitys korostuu. Yleisen periaatteen mukaan sama tieto voidaan välittää kielessä joko syntagmaattisesti tai paradigmaattisesti. Synonymiassa tämä näkyy esimerkiksi lauseissa *Ajan New Yorkiin (I'm driving to New York)* ja *Lennän New Yorkiin (I'm flying to New York)* sekä *Menen New Yorkiin autolla (I'm going to New York by car)* ja *Menen New Yorkiin lentokoneella (I'm going to New York by air)*. Kahden ensimmäisen lauseen välillä merkityksiä erottavana tekijänä on paradigmaattinen valinta verbien *ajaa* ja *lentää* välillä ja kahden jälkimmäisen lauseen välillä sen sijaan *mennä*-verbin syntagmaattinen mukauttaminen.

### 2.3.2 Verbien HAVAITA ja HUOMATA synonymia

Tässä tutkimuksessa tarkasteltavien kvasirakenteiden keskenään varsin samanmerkityksiset vartaloverbit HAVAITA ja HUOMATA ovat toistensa lähisynonyynejä. Näiden kahden sanan semanttiset yhtäläisyydet sekä toisaalta synonyymisyyden vaillinaisuus ilmenevät esimerkiksi Kielitoimiston sanakirjan (KS s.v. *havaita*, s.v. *huomata*; kuviot 1 ja 2) esittämissä selitteissä sekä useissa esimerkkilauseissa, joissa kyseisten verbien sijoittaminen toistensa tilalle (ks. Kangasniemi 1997: 43) muuttaisi lauseen merkitystä vähintäänkin konnotaatioltaan. Esimerkiksi lause *En huomannut pyytää kuitausta* olisi jopa kielitajun vastainen, jos HUOMATA-verbi korvattaisiin HAVAITA-verbillä. HAVAITA-verbin selitteessä korostuu enemmän 'aistiminen', kun taas HUOMATA-verbin selitteessä 'merkille paneminen', 'keksiminen' ja 'käsittäminen'. Verbien väliset semanttiset erot eivät kuitenkaan välttämättä sellaisinaan heijastu kvasirakenteisiin, jotka on muodostettu kyseisten vartaloverbien avulla. Tämän tutkimuksen tulokset voivatkin tuoda hieman uutta ja kiinnostavaa tietoa siitä, missä määrin verbien väliset merkityserot ilmenevät myös niistä muodostettujen produktiivisten rakenteiden merkityksissä.

## havaita<sup>69</sup>

### TAIVUTUS ▾

1. huomata aistien avulla, aistia, erottaa; us. = **nähdä**.

*Hätkähti minut havaitessaan.*

#### NÄYTÄ VÄHEMMÄN ESIMERKKEJÄ ▾

*Havaitsi huoneessa käydyn.*

*Vain mikroskoopilla havaittava, havaittavissa.*

*Tuskin havaittava savunhaju.*

**Laajemmin** myös laitteen tms. vastaavanlaisesta toiminnasta.

*Liiketunnistin havaitsi oven avaamisen.*

2. huomata pohdinnan t. tutkimuksen perusteella, saada selville, oivaltaa, todeta, tajuta.

*Havaita taudin oireita.*

#### NÄYTÄ VÄHEMMÄN ESIMERKKEJÄ ▾

*Työssä ei havaittu pienintäkään edistymistä.*

*Hyväksi havaittu tapa.*

*Tutkimuksessa on havaittu, että – –.*

© 2020 Kotimaisten kielten keskus ja Kielikone Oy

KUVIO 1 Kielitoimiston sanakirjan (KS) selitteet ja esimerkkilauseet hakusanalle *havaita*.

## huomata<sup>73</sup>

### TAIVUTUS ▾

havaita, erottaa, saada selville, keksiä, panna merkille; oivaltaa, käsittää, ymmärtää, tajuta, älytä, todeta.

*Huomasin hänet jo kaukaa.*

#### NÄYTÄ VÄHEMMÄN ESIMERKKEJÄ ▾

*Eroa tuskin huoma.*

*Vika jäi huomaamatta.*

*Mitään epäilyttävää ei huomattu.*

*Huomattakoon, että – –.*

*Huomaa heti, ettet ole kunnossa.*

*Ei ollut huomaavinaanakaan.*

*Jo hänen äänestään huomaa, että – –.*

*Huomasi tulleen petetyksi.*

*Huomasi vihdoinkin lähteä.*

*En huomannut pyytää kiittausta.*

*Huomata jkn lahjakkuus.*

*Huomata jk hyväksi, puutteelliseksi.*

*Poika oli nyreissään, kun häntä ei huomattu tarpeeksi hänelle ei osoitettu huomiota.*

*Huomaa! (lyhenne: huom.) kirjallisessa esityksessä vars. huomautuksen alussa t. tähdentävänä ilmauksena.*

*Huomaa käsien asento!*

*Huom.! Selitys sivulla 4.*

*Huomaamatta huomiota herättämättä, salaa; epähuomiossa, vahingossa.*

*Livahti huomaamatta ulos.*

*Otin huomaamatta(ni) väärän avaimen.*

© 2020 Kotimaisten kielten keskus ja Kielikone Oy

KUVIO 2 Kielitoimiston sanakirjan (KS) selite ja esimerkkilauseet hakusanalle *huomata*.

### 2.3.3 Aiempia fraseologisia korpustutkimuksia suomenkielisten sanojen synonymiasta

Suomen kielen lähisynonyymien fraseologiaa korpuslingvistisin keinoin on tarkastellut erityisesti Jarmo Jantunen. Hän on tutkinut lähisynonyymisiä adjektiiveja TÄRKEÄ ja KESKEINEN (Jantunen 2001), käännessuomen synonymiaa astemääritteiden HYVIN, KOVIN ja OIKEIN kautta (Jantunen 2004) sekä verbien HIEVAHTAA ja LIIKAHTAA synonyymiaa oppimisen ja opettamisen näkökulmasta (Jantunen 2009a). Hän tunnisti, että lähisynonyymejä erottavia tekijöitä ovat muun muassa eri tekstilajien välinen vaihtelu, monet kotekstuaaliset valintasäännöt (Jantunen 2001: 188–189), kontekstuaalisen vaihtelun määrä (Jantunen 2004: 225) sekä erot taivutusmuotopreferensseissä, kollokaattien semantiikassa, kieltohakuisuuden asteessa ja sanasta muodostettavien johdosten määrässä (Jantunen 2009a: 371–373).

Lisäksi lähisynonymian fraseologiseen ja korpuslingvistiseen tutkimiseen on keskitytty muutamissa opinnäytetöissä. Tuuri (2010) ja Suoraniemi (2011) ovat tutkineet adpositioiden VUOKSI ja TAKIA synonymiaa, johon Tuuri (2014) perehtyi myöhemmässä tutkimuksessaan uudelleen analysoimalla edellisten lisäksi myös adpositioiden TÄHDEN, ANSIOSTA ja JOHDOSTA lähisynonymiaa. Näiden adpositioiden piirteet ovat analyysin mukaan varsin samanlaisia keskenään. Merkittävimmiksi synonymiaa erottaviksi tekijöiksi osoittautuivat sanan TÄHDEN matalampi korpusfrekvenssi sekä sanojen ANSIOSTA ja JOHDOSTA poikkeukselliset kontekstuaaliset rajoitukset ja semanttiset erityispiirteet. (Tuuri 2014.) Pirkola (2016) on tutkinut modaalisanon EHKÄ ja MAHDOLLISESTI synonymiaa ja havainnut, että sanoja yhdistäviä piirteitä ovat esimerkiksi pitkälti yhteiset semanttiset preferenssit, kuten merkitysryhmät 'aika', 'prosessi' ja 'määrä'. Vastaavasti synonymiaa rikkovia piirteitä ovat analyysin mukaan erot kollokaattien jakautumisessa eri sanaluokkiin sekä muutama semanttisen preferenssin merkitysryhmä. EHKÄ-sanassa korostuu enemmän subjektiivinen modaalisuus ja yksilön tieto, mikä ilmenee esimerkiksi suhteellisuuteen ja subjektiivisiin tulkintoihin viittaavien adjektiivivi- ja adverbikollokaattien kautta. MAHDOLLISESTI-sanassa sen sijaan korostuu objektiivinen modaalisuus ja yhteisön tieto, mikä nousee esiin esimerkiksi semanttisten preferenssien 'yhteiskunnallisuus, institutionaalisuus' ja 'konkreettiset oliot' kautta, jotka viittaavat koko yhteisöä koskeviin tai konkreettisesti havaittavissa oleviin asioihin. Laakkonen (2017) on puolestaan tarkastellut sanojen KEHO, VARTALO ja RUUMIS synonymiaa. Hänkin tunnisti lekseemien välistä synonymiaa rikkovana tekijänä kollokaattien jakautumisen eri sanaluokkiin, ja lisäksi eroja ilmeni monissa semanttisen preferenssin merkitysryhmissä. Kaikille kolmelle lekseemille yhteisiä merkitysryhmiä sen sijaan ovat 'yksilöiminen' ja 'suhtautuminen'. Kangas (2018) on analysoinut adverbien TÄYSIN ja KOKONAAN synonymiaa. Hänen mukaansa kyseisten adverbien kollokaateista – varsinkin verbikollokaateista – vain harva on molemmille yhteisiä, joten erot adverbien kollokaatioissa rikkovat synonymiaa. Yhteistä näille adverbeille on puolestaan lievästi negatiivinen semanttinen prosodia.

## 3 AINEISTO JA METODIT

### 3.1 Aineisto

Tutkimukseni aineistona käytin CSC:n ylläpitämän Kielipankin Suomi24-korpusta (Aller Media Oy 2020), joka on valmiiksi lemmatisoitu ja koostuu verkkokeskusteluista aikaväliltä 1.1.2001–31.12.2017. Suomi24 Keskustelu on kaikille avoin ja maksuton palvelu, jossa käyttäjät voivat käydä foorumilla keskustelua joko omalla nimellään, nimimerkillä tai täysin anonyymisti. Rekisteröityneiden käyttäjien kirjoittamia on noin seitsemän prosenttia aineiston kaikista kommentista. (Suomi24 2020; Lagus, Pantzar, Ruckenstein & Ylisiurua 2016: 9, 36–37.)

Aineiston etuna on tutkimukseni näkökulmasta laajuus: se sisältää jopa 3 470 785 562 sanetta<sup>3</sup>. HAVAITA-kvasirakenteita esiintyy aineistossa duplikaattivirkkeiden poistamisen (ks. alaluku 3.3) jälkeen 1 982 ja HUOMATA-kvasirakenteita 8 873. Paljon suppeammilla aineistoilla esiintymiä ei välttämättä olisi määrälliseen analyysiin riittävästi, sillä kvasirakenne on suhteellisen harvinainen (ks. Salminen 2000: 43, 45). Käyttämäni aineiston heikkoutena on, että se tarjoaa tietoa lähinnä verkkokeskusteluille tyypillisestä kielenkäytöstä, eivätkä tutkimuksen tulokset siksi ole välttämättä yleistettävissä kuvaamaan muille tekstilajeille tyypillistä kielenkäyttöä kovin täsmällisesti.

Suomi24-foorumien kirjoittajia voidaan pitää verkkoyhteisönä, johon sisältyy useita pienempiä yhteisöjä. Erityisen tiiviitä ovat esimerkiksi lesbopalstojen yhteisö sekä 70-vuotiaiden palstojen yhteisö. Suomi24 perustuu vähemmän kaveriverkostoihin ja oman identiteetin rakentamiseen kuin sosiaalisen median verkostopalvelut, kuten Facebook. Yhteisöllisyyden kokemuksen lisäksi käyttäjät kokevat saavansa toisiltaan vertaistukea ja löytävänsä kokemuseräistä tietoa. Osa kirjoittaa foorumille edistääkseen omaa yritystään tai poliittisia näkemyksiään, kun taas toiset kirjoittavat

---

<sup>3</sup> Tässä tutkimuksessa tarkoitan *saneella* merkkijonoa, joka sisältää vähintään yhden latinalaisen kirjaimen (ml. ääkköset) tai arabialaisen numeron.

puhtaasti ajanvietteenä. (Lagus ym. 2016: 10; Harju 2018: 51–52.) Citizen Mindscapes -hankkeessa toteutetun verkkokyselyn perusteella tyypillinen Suomi24-foorumin käyttäjä on keski-ikäinen kaupungissa avio- tai avopuolionsa kanssa asuva mies. Naisia kyselyyn vastanneista oli 36 %. Peräti 40 % vastaajista ilmoitti olevansa tyytymättömiä suomalaisen yhteiskunnan nykytilaan. (Harju 2017.)

Suomi24-foorumin keskustelut jäsentyvät hierarkkisesti, siten että päätason teema-alueet jakautuvat useisiin pienempiin alatasoihin, ja alimpana tasona ovat itse keskusteluketjut. Jokainen palstalle lähetetty kommentti joko avaa uuden keskusteluketjun tai kommentoi jo olemassa olevaa ketjua. Kunkin keskusteluketjun tekninen maksimi on 500 kommenttia, jonka täytyttyä ketju sulkeutuu. (Lagus ym. 2016: 6–7.)

Keskustelupalstan moderaattorit poistavat ketjujen avauksia ja kommentteja aktiivisesti – pääasiassa kävijöiden esittämien poistopyyntöjen perusteella. Pienten paikkakuntien palstat sekä alkoholiin, työttömyyteen tai potkuihin liittyvät keskustelut vaativat paljon moderointia. Niillä esiintyy enemmän toisten keskustelijoiden tahallista ärsyttämistä ja provosointia, josta monilla foorumin käyttäjällä on kielteisiä kokemuksia. (Lagus ym. 2016: 9; Harju 2017; Harju 2018: 51.) Esimerkiksi harrastuksiin tai luonnontieteisiin liittyvät palstat pysyvät paremmin siisteinä eivätkä vaadi moderointia. Poistetut tekstit eivät sisälly käyttämäni Suomi24-korpukseen. (Lagus ym. 2016: 9–10.)

Ketjujen, kommenttien ja sanojen määrissä mitattuna suosituin aihealue on nimeltään *yhteiskunta*, johon kuuluvat suositut teemat *politiikka* ja *uskonto*, ja toiseksi suosituin aihealue on *suhteet* (Lagus ym. 2016: 24–25). Tyypillisiä puheenaiheita ovat keskustelijoiden yhteiset kiinnostuksen kohteet, kuten harrastukset, terveysongelmat, sukupuolinen suuntautuminen sekä perheen perustaminen (mts. 5). Siksi on perusteltua pitää todennäköisenä, että kyseiset teemat nousevat keskimääräistä enemmän esiin Suomi24-keskusteluaineistoja analysoitaessa. Tämä onkin syytä ottaa huomioon myös tämän tutkimuksen tuloksia yleistettäessä.

## 3.2 Korpuslingvistiikka

Käytin tutkimuksessani korpuslingvistisiä menetelmiä. Korpuslingvistiikka keskittyy tarkastelemaan kielenkäyttöä ja siinä esiintyviä kaavoja empiirisesti luonnollisten tekstien avulla. Aineistona käytetään tiettyjen periaatteiden mukaan koottuja, sähköisesti koodattuja, laajoja luonnollisten tekstien kokoelmia eli korpuksia. Korpukset ovat usein annotoituja, eli niihin on systemaattisesti lisätty lingvististä tietoa esimerkiksi kunkin sanan sanaluokasta. Korpusten analysoinnissa hyödynnetään tietokoneella tehtäviä automaattisia ja vuorovaikutteisia tekniikoita, jotka nopeuttavat ja tehostavat analyysia. Analyysin tueksi määrällisten tulosten merkitsevyyttä voidaan arvioida tilastollisten testien avulla. Korpuslingvistiikassa määrälliseen kielentutkimukseen tulee aina yhdistää myös laadullista analyysia: määrällisistä havainnoista, kuten

kielellisten ilmausten esiintymisfrekvensseistä, tehdään laadullisia tulkintoja ja päätelmiä kielenkäytön malleista. (Biber ym. 2006: 4–5, 9; Baker 2006: 1–2.) Tässä tutkimuksessa käyttämiäni laadullisia menetelmiä käsittelen määrällisten analyysimenetelmien ohessa kahdessa seuraavassa alaluvussa.

Korpustutkimus voidaan jakaa korpuspohjaiseen eli korpusperustaiseen tutkimukseen ja korpusvetoiseen eli korpuslähtöiseen tutkimukseen (ks. Jantunen 2009b). Korpuspohjaisessa tutkimuksessa korpuksen avulla testataan, selitetään ja ilmentetään teoreettisia teorioita ja kuvauksia, jotka on muotoiltu jo ennen korpuksen tarkastelemista (Tognini-Bonelli 2001: 65). Korpusvetoisessa tutkimuksessa sitä vastoin pyritään systemaattisesti muodostamaan kieltä koskevia teoreettisia kategorioita korpusaineistossa esiintyvien toistuvien mallien perusteella (mts. 87). Aineistolla on siis korpusvetoisessa tutkimuksessa hyvin merkittävä painoarvo, ja korpus pyritään huomioimaan kielenkäytön kuvauksissa ja teoreettisissa esityksissä laajasti ja kokonaisvaltaisesti (mts. 84). Korpuspohjainen tutkimus sopii erityisen hyvin synonymian tutkimiseen ja kielenkäytön mallien välisten systemaattisten erojen paljastamiseen (Biber ym. 2006: 43), ja siksi perustan myös tämän tutkimukseni korpuspohjaiseen analyysiin. Lähisynonyymien välisiä eroavaisuuksia voidaan tutkia empiirisesti tarkastelemalla niiden käyttöä ja kontekstuaalisia preferenssejä korpuksissa. Korpuksia analysoimalla on myös mahdollista vertailla, miten esimerkiksi natiivipuhujan intuitio lähisynonyymisten sanojen merkityseroista poikkeaa autenttiseen kielenkäyttöön perustuvan tutkimuksen tuloksista. (Mts. 24–25.) Sinclairin (1991: 4) mukaan kielenkäyttäjän intuitiiviset käsitykset omasta kielenkäytöstään osoittautuvat harhaanjohtaviksi, kun kielenkäyttöä tarkastellaan empiirisesti. Intuitio ei siis ole täysin luotettava kieltä ja kielenkäyttöä koskevan tiedon lähde, ja sitä tulisi käyttää ennemminkin kielenkäyttöä koskevien tutkimusnäyttöjen arvioimiseen kuin niiden luomiseen (mts. 39; Tognini-Bonelli 2001: 86).

### 3.3 Kollokaatioanalyysi ja tilastolliset testit

Käytin kollokaatioanalyysini noodina kvasirakenteen pääverbiä eli sitä osaa, joka koostuu HAVAITA- tai HUOMATA-verbistä sekä nipputunnuksesta *-vinA-+POS*. Noodina oli siis aina vain yksi sana, enkä sisällyttänyt noodiin OLLA-apuverbiä. Valintani perustuu siihen, että apuverbi ei aina esiinny pääverbin vieressä, ja kollokaattien määrittäminen olisi kaksiosaista noodia käytettäessä monimutkaisempaa. Valinnan vaikutus tuloksiin tuskin on kovin merkittävä, koska pääverbin kollokaatteina esiintyvät sanat ovat paljolti samoja kuin sen lähellä esiintyvän apuverbin kollokaatit. *Kvasirakenteen kollokaateista* kirjoittaessani viittaan tässä tutkielmassa siis nimenomaan pääverbin kollokaatteihin.

Kollokaatioanalyysia varten on tarpeen määrittää tarkastelualue eli kollokaatioväli. Sen perusteella määrättyä, kuinka kaukana noodista esiintyvät sanat



lasketaan analyysissa kollokaateiksi. Churchin ja Hanksin (1990: 23) mukaan tarkastelualueen koko vaikuttaa analyysiin siten, että pienemmän tarkastelualueen avulla kiinteät ilmaukset, kuten idiomit, tunnistetaan helpommin, kun taas suurempi tarkastelualue korostaa esimerkiksi semanttisia suhteita. Määritin kollokaatioanalyysilleni tarkastelualueeksi 4:4. Se tarkoittaa, että laskin analyysiini mukaan sanat, jotka ovat enintään neljän sanan päässä noodista: neljä sanaa noodin vasemmalta puolelta ja neljä oikealta puolelta. Näin pääsin analyysissani riittävän hyvin käsiksi tuloksissa esiintyvien lauseiden semanttisiin suhteisiin. Kollokaatit voivat olla missä tahansa muodossa, sillä käsittelin kollokaatteja leksemeinä, en tiettyinä sananmuotoina. Myös noodina toimivaan kvasirakenteen pääverbiin voi liittyä mitä tahansa suffikseja.

Tutkimustulosten määrällisessä analysoinnissa hyödynsin tilastollisia menetelmiä: vertailin HAVAITA- ja HUOMATA-kvasirakenteiden kollokaatteja keskenään t-testin ja MI-testin avulla. Käytin MI-testiä mitatakseni sanojen välisten yhteyksien voimakkuutta. MI-testissä verrataan noodin ja kollokaatin yhteisesiintymän todennäköisyyttä sen todennäköisyyteen, että ne olisivat riippumattomia toisistaan. MI-testi saattaa siis nostaa esiin myös sellaisia kollokaatteja, joiden absoluuttinen esiintymisfrekvenssi on pieni. (Clear 1993: 279–280.) Jotta yksittäiset satunnaiset esiintymät eivät vääristäisi tuloksia, määritin absoluuttisen esiintymisfrekvenssin raja-arvoksi 5, eli en ottanut analyysissani huomioon kollokaatteja, jotka esiintyvät aineistossa alle 5 kertaa. Raja-arvon asettaminen vähentää esimerkiksi sen todennäköisyyttä, että yksittäisen tekstintuottajan idiosynkraattiset valinnat korostuisivat tuloksissa. Vastaavasti t-testi korostaa herkemmin kollokaatteja, joiden absoluuttinen frekvenssi on korkeampi kuin MI-testin korostamilla kollokaateilla. Kieliopillisia, taajaan esiintyviä sanoja, kuten persoonapronomineja ja prepositioita, nousee herkästi esiin t-testin tuloksissa, kun taas MI-testi antaa enemmän painoa idiosynkraattisille kollokaateille ja on hyvä keino kiinteiden sanayhdistelmien tunnistamiseen. (Mts. 281.) Koska testeillä on keskenään erilaiset vahvuudet ja heikkoudet, ne toimivat tutkimuksessani yhdessä hyödyllisinä työkaluina tilastollisen merkitsevyyden tarkastelemisessa ja yhden testin aiheuttaman vääristymän painoarvo tutkimuksen tuloksissa on pienempi.

Tilastollisia testejä tehdessäni käsittelin aineistoani kielellisesti analysoidussa VRT-formaatissa (ks. Lagus ym. 2016: 13–15) Mahti-supertietokoneen avulla. Laskin t- ja MI-testien tekemiseen tarvittavat tiedot sekä näiden testien tulokset Kumpulaisen (2020) Python-ohjelmointikielellä koodaamien tiedostojen avulla. Noodien laskemiseen käytin kahta säännöllistä lauseketta:

```
havaitsevina(ni | si | an | mme | nne | nsa)[a-z\ -]*
```

```
huomaavina(ni | si | an | mme | nne | nsa)[a-z\ -]*.
```

Näitä lausekkeita käyttäessäni poistin lisäksi kirjainkokojen vaikutuksen tuloksiin. Ensin Python-tiedostot etsivät hakulausekkeen avulla kaikki konkordanssit HAVAITA- ja HUOMATA-kvasirakenteille (kaavoissa tunnus *a*), laskivat kaikkien kollokaatteina

esiintyvien sanojen esiintymisfrekvenssit määrittämälläni 4:4-tarkastelualueella ( $k$ ), laskivat kaikkien kollokaatteina esiintyvien sanojen esiintymisfrekvenssit koko korpuksessa ( $h$ ) sekä laskivat korpuksen kaikkien saneiden yhteismäärän ( $N$ ). Sitten Python-tiedostot laskivat näiden keräämiensä lukujen perusteella t- ja MI-testien tulokset eli testien arvot kullekin aineistossa vähintään 5 kertaa esiintyvälle kollokaatille.

Laskiessani kullekin kollokaatille absoluuttisia esiintymisfrekvenssejä en luke-  
nut laskuihini mukaan konkordanssiduplikaatteja eli niiden virkkeiden identtisiä kaksoiskappaleita, joissa tarkastelun kohteena oleva kollokaatti esiintyy. HAVAITA-kvasirakenteen duplikaattivirkkeitä poistin 15 ja HUOMATA-kvasirakenteen duplikaattivirkkeitä 61. Suomi24-keskusteluaineistossa duplikaattivirkkeitä esiintyy erityisesti siksi, että palstan teknisten ominaisuuksien vuoksi sama viesti saattaa toistua aineis-  
tossa useamman kerran, jos joku on vastannut siihen. Kun jätin konkordanssidupli-  
kaatit pois laskuista, ne eivät päässeet vääristämään tilastollisten testien tuloksia. On  
teoriassa mahdollista, että jokin konkordanssivirke esiintyisi sattumalta täysin ident-  
tisenä useamman kerran. Tällaiset tapaukset ovat kuitenkin varsin epätodennäköisiä  
ottaen huomioon kvasirakenteen harvinaisuuden.

Tein t-testin ja MI-testin jokaiselle HAVAITA- tai HUOMATA-kvasirakenteen kollo-  
kaattina esiintyvälle sanalle. Tarkastelin testeissä seuraavia tapahtumia kunkin kor-  
puksessa esiintyvän lekseemin osalta ja molempien tarkasteltavien kvasirakenteiden  
osalta:

T1: satunnaisesti valittu sane on se sana, joka on tarkastelun kohteena.

T2: satunnaisesti valittu sane esiintyy tarkasteltavan kvasirakenteen kontekstissa (tarkaste-  
luvälillä 4:4).

Tapahtuman T1 todennäköisyys on siis

$$P(T1) = \frac{k}{N},$$

jossa

$k$  = tarkastelun kohteena olevan sanan esiintymät yhteensä koko korpuksessa

$N$  = kaikkien saneiden yhteismäärä koko korpuksessa.

Ja tapahtuman T2 todennäköisyys on

$$P(T2) = \frac{h}{N},$$

jossa

$h$  = kaikkia saneita yhteensä tarkastelun kohteena olevan kvasirakenteen kontekstissa (tar-  
kasteluvälillä 4:4)

$N$  = kaikkien saneiden yhteismäärä koko korpuksessa.

### 3.3.1 t-testi

Nollahypoteesi t-testissä on, että tapahtumat T1 ja T2 ovat toisistaan riippumattomia. Tämän oletuksen valossa todennäköisyys sille, että tapahtumat T1 ja T2 pätevät samaan aikaan – eli että satunnaisesti valittu sane on tarkastelun kohteena oleva sana ja se esiintyy tarkasteltavan kvasirakenteen kollokaattina – on

$$P_0(T1 \text{ ja } T2) = P(T1)P(T2) = \frac{k}{N} \cdot \frac{h}{N} = \frac{kh}{N^2}.$$

Tein t-testin siten, että laskin ensin kaikkien HAVAITA- tai HUOMATA-kvasirakenteen kotekstissa tarkasteluvälillä 4:4 esiintyvien sanojen esiintymät kyseisten kvasirakenteiden kollokaatteina sekä niiden lukumäärät koko aineistossa. Jos t-testin nollahypoteesi pitäisi paikkansa, tarkastelun kohteena olevan sanan tulisi esiintyä noodin eli tarkasteltavan kvasirakenteen kotekstissa keskimäärin yhtä usein kuin jos sanat olisivat satunnaisessa järjestyksessä, jolloin korrelaatiota tarkasteltavan sanan ja tarkasteltavan kvasirakenteen välillä ei olisi. Nollahypoteesin antamaa todennäköisyyttä  $P_0(T1 \text{ ja } T2)$  verrataan mitattuun todennäköisyyteen  $P(T1 \text{ ja } T2)$ , joka on siis tarkasteltavan sanan suhteellinen esiintymisfrekvenssi tarkasteltavan kvasirakenteen kollokaattina tarkasteluvälillä 4:4.

$$P(T1 \text{ ja } T2) = \frac{a}{N},$$

jossa

$a$  = tarkastelun kohteena olevan sanan absoluuttinen frekvenssi tarkasteltavan kvasirakenteen kotekstissa (tarkasteluvälillä 4:4)

$N$  = kaikkien saneiden yhteismäärä koko korpuksessa.

Käytin t-testin laskemisessa kaavaa

$$\frac{x - \mu}{\sqrt{\frac{x(1-x)}{N}}} \text{ (ks. Manning \& Schütze 1999: 164–165),}$$

jossa

$x = P(T1 \text{ ja } T2)$  eli suhteellinen esiintymisfrekvenssi (tarkasteltavan kvasirakenteen kollokaattina tarkasteluvälillä 4:4)

$\mu = P_0(T1 \text{ ja } T2)$  eli nollahypoteesin mukainen suhteellisen frekvenssin odotusarvo

$N$  = kaikkien saneiden yhteismäärä koko korpuksessa.

### 3.3.2 MI-testi

MI-testin laskemiseen puolestaan käytin kaavaa

$$\log_2 \left( \frac{P(T1 \text{ ja } T2)}{P(T1)P(T2)} \right) \text{ (ks. Manning \& Schütze 1999: 178–179).}$$

MI-testissä siis verrataan sanan mitattua suhteellista esiintymisfrekvenssiä siihen, mikä odotettu suhteellinen esiintymisfrekvenssi olisi sanojen ollessa riippumattomia toisistaan.

### 3.3.3 Esimerkki t- ja MI-testien laskemisesta

Esimerkiksi kollokaatin PIENI tulokset HAVAITA-kvasirakenteen kontekstissa tarkasteluvälillä 4:4 laskettiin näin:

$$N = \text{koko korpuksen saneiden yhteismäärä} = 3\,470\,785\,562$$

$$a = \text{absoluuttinen frekvenssi sanalle pieni HAVAITA-kvasirakenteen kontekstissa tarkasteluvälillä 4:4} = 108$$

$$k = \text{pieni-sanan esiintymät yhteensä koko korpuksessa} = 3\,531\,693$$

$$h = \text{kaikkia saneita yhteensä HAVAITA-kvasirakenteen kontekstissa tarkasteluvälillä 4:4} = 12\,574$$

Tällöin tapahtuman  $P(T1)$  todennäköisyys on siis

$$P(T1) = \frac{k}{N} = \frac{3\,531\,693}{3\,470\,785\,562} = 0,001017 \dots$$

ja tapahtuman  $P(T2)$  todennäköisyys on

$$P(T2) = \frac{h}{N} = \frac{12\,574}{3\,470\,785\,562} = 0,000003622 \dots$$

Näiden avulla laskin t-testin:

$$x = P(T1 \text{ ja } T2) = \frac{a}{N} = \frac{108}{3\,470\,785\,562} = 0,000000031 \dots$$

$$\mu = P_0(T1 \text{ ja } T2) = P(T1)P(T2) = \frac{k}{N} \cdot \frac{h}{N} = \frac{kh}{N^2} = \frac{3\,531\,693 \cdot 12\,574}{3\,470\,785\,562^2} = 0,000000003686 \dots$$

$$\frac{x - \mu}{\sqrt{\frac{x(1-x)}{N}}} = \frac{0,000000031 \dots - 0,000000003686 \dots}{\sqrt{\frac{0,000000031 \dots \cdot (1 - 0,000000031 \dots)}{3\,470\,785\,562}}} = 9,16113 \dots \approx 9,16$$

ja MI-testin:

$$\log_2 \left( \frac{P(T1 \text{ ja } T2)}{P(T1)P(T2)} \right) = \log_2 \left( \frac{0,000000031 \dots}{0,001017 \dots \cdot 0,000003622 \dots} \right) = 3,07741 \dots \approx 3,08.$$

Vastaavasti laskin t- ja MI-testien arvot kaikille muille sanoille, jotka esiintyvät HAVAITA- tai HUOMATA-kvasirakenteen kontekstissa tarkasteluvälillä 4:4.

### 3.3.4 Tilastollisten testien tulosten kokoaminen ja kollokaattien laadullinen analyysi

Kollokaattien määrällisen analyysin tuloksia kootessani poistin kollokaattien joukosta käsin sellaisia sananmuotoja, jotka on virheellisesti lemmatisoitu puhekielisyyden, kirjoitusvirheen tai sanan lyhentämisen vuoksi omiksi lekseemeikseen, kuten *Ite*, *ees*, *,,ei*, *,olen*, *USAn*, *viimeaikoina*, *äläkä*, *ns.* ja *etta*. Kun vertailin HAVAITA- ja HUOMATA-

kvasirakenteiden kollokaatteja keskenään, en keskittynyt tarkastelemaan t- ja MI-testien arvojen absoluuttisia tilastollisia merkitsevyyksiä. Tämä valinta perustuu siihen, että tieto merkitsevyyden tasosta itsessään ei ole kovin hyödyllistä, ja erityisesti t-testin avulla absoluuttisia tilastollisia merkitsevyyksiä on vaikea arvioida. Kiinnostavampaa ja hyödyllisempää on tarkastella ja vertailla keskenään kollokaattien saamia arvoja tilastollisissa testeissä sekä erityisesti kollokaattien keskinäisiä järjestyksiä eri taulukoissa. (Barnbrook 1996: 98; Manning & Schütze 1999: 166.) Asetin kuitenkin t-testin raja-arvoksi 2,00 ja MI-testin raja-arvoksi 1,58 rajatakseni tutkimuksen ulkopuolelle sanat, joita ei testitulosten perusteella selvästikään voi pitää tarkastelemilleni kvasirakenteille tyypillisinä (ks. Barnbrook 1996: 99; ks. Jantunen 2001). Tulkitsin suoraan kollokaattien t- ja MI-testien tarjoamista merkitsevyyksisarvoista, mitkä kollokaatit ovat tyypillisiä sekä HAVAITA- että HUOMATA-kvasirakenteille ja mitkä vain jommallekummalle. Tilastollisia merkitsevyyksiä käsitellessäni kirjoitan siis kollokaattien *tyypillisyydestä* korostaakseni t- ja MI-testien tarjoamien merkitsevyyksisarvojen suhteellisuutta. Koska HUOMATA-kvasirakenteita esiintyy aineistossa moninkertaisesti enemmän kuin HAVAITA-kvasirakenteita ja koska absoluuttiset frekvenssit voivat vaikuttaa t-testin merkitsevyyksisarvoihin, eri kvasirakenteiden t-arvot eivät välttämättä ole vertailukelpoisia keskenään. Siksi keskityn t-testitulosten analyysissä enemmän kollokaattien suhteellisiin järjestyksiin kuin varsinaisiin t-arvoihin.

Kollokaattien t- ja MI-arvojen sekä absoluuttisten frekvenssien tarkastelemisen lisäksi analysoin kollokaatteja laadullisesti tarkastelemalla niitä konkordanssikorpuksen avulla tekstiyhteyksissään voidakseni tulkita niiden merkityksiä asianmukaisesti. Laadullisessa analyysissä nostin esiin kiinnostavia havaintoja esimerkiksi sanaluokkajaon ja lauseenjäsennyksen näkökulmasta sekä keskittymällä substantiivikollokaattien jaollisuuteen (ks. VISK § 554–556) ja niiden konkreettisuuteen tai abstraktisuuteen (ks. VISK § 551), adjektiivikollokaattien absoluuttisuuteen tai suhteellisuuteen (ks. VISK § 605–607) sekä verbikollokaattien semanttisiin piirteisiin (ks. VISK § 445) ja sija- ja persoonamuotoihin. Kollokaattien jaottelu eri sanaluokkiin perustuu Ison suomen kieliopin (VISK § 438) esitykseen, jonka keskeisimpänä kriteerinä on taiputus. Siinä sanat jaetaan kolmeen pääryhmään, joita ovat sijan ja luvun mukaan taipuvat nominit (eli substantiivit, adjektiivit, pronominit ja numeraalit), persoonan, tempuksen ja moduksen mukaan taipuvat verbit sekä taipumattomat tai vaillinaisesti taipuvat adpositiot, adverbit ja partikkelit.

## 4 TULOKSET

### 4.1 HAVAITA- ja HUOMATA-kvasirakenteiden yhteiset merkitsevät kollokaatit

#### 4.1.1 MI-testi

MI-testin tulosten mukaan (taulukko 1) HAVAITA- ja HUOMATA-kvasirakenteiden synonymiaa yhdistäviä, molemmille rakenteille hyvin tyypillisiä, 40 merkitsevimmän kollokaatin joukkoon kuuluvia substantiivikollokaatteja ovat SARKASMI, IRONIA, PIIRRE, ILMIÖ, KIINNOSTUS ja SÄVY. Kiinnostavaa on, että näistä kaikki ovat merkitykseltään abstrakteja. Ivallisuuteen viittaavat substantiivit SARKASMI ja IRONIA, jotka ovat merkitykseltään hyvin lähellä toisiaan, ovat kaikkein merkitsevimpiä molemmille kvasirakenteille yhteisiä kollokaatteja. Niillä viitataan useimmiten jonkin edeltävän viestin ivalliseen sävyyn, kuten esimerkeissä 1 ja 2. Niiden MI-arvo onkin molempien rakenteiden kollokaattina yli 5 eli poikkeuksellisen korkea – HAVAITA-kvasirakenteen kollokaattina peräti yli 7. Ivallisuuteen viittaavia kollokaatteja voi siis perustellusti pitää eräänä HAVAITA- ja HUOMATA-kvasirakenteiden synonymiaa yhdistävänä tekijänä. Myös kollokaatit PIIRRE ja SÄVY ovat suhteellisen paljon tyypillisempiä HAVAITA- kuin HUOMATA-kvasirakenteelle. ILMIÖ ja KIINNOSTUS sen sijaan ovat likimain yhtä tyypillisiä kummallekin rakenteelle. PIIRRE esiintyy aineistossa tyypillisesti monikon partitiivissa, HAVAITA- tai HUOMATA-kvasirakenteen objektina, kuten esimerkeissä 3 ja 4. Substantiivikollokaateista jaollisia eli ei-laskettavia ovat SARKASMI, IRONIA ja KIINNOSTUS sekä joissakin tapauksissa SÄVY (jaollisuudesta ja jaottomuudesta ks. VISK § 554–556).

1) Olin vähän *havaitsevinani* **sarkasmia** tuossa viestissä.

2) Olen *huomaavinani* lievää **ironiaa** kirjoituksestasi.

3) Sellaista kunnioittavat ja pelkäävät kaikenlaiset miehet statuksesta riippumatta, ja samoja **piirteitä** olen ollut *huomaavinani* myös Venäjällä käydessä.

4) Olen *havaitsevinani* sinussa autoritaarisia **piirteitä**.

TAULUKKO 1 HAVAITA- ja HUOMATA-kvasirakenteiden 40 merkitsevintä kollokaattia MI-arvon mukaan (molemmille kvasirakenteille yhteiset kollokaatit lihavoitu).

kollokaatti (HAVAITA-kvasirakenne)	MI-arvo	kollokaatti HUOMATA-kvasirakenne)	MI-arvo
<b>hienoinen</b>	<b>9,06</b>	tendenssi	7,48
<b>sarkasmi</b>	<b>7,65</b>	<b>hienoinen</b>	<b>6,37</b>
<b>ironia</b>	<b>7,57</b>	häivähdys	6,33
jonkinasteinen	6,44	sivusilmä	6,21
yhtäläisyys	6,16	<b>ironia</b>	<b>5,65</b>
<b>lievä</b>	<b>6,10</b>	vahingonilo	5,49
tietynlainen	5,66	liikehdintä	5,24
<b>piirre</b>	<b>5,39</b>	kireys	5,18
ristiriita	5,37	vilkuilla	5,08
katkeruus	5,23	<b>sarkasmi</b>	<b>5,07</b>
<b>jonkinlainen</b>	<b>5,11</b>	pilke	4,96
<b>välistä</b>	<b>5,01</b>	moikata	4,86
epätoivo	4,96	tervehtiä	4,79
älyllinen	4,95	<b>välistä</b>	<b>4,31</b>
pyrkimys	4,53	flirtti	4,30
<b>sävy</b>	<b>4,46</b>	punastua	4,28
kateus	4,40	poispäin	4,28
rivi	4,06	vastakaiku	4,03
kirjoitus	3,89	katse	4,01
<b>ilmiö</b>	<b>3,88</b>	äläkä	3,99
<b>kiinnostus</b>	<b>3,66</b>	etäinen	3,97
<b>toisinaan</b>	<b>3,62</b>	<b>piirre</b>	<b>3,86</b>
havaita	3,45	ilme	3,68
ajattelu	3,18	ilkeys	3,67
tietty	3,17	kirkastua	3,64
asenne	3,16	halveksunta	3,61
pieni	3,08	<b>kiinnostus</b>	<b>3,61</b>
tällainen	3,05	<b>ilmiö</b>	<b>3,60</b>
kommentti	3,03	vastaantulija	3,57
lause	3,02	<b>lievä</b>	<b>3,56</b>
kehitys	2,92	teeskennellä	3,55
uskonnollinen	2,68	<b>toisinaan</b>	<b>3,52</b>
palsta	2,60	jonkunlainen	3,46
jotenkin	2,56	flirttailla	3,45
muutos	2,55	erinäinen	3,44
kirjoittaja	2,55	molemminpuolinen	3,44
merkki	2,52	vältellä	3,35
joskus	2,52	<b>jonkinlainen</b>	<b>3,33</b>
samanlainen	2,51	<b>sävy</b>	<b>3,30</b>
hieman	2,49	ele	3,29

HAVAITA- ja HUOMATA-kvasirakenteiden välistä synonymiaa yhdistäviä, molemmille rakenteille tyypillisiä adjektiivikollokaatteja puolestaan ovat MI-testin tulosten perusteella vähäisyyttä ilmaisevat, keskenään synonyymiset adjektiivit HIENOINEN ja LIEVÄ sekä kaltaisuutta ilmaiseva JONKINLAINEN. HIENOINEN ja LIEVÄ ovat suhteellisia adjektiiveja, jotka komparoituvat. Kollokaatti JONKINLAINEN on indefiniittisestä kvanttoripronominista *jokin* johdettu proadjektiivi, joka usein ilmaisee kvasirakenteelle ominaista epämääräisyyttä (ks. Salminen 2000: 132), kuten esimerkissä 5. JONKINLAINEN voi myös kollokaattien HIENOINEN ja LIEVÄ tavoin ilmaista määrää, jolloin sen merkitys on 'jonkin verran' (KS s.v. *jonkinlainen*). MI-testin perusteella adjektiivia HIENOINEN voidaan pitää kaikkein merkitsevimpänä kollokaattina tarkastelemilleni kvasirakenteille: HAVAITA-kvasirakenteen kollokaattina sen MI-arvo on kaikista korkein, 9,06, ja HUOMATA-kvasirakenteen kollokaattina toiseksi korkein, 6,37. Myös kollokaatit LIEVÄ ja JONKINLAINEN ovat MI-arvojen mukaan merkittävästi tyypillisempiä HAVAITA-kvasirakenteelle. Aineistossa kyseiset adjektiivikollokaatit esiintyvät tyypillisesti partitiivimuodossa, attribuuttina jollekin merkitykseltään negatiiviselle HAVAITA- tai HUOMATA-kvasirakenteen substantiiviobjektille (esimerkit 5, 6 ja 7).

- 5) Olen *havaitsevinani* kirjoituksessani jo **jonkinlaista** epätoivoa ja paniikkia.
- 6) Olin *huomaavinani* **hienoista** kiukkuakin kannanotossani.
- 7) Näin sivustaseuraavana olen *havaitsevinani* **lievää** vittuuntumista sponzorinketkulla?

MI-testin perusteella sekä HAVAITA- että HUOMATA-kvasirakenteen kanssa hyvin voimakkaasti kollokoivat edellä esitettyjen substantiivien ja adjektiivien lisäksi myös kaksi adverbia: aikaa ilmaiseva TOISINAAN sekä idiomaattisessa yhteydessä esiintyvä VÄLISTÄ. Kollokaatti VÄLISTÄ esiintyy aineistossa yleensä osana semanttisesti epäkompositionaalista ilmausta *rivien välistä*, kuten esimerkissä 8. Ilmauksella viitataan siis sanomatta tai kirjoittamatta jätettyjen asioiden päättelemiseen ja arvaamiseen (KS s.v. *rivi*). VÄLISTÄ kollokoi hieman voimakkaammin HAVAITA- kuin HUOMATA-kvasirakenteen kanssa.

- 8) Jos ei nyt aina ihan suoraan, niin kuitenkin rivien **välistä** olen *huomaavinani* vahin-  
goniloa minun helvettiin joutumisestani.

#### 4.1.2 t-testi

HAVAITA- ja HUOMATA-kvasirakenteiden synonymiaa yhdistäviä nominikollokaatteja, jotka ovat hyvin merkitseviä molemmille rakenteille, ovat t-testin tulosten mukaan (taulukko 2) substantiivit ILMIÖ ja VIESTI, adjektiivit PIENI ja SELLAINEN sekä pronominit JOKIN ja ITSE. Abstrakti ja jaoton substantiivi ILMIÖ, joka nousi esiin myös MI-testin tuloksissa, kohdassa 4.1.1, on t-testinkin perusteella suunnilleen yhtä merkitsevä kollokaatti kummallekin tarkastelemalleni kvasirakenteelle. Niin ikään abstrakti ja jaoton



substantiivi VIESTI puolestaan on yleisempi HAVAITA-kvasirakenteen kollokaattina: se on HAVAITA-kvasirakenteen kollokaattina 22. yleisin ja HUOMATA-kvasirakenteen kollokaattina 37. yleisin. HAVAITA- ja HUOMATA-kvasirakenteiden substantiivikollokaattien abstraktisuus, joka nousi esiin myös MI-testituloksissa (ks. kohta 4.1.1) voidaan lukea rakenteiden synonymiaa yhdistäväksi tekijäksi.

Vähäisyyttä ilmaisevan adjektiivikollokaatin PIENI merkitys viittaa useissa tapauksissa jonkin – yleensä negatiivisen – tunteen voimakkuuteen, kuten esimerkissä 9. Sanan PIENI merkitys vaikuttaa olevan aineistossa toisinaan ivallinen siinä mielessä, että sitä käytetään myös silloin, kun viitataan johonkin suurempaan kuin 'pieni'. Demonstratiivipronominista SE johdettu proadjektiivi SELLAINEN esiintyy yleensä yksikön partitiivissa – joko kvasirakenteen pääverbin objektina, kuten esimerkissä 10, tai objektin attribuuttina. Sillä viitataan anaforisesti asiaan, henkilön tai tilanteen edustamaan luokkaan tai lajiin (VISK § 629). Indefiniittinen kvanttoripronomini JOKIN on HAVAITA-kvasirakenteen kollokaattina kaikkein merkitsevin nominikollokaatti, ja HUOMATA-kvasirakenteen kollokaattinakin se on nomineista toiseksi merkitsevin. Se saa t-testissä molempien kvasirakenteiden kollokaattina arvoksi yli 10. Se on kvasirakenteesiintymien lukumääriin suhteutettuna selvästi tyypillisempi kollokaatti HAVAITA- kuin HUOMATA-kvasirakenteelle. JOKIN esiintyy tyypillisesti niin ikään yksikön partitiivissa eli muodossa *jotain*. Tällöin kvasirakenne on yleensä yksikön 1. persoonassa ja puhujan aistihavaintoihin liittyy kvasirakenteelle ominaiseen tapaan epävarmuutta tai epämääräisyyttä (ks. Salminen 2000: 132). Tämä käy ilmi paitsi esimerkissä 10 myös esimerkissä 11, johon sisältyy myös tyypillinen kollokaattiesiintymä refleksiivipronominista ITSE. Kollokaatti ITSE viittaa aineistossa useimmiten puhujaan tai kirjoittajaan itseensä, kun hän haluaa korostaa omia aistihavaintojaan.

9) Olen *havaitsevinani* kirjoituksestasi **pien**tä narkästyistä siitä, että monet ovat kristinuskon nimissä toisiaan lahdanneet.

10) Olin jo silloin *havaitsevinani*, **jotain sellaista** Orinmattilan uroksissa.

11) Mutta **itse** olen ollut *havaitsevinani* **jotain** pakkomielleisyyteen viittaavaa, tai miten sen nyt ilmaisisin...

TAULUKKO 2 HAVAITA- ja HUOMATA-kvasirakenteiden 40 merkitsevintä kollokaattia MI-arvon mukaan (molemmille kvasirakenteille yhteiset kollokaatit lihavoitu).

kollokaatti (HAVAITA-kvasirakenne)	t-arvo	kollokaatti (HUOMATA-kvasirakenne)	t-arvo
<b>olla</b>	<b>42,13</b>	<b>olla</b>	<b>85,20</b>
<b>että</b>	<b>19,50</b>	<b>että</b>	<b>43,28</b>
<b>jokin</b>	<b>10,10</b>	ei	35,06
<b>pieni</b>	<b>9,16</b>	hän	12,17
kirjoitus	7,04	<b>jokin</b>	<b>12,05</b>
jonkinlainen	6,37	ettäei	11,62
<b>joskus</b>	<b>5,89</b>	<b>joskus</b>	<b>10,49</b>
palsta	5,79	<b>mutta</b>	<b>9,88</b>

tällainen	5,63	minä	9,63
piirre	5,43	<b>pieni</b>	<b>9,49</b>
lievä	5,40	<b>itse</b>	<b>9,24</b>
myös	5,36	nainen	8,83
<b>sellainen</b>	<b>5,13</b>	huomata	8,69
<b>ainakin</b>	<b>5,04</b>	kun	8,63
<b>itse</b>	<b>4,93</b>	<b>sellainen</b>	<b>8,45</b>
sama	4,93	<b>ainakin</b>	<b>8,31</b>
<b>kyllä</b>	<b>4,87</b>	<b>kyllä</b>	<b>7,78</b>
tietty	4,79	välillä	7,75
ristiriita	4,68	palsta	7,72
<b>mutta</b>	<b>4,59</b>	edes	7,47
kuitenkin	4,26	koira	7,43
<b>viesti</b>	4,24	katse	6,89
hieman	4,19	jonkinlainen	6,74
kommentti	4,02	tai	6,73
sarkasmi	3,98	kirjoitus	6,59
katkeruus	3,89	<b>jotenkin</b>	<b>6,46</b>
tämä	3,87	piirre	6,45
<b>jotenkin</b>	<b>3,81</b>	<b>ilmiö</b>	<b>6,36</b>
merkki	3,60	täällä	6,22
hienoinen	3,60	mä	5,98
moni	3,60	mies	5,96
havaita	3,52	sama	5,76
rivi	3,39	kiinnostus	5,73
viime	3,37	kukaan	5,61
<b>ilmiö</b>	<b>3,36</b>	vaikka	5,42
kateus	3,30	muutos	5,34
ironia	3,30	<b>viesti</b>	5,14
samanlainen	3,19	tervehtiä	5,10
jopa	3,11	muka	5,07
tietynlainen	3,10	keskustelu	5,00

Nominikollokaattien lisäksi molempien tarkastelemieni kvasirakenteiden pääverbeille yhteisiä, hyvin merkitseviä kollokaatteja ovat verbi OLLA, adverbit JOSKUS ja JOTENKIN sekä partikkelit ETTÄ, AINAKIN, KYLLÄ ja MUTTA. Esiintymisfrekvenssien ja t-testin arvojen mukaan mitattuna verbi OLLA on ylivoimaisesti yleisin ja merkitsevin kollokaatti sekä HAVAITA- että HUOMATA-kvasirakenteen pääverbin kotekstissa. Tämä ei ole lainkaan yllättävää, sillä OLLA-apuverbi on välttämätön osa kvasirakennetta (ks. VISK § 451). Muiden kuin apuverbinä toimivien OLLA-verbien kollokointia tarkastelemieni kvasirakenteiden kanssa ei siis voida menetelmällisistä syistä suoraan havaita näistä tuloksista. Kollokaatti JOSKUS on ajankohdan adverbii, jonka merkitykseen sisältyy usein epämääräisyyttä (VISK § 649). Myös tapaa kvantifioiva proadverbikollokaatti JOTENKIN implikoi usein epämääräisyyttä. HAVAITA- ja HUOMATA-kvasirakenteen kollokaattina sitä käytetään melko usein virkkeen ensimmäisenä sanana

viittaamassa kvasirakenteen ilmaiseman aistihavainnon tapaan tai sen epämääräisyyteen, kuten esimerkeissä 12 ja 13.

12) **Jotenkin** olen ollut *havaitsevinani*, että kolmion käyttö olisi nykyisin unohtunut monelta.

13) **Jotenkin** olin *huomaavinani*, että kivien paikat olisivat selvillä hyvinkin kaukaa historiasta ja että se on ajansaatossa muutenkin kuin kansallisylpeyden takia muuttanut muotoaan.

Partikkelikollokaatit ETTÄ ja MUTTA ovat tarkastelemieni kvasirakenteiden kotteksteissa usein esiintyviä konjunktioita: Yleiskonjunktio ETTÄ on selvästi yleisin noodin oikeanpuoleisen kotelstin kollokaatti sekä HAVAITA- että HUOMATA-kvasirakenteen yhteydessä, ja se aloittaa lähes aina sivulauseen, joka toimii objektina eli on havaitsemisen tai huomaamisen kohde, kuten esimerkissä 14. Tämän lisäksi ETTÄ on absoluuttisen frekvenssin perusteella kummankin kvasirakenteen toiseksi yleisin kollokaatti. Kollokaatti MUTTA on puolestaan rinnastuskonjunktio, joka on yleisempi noodin vasemmanpuoleisessa kotelstissa. Se esiintyy tavallisesti adversatiivisuutta tai konsessiivisuutta ilmaisevissa virkkeissä eli sellaisissa, joissa ilmaistaan vastakohtaisuutta tai rajoitusta, kuten esimerkissä 15, tai myönnytystä, kuten esimerkissä 16 (ks. VISK § 1103; KS s.v. *mutta*).

14) Olen ollut *huomaavinani*, **että** olet valveutunut ja paljon oivaltanut Raamatun perustuuksia.

15) Olen toistaiseksi antanut asian olla ja en ole henkilöä *huomaavinani*, **mutta** kieltämättä se vähän rajoittaa omaa elämää.

16) Kärsimättömyyttä olin vain *havaitsevinani* kesksutelussa, **mutta** sitä meissä on kyllä jokaisessa.

Fokuspartikkeli AINAKIN koskee asteikon alkupäätä ja esiintyy tarkastelemieni kvasirakenteiden kollokaattina useimmiten merkityksessä 'vähintään', kuten esimerkissä 17 (ks. VISK § 840). Kollokaatti KYLLÄ on modaalipartikkeli, jota käytetään paljon varsinkin puhutussa kielessä. Sen avulla pyritään yleensä tuomaan lausumaan vakuuttelua tai häivyttämään negatiivista implikaatiota, kuten esimerkissä 18 (ks. VISK § 1609).

17) Olen ollut *havaitsevinani* tällaista asennetta **ainakin** poliisin, terveydenhuollon että koulutusalan viranhaltijoissa.

18) Sinkkuna ollessani olin **kyllä** *huomaavinani*, että jotkut kauniitkin naiset menivät jotenkin lukkoon seurassani.

## 4.2 HAVAITA-kvasirakenteen merkitsevät kollokaatit

### 4.2.1 MI-testi

HAVAITA-kvasirakenteelle hyvin tyypillisiä, 40 merkitsevimmän kollokaatin joukkoon kuuluvia substantiivikollokaatteja, jotka eivät ole HUOMATA-kvasirakenteelle tyypillisiä (eivät ole 80 merkitsevimmän kollokaatin joukossa), ovat YHTÄLÄISYYS, EPÄTOIVO, PYRKIMYS, AJATTELU, KIRJOITUS, ASENNE, KOMMENTTI, LAUSE, KEHITYS, PALSTA, KIRJOITTAJA ja MERKKI (taulukko 3). Nämä substantiivikollokaatit siis erottavat HAVAITA- ja HUOMATA-kvasirakenteiden välistä synonymiaa. Näistä hyvin moni on merkitykseltään abstrakteja substantiiveja, ja selvästi konkreettiseksi voidaan lukea vain KIRJOITTAJA. Substantiivikollokaateista selvästi ainakin KIRJOITUS, KOMMENTTI, LAUSE, PALSTA ja KIRJOITTAJA usein viittaavat aineistossa Suomi24-keskustelufoorumilla käytävään keskusteluun, mikä lienee pääteltävissä jo kyseisten kollokaattien semanttisista merkityksistä. Tämä ilmenee esimerkeissä 19, 20, ja 21, joissa siis käydään metakeskustelua eli keskustellaan itse keskustelemisesta. Lisäksi mielenkiintoinen havainto on se, että kyseisten substantiivikollokaattien joukossa on kaksi johdosta samasta sanasta, KIRJOITTAA-verbistä.

19) Mutta joo, tää on tämmöstä, jonkinlaista epätarkkuutta olin *havaitsevinani* psykopaatteihin tuossa **lauseessasi**

20) **Kommenttisi** taustalta olen *havaitsevinani* sinulla tai läheisellä ystävälläsi syvän masenuksen poikasta.

21) Tämä ei liene ainut avauksesi tällä **palstalla**, jotain tuttua olen *havaitsevinani*?

TAULUKKO 3 HAVAITA-kvasirakenteiden 40 merkitsevintä kollokaattia MI-arvon mukaan (HAVAITA-kvasirakenteelle [HUOMATA-kvasirakenteeseen verrattuna] paljon tyypillisemmät kollokaatit lihavoitu; vain HAVAITA-kvasirakenteelle tyypilliset kollokaatit merkitty tähdellä \*).

kollokaatti (HAVAITA-kvasirakenteeseen)	kollokaattiesiintymiä, absoluuttinen frekvenssi $f$	esiintymiä koko aineistossa, absoluuttinen frekvenssi $f$	MI-arvo	t-arvo
hienoinen	13	6711	9,06	3,60
sarkasmi	16	22059	7,65	3,98
ironia	11	15988	7,57	3,30
<b>jonkinasteinen*</b>	<b>6</b>	<b>19136</b>	<b>6,44</b>	<b>2,42</b>
<b>yhtäläisyys*</b>	<b>5</b>	<b>19315</b>	<b>6,16</b>	<b>2,20</b>
lievä	30	120420	6,10	5,40
<b>tietynlainen</b>	<b>10</b>	<b>54454</b>	<b>5,66</b>	<b>3,10</b>
piirre	31	204337	5,39	5,43
<b>ristiriita</b>	<b>23</b>	<b>153000</b>	<b>5,37</b>	<b>4,68</b>
<b>katkeruus</b>	<b>16</b>	<b>117307</b>	<b>5,23</b>	<b>3,89</b>
jonkinlainen	43	343254	5,11	6,37
välistä	8	68425	5,01	2,74

<b>epätoivo*</b>	<b>5</b>	<b>44462</b>	<b>4,96</b>	<b>2,16</b>
<b>älyllinen</b>	<b>6</b>	<b>53412</b>	<b>4,95</b>	<b>2,37</b>
<b>pyrkimys*</b>	<b>5</b>	<b>59851</b>	<b>4,53</b>	<b>2,14</b>
sävy	7	87538	4,46	2,53
<b>kateus</b>	<b>12</b>	<b>156687</b>	<b>4,40</b>	<b>3,30</b>
<b>rivi</b>	<b>13</b>	<b>215868</b>	<b>4,06</b>	<b>3,39</b>
<b>kirjoitus*</b>	<b>57</b>	<b>1064647</b>	<b>3,89</b>	<b>7,04</b>
ilmiö	13	244026	3,88	3,36
kiinnostus	9	196475	3,66	2,76
toisinaan	6	134550	3,62	2,25
havaita	15	377623	3,45	3,52
<b>ajattelu*</b>	<b>7</b>	<b>212945</b>	<b>3,18</b>	<b>2,35</b>
<b>tietty*</b>	<b>29</b>	<b>892025</b>	<b>3,17</b>	<b>4,79</b>
<b>asenne*</b>	<b>12</b>	<b>370956</b>	<b>3,16</b>	<b>3,08</b>
<b>pieni*</b>	<b>108</b>	<b>3531693</b>	<b>3,08</b>	<b>9,16</b>
<b>tällainen*</b>	<b>41</b>	<b>1370279</b>	<b>3,05</b>	<b>5,63</b>
<b>kommentti*</b>	<b>21</b>	<b>710549</b>	<b>3,03</b>	<b>4,02</b>
<b>lause*</b>	<b>8</b>	<b>271448</b>	<b>3,02</b>	<b>2,48</b>
<b>kehitys*</b>	<b>9</b>	<b>327326</b>	<b>2,92</b>	<b>2,60</b>
<b>uskonnollinen*</b>	<b>6</b>	<b>258680</b>	<b>2,68</b>	<b>2,07</b>
<b>palsta*</b>	<b>48</b>	<b>2182503</b>	<b>2,60</b>	<b>5,79</b>
<b>jotenkin*</b>	<b>21</b>	<b>983516</b>	<b>2,56</b>	<b>3,81</b>
<b>muutos</b>	<b>12</b>	<b>563947</b>	<b>2,55</b>	<b>2,87</b>
<b>kirjoittaja*</b>	<b>11</b>	<b>518900</b>	<b>2,55</b>	<b>2,75</b>
<b>merkki*</b>	<b>19</b>	<b>913190</b>	<b>2,52</b>	<b>3,60</b>
joskus	51	2457800	2,52	5,89
samanlainen	15	726243	2,51	3,19
hieman	26	1279822	4,19	2,49

MERKKI esiintyy aineistossa yleensä monikon partitiivissa, HAVAITA- tai HUOMATA-kvasirakenteen objektina, abstraktissa merkityksessä 'osoitus tai todiste jostakin' (KS s.v. *merkki*), kuten esimerkissä 22. Suurimpia MI-arvoja vain HAVAITA-kvasirakenteelle tyypillisistä substantiivikollokaateista saavat YHTÄLÄISYYS (6,16), EPÄTOIVO (4,96) ja PYRKIMYS (4,53). Näiden kolmen sanan absoluuttinen frekvenssi HAVAITA-kvasirakenteen kollokaattina on kuitenkin varsin alhainen – ne juuri ja juuri yltävät asettamaan raja-arvoon 5. Kollokaattien KIRJOITUS ja PALSTA absoluuttinen frekvenssi puolestaan on varsin korkea, ja etenkin KIRJOITUS saa myös varsin korkean MI-arvon, 3,89. KIRJOITUS esiintyy HAVAITA-kvasirakenteen kollokaattina usein inessiivissä, kuten esimerkissä 23.

22) Itse olin *havaitsevinani* tästä naisesta ihastumisen **merkkejä** alussa, mutta kuitenkin en vaan jotenkin voinut uskoa sitä enkä noteerannut sitä oikein mitenkään.

23) Olin kuitenkin *havaitsevinani* alkuperäisessä **kirjoituksessasi** yrityksen osoittaa, että avioerolapsestakin voi tulla juristi.

Myös substantiivit RISTIRIITA, KATKERUUS, KATEUS ja RIVI kollokoivat huomattavasti voimakkaammin HAVAITA-kvasirakenteen kanssa (kuuluvat 20 merkitsevimmän kollokaatin joukkoon), mutta ne ovat suhteellisen tyypillisiä myös HUOMATA-kvasirakenteelle (kuuluvat 80 merkitsevimmän mutta eivät 40 merkitsevimmän kollokaatin joukkoon). Niiden kaikkien MI-arvo on HAVAITA-kvasirakenteen kollokaattina hyvin korkea, yli 4, ja RISTIRIITA ja KATKERUUS yltyvätkin kymmenen merkitsevimmän HAVAITA-kvasirakenteen kollokaatin joukkoon. HUOMATA-kvasirakenteen kollokaatteina nämä neljä substantiivia saavat MI-arvoja 2,5:n ja 3:n väliltä. RIVI on HAVAITA-kvasirakenteen tilastollisesti merkitsevä kollokaatti lähinnä siksi, että idiomaattinen ilmaus *rivien välistä* on rakenteen kotekstissa varsin yleinen. Näin ollen näistäkin neljästä substantiivikollokaatista kaikki ovat merkitykseltään abstrakteja.

HAVAITA-kvasirakenteen merkitsevien kollokaattien joukossa on paljon sekä jaollisia että jaottomia substantiiveja. Jaollisia ovat EPÄTOIVO, AJATTELU, ASENNE, KEHITYS, KATKERUUS, KATEUS ja RISTIRIITA ja jaottomia puolestaan KIRJOITUS, KOMMENTTI, LAUSE, PALSTA, KIRJOITAJA, MERKKI, RIVI ja MUUTOS. Rajatapauksiksi voidaan lukea kollokaatit YHTÄLÄISYYS ja PYRKIMYS.

Adjektiiveista HAVAITA-kvasirakenteen kanssa voimakkaasti kollokoivia mutta HUOMATA-kvasirakenteen kanssa korkeintaan heikosti kollokoivia ovat MI-testin mukaan JONKINASTEINEN, TIETTY, PIENI, TÄLLAINEN ja USKONNOLLINEN. Näistä lähes kaikki ilmaisevat tarkastelemieni kvasirakenteiden kollokaatteina absoluuttista ominaisuutta – ainoastaan PIENI on suhteellinen adjektiivi. Indefiniittisestä kvanttoripronominista JOKIN johdettu proadjektiivi JONKINASTEINEN (vrt. JONKINLAINEN kohdassa 4.1.1) kollokoi HAVAITA-kvasirakenteen kanssa selvästi voimakkaammin, ja sen MI-arvo on 6,44, mutta absoluuttinen frekvenssi vain 6. Yksilöivä, ei-kuvaileva, spesifistä indefiniittisyyttä ilmaiseva TIETTY, kokoon viittaava PIENI sekä demonstratiivinen proadjektiivi TÄLLAINEN sen sijaan esiintyvät HAVAITA-kvasirakenteen kanssa varsin usein – PIENI peräti 108 kertaa. Niiden MI-arvo on kuitenkin selvästi pienempi, mutta silti merkittävän korkea, yli 3. JONKINASTEINEN, TIETTY ja PIENI esiintyvät aineistossa tyypillisesti yksikön partitiivissa, substantiiviobjektin attribuutteina, kuten esimerkeissä 24 ja 25. HAVAITA-kvasirakenteen kollokaatteina hyvin merkitseviä ovat lisäksi adjektiivit TIETYNLAINEN (MI-arvolla 5,66) ja ÄLYLLINEN (MI-arvolla 4,95), jotka ovat suhteellisen merkitseviä myös HUOMATA-kvasirakenteen kollokaatteina.

24) **Jonkinasteista** pinnallisuutta olin *havaitsevinani* tällaisessa päättelyssä.

25) **Tiettyä** väheksymistä olin *havaitsevinani*, mutta eiköhän sekin muutu.

HAVAITA-kvasirakenteelle poikkeuksellisen tyypillinen verbi on mentaaliverbi HAVAITA, jonka MI-arvo HAVAITA-kvasirakenteen kollokaattina on verbeistä selvästi korkein, 3,45. Kyse on siis HAVAITA-verbin itsekollokaatiosta eli kollokaatiosta itsensä kanssa (ks. esim. Daley, Jones, Sinclair & Krishnamurthy 2004: 83). HAVAITA-verbi esiintyy kollokaattina tyypillisesti samalla, kun ilmaistaan 'havaitsemisen'

ja 'havaitsevinaan olemisen' välistä disjunkttiivista suhdetta eli vaihtoehtosuhdetta rinnastuskonjunktion TAI avulla, kuten esimerkissä 26. HUOMATA-kvasirakenteen kanssa HAVAITA-verbi ei sen sijaan kollokoi kovin voimakkaasti.

26) Hyvä aihe tämä, itse olen **havainnut** tai ollut *havaitsevinani* että kyseisessä vaunussa 9413 tahtoo tyhjäkäynti pudota kovin pienille kierroksille vaikka kuinka olen huomannut että kuljettajat yrittävät säätää.

#### 4.2.2 t-testi

Merkitseviä kollokaatteja, jotka ovat t-testin tulosten mukaan (taulukko 4) tyypillisiä HAVAITA-kvasirakenteelle (kuuluvat 40 merkitsevimmän kollokaatin joukkoon) mutta eivät HUOMATA-kvasirakenteelle (eivät kuulu 80 merkitsevimmän kollokaatin joukkoon), ovat abstrakti ja jaollinen substantiivi KATKERUUS, suhteellinen adjektiivi HIE-NOINEN ja esimerkissä 27 esiintyvä mentaaliverbi HAVAITA – jotka nousivat korkean MI-arvonsa vuoksi esiin jo kohdassa 4.2.1 – sekä demonstratiivipronomini TÄMÄ ja fokuspartikkeli JOPA. Näitä lekseemejä voidaan t-arvojen perusteella pitää HAVAITA- ja HUOMATA-kvasirakenteiden välistä synonymiaa rikkovina kollokaatteina. Kollokaatti TÄMÄ on 4:nneksi yleisin kaikista HAVAITA-kvasirakenteen kollokaateista, sillä sen absoluuttinen frekvenssi on peräti 147. Kollokaatti JOPA esiintyy tyypillisesti ilmauksissa, joissa puhuja haluaa asettaa jonkin ominaisuuden tai asiointilan asteikon loppu- eli yläpään (ks. VISK § 849). Monesti se ominaisuus, jota kollokaatti JOPA korostaa, on lauseen objektina, kuten esimerkissä 27.

27) Ilmassa olen ollut *havaitsevinani jopa* vihamielisyyttä.

TAULUKKO 4 HAVAITA-kvasirakenteiden 40 merkitsevintä kollokaattia t-arvon mukaan (HAVAITA-kvasirakenteelle [HUOMATA-kvasirakenteeseen verrattuna] paljon tyypillisemmät kollokaatit lihavoitu; vain HAVAITA-kvasirakenteelle tyypilliset kollokaatit merkitty tähdellä \*).

kollokaatti (HAVAITA-kvasirakente)	kollokaatti- esiintymiä, ab- soluuttinen frekvenssi <i>f</i>	esiintymiä koko aineistossa, abso- luuttinen frek- venssi <i>f</i>	t-arvo	MI-arvo
olla	3131	2,14E+08	42,13	2,02
että	636	39824984	19,50	2,14
jokin	167	10054466	10,10	2,20
pieni	108	3531693	9,16	3,08
<b>kirjoitus</b>	<b>57</b>	<b>1064647</b>	<b>7,04</b>	<b>3,89</b>
<b>jonkinlainen</b>	<b>43</b>	<b>343254</b>	<b>6,37</b>	<b>5,11</b>
joskus	51	2457800	5,89	2,52
<b>palsta</b>	<b>48</b>	<b>2182503</b>	<b>5,79</b>	<b>2,60</b>
<b>tällainen</b>	<b>41</b>	<b>1370279</b>	<b>5,63</b>	<b>3,05</b>
<b>piirre</b>	<b>31</b>	<b>204337</b>	<b>5,43</b>	<b>5,39</b>
<b>lievä</b>	<b>30</b>	<b>120420</b>	<b>5,40</b>	<b>6,10</b>
<b>myös</b>	<b>85</b>	<b>9832708</b>	<b>5,36</b>	<b>1,25</b>

sellainen	53	4319831	5,13	1,76
ainakin	46	3261212	5,04	1,96
itse	100	13982938	4,93	0,98
<b>sama</b>	<b>67</b>	<b>7355007</b>	<b>4,93</b>	<b>1,33</b>
kyllä	78	9660595	4,87	1,16
<b>tietty</b>	<b>29</b>	<b>892025</b>	<b>4,79</b>	<b>3,17</b>
<b>ristiriita</b>	<b>23</b>	<b>153000</b>	<b>4,68</b>	<b>5,37</b>
mutta	135	22557328	4,59	0,72
<b>kuitenkin</b>	<b>40</b>	<b>3610904</b>	<b>4,26</b>	<b>1,61</b>
viesti	<b>28</b>	<b>1535850</b>	<b>4,24</b>	<b>2,33</b>
<b>hieman</b>	<b>26</b>	<b>1279822</b>	<b>4,19</b>	<b>2,49</b>
<b>kommentti</b>	<b>21</b>	<b>710549</b>	<b>4,02</b>	<b>3,03</b>
<b>sarkasmi</b>	<b>16</b>	<b>22059</b>	<b>3,98</b>	<b>7,65</b>
<b>katkeruus*</b>	<b>16</b>	<b>117307</b>	<b>3,89</b>	<b>5,23</b>
<b>tämä*</b>	<b>147</b>	<b>27634109</b>	<b>3,87</b>	<b>0,55</b>
jotenkin	21	983516	3,81	2,56
<b>merkki</b>	<b>19</b>	<b>913190</b>	<b>3,60</b>	<b>2,52</b>
<b>hienoinen*</b>	<b>13</b>	<b>6711</b>	<b>3,60</b>	<b>9,06</b>
<b>moni</b>	<b>40</b>	<b>4763827</b>	<b>3,60</b>	<b>1,21</b>
<b>havaita*</b>	<b>15</b>	<b>377623</b>	<b>3,52</b>	<b>3,45</b>
<b>rivi</b>	<b>13</b>	<b>215868</b>	<b>3,39</b>	<b>4,06</b>
<b>viime</b>	<b>18</b>	<b>1016523</b>	<b>3,37</b>	<b>2,29</b>
ilmiö	13	244026	3,36	3,88
<b>kateus</b>	<b>12</b>	<b>156687</b>	<b>3,30</b>	<b>4,40</b>
<b>ironia</b>	<b>11</b>	<b>15988</b>	<b>3,30</b>	<b>7,57</b>
<b>samanlainen</b>	<b>15</b>	<b>726243</b>	<b>3,19</b>	<b>2,51</b>
<b>jopa*</b>	<b>22</b>	<b>2050378</b>	<b>3,11</b>	<b>1,57</b>
<b>tietynlainen</b>	<b>10</b>	<b>54454</b>	<b>3,10</b>	<b>5,66</b>

HAVAITA-kvasirakenteen hyvin merkitsevistä kollokaateista, jotka ovat jokseenkin merkitseviä myös HUOMATA-kvasirakenteen kotekstissa, valtaosa on joko substantiiveja tai adjektiiveja. Näitä ovat MI-testin jo esiin nostamat substantiivit KIRJOITUS, PALSTA, RISTIRIITA, KOMMENTTI, MERKKI, RIVI ja KATEUS ja adjektiivit JONKINLAINEN ja LIEVÄ, joita analysoin kohdassa 4.1.1 molemmille kvasirakenteille yhteisinä kollokaatteina, sekä substantiivit PIIRRE, SARKASMI ja IRONIA ja adjektiivit TÄLLAINEN, TIETTY ja TIETYNLAINEN, joita analysoin kohdassa 4.2.1 HAVAITA-kvasirakenteelle erityisen tyyppillisinä kollokaatteina. Näistä kaikkein merkitsevimpiä HAVAITA-kvasirakenteelle ovat t-testin mukaan substantiivi KIRJOITUS, joka on 5:nneksi merkitsevin kollokaatti, sekä adjektiivi JONKINLAINEN, joka on 6:nneksi merkitsevin kollokaatti. Edellä mainitut substantiivikollokaatit ovat pitkälti abstrakteja, ja niistä kaikki paitsi KATEUS ovat jaottomia.

Jo MI-testin tuloksissa esiin nousseiden nominikollokaattien lisäksi t-testin tuloksissa uusina HAVAITA-kvasirakenteen nominikollokaatteina esiin nousevat adjektiivit VIIME ja SAMANLAINEN sekä pronominit SAMA ja MONI. Näistä korkein t-arvo on pronominilla SAMA, joka on t-testin tulosten mukaan HAVAITA-kvasirakenteen 16:nneksi merkitsevin kollokaatti. Taipumaton adjektiivikollokaatti VIIME esiintyy



lähies aina jonkin aikaan viittaavan sanan edellä, jolloin sen merkitys on 'edellinen' tai 'viimeksi kulunut' (ks. KS s.v. *viime*), kuten esimerkissä 28. Kaltaisuutta ilmaiseva vertailuadjektiivi SAMANLAINEN on johdettu indefiniittipronominista SAMA. Sekä SAMA että SAMANLAINEN viittaavat HAVAITA-kvasirakenteen kollokaatteina yleensä identtisyyteen, vastaavuuteen tai yhtäläisyyteen, kuten esimerkeissä 29 ja 30. Kollokaatti MONI esiintyy HAVAITA-kvasirakenteen kollokaattina yleensä pronomina, merkityksessä 'usea', kuten esimerkissä 31, mutta joskus myös substantiivina, kuten esimerkissä 32 (ks. KS s.v. *moni*).

28) Olen ollut **viime** vuosina *havaitsevinani* yhteiskunnan jakautumista yhä näkyvämmiin kahteen luokkaan.

29) Tosin olen ollut *havaitsevinani*, että tämä **sama** ongelma tiedonvälityksen puolueellisuudesta vaivaa koko Eurooppaa.

30) Siellä oli **samanlaista** idealismia kuin olin *havaitsevinani* sinun kirjoituksessasi.

31) Olen ollut *havaitsevinani* **monella** naisella samankaltaisia tunteja...

32) Lisäksi palstaa lukiessani olen ollut *havaitsevinani*, että **moni** on alkanut jopa uskomaan tuota herjauspostajaa, niin kuin niillä olisi jotain totuus pohjaa!

Muita t-testin tuloksissa esiin nousevia HAVAITA-kvasirakenteen kollokaatteja ovat adverbi HIEMAN sekä partikkelit MYÖS ja KUITENKIN. HIEMAN on kvanttoriadverbi, joka ilmaisee ominaisuuden vähyyttä (ks. VISK § 657). MYÖS on inklusiivinen fokuspartikkeli, joka on t-testin mukaan HAVAITA-kvasirakenteelle erittäin tyypillinen – rakenteen 12:ksi merkitsevin kollokaatti. Partikkeli KUITENKIN on puolestaan konsessiivisen suhteen konnektiivi. Se esiintyy HAVAITA-kvasirakenteen kollokaattina noodin vasemmalla puolella, kuten esimerkissä 33. Partikkelit MYÖS ja KUITENKIN esiintyvät HAVAITA-kvasirakenteen kollokaatteina lähes aina myönteisissä muodoissa, ja niiden kieltomuotoja *myöskään* ja *kuitenkaan* esiintyy kutakin vain 2 kappaletta.

33) Olen **kuitenkin** *havaitsevinani* ajattelussasi sitä samaa 'urautumista' mistä syytät meitä kaupunkilaisia.

## 4.3 HUOMATA-kvasirakenteen merkitsevät kollokaatit

### 4.3.1 MI-testi

HAVAITA- ja HUOMATA-kvasirakenteiden synonymiaa rikkovia, hyvin merkitseviä HUOMATA-kvasirakenteen substantiivikollokaatteja, jotka HAVAITA-kvasirakenteen<sup>4</sup>

---

<sup>4</sup> On teoriassa myös mahdollista, että kyseiset substantiivit ovat yhtä tyypillisiä tai jopa tyypillisempiä HAVAITA-kvasirakenteelle, mutta saavuttavat absoluuttisen frekvenssin raja-arvon 5 vain HUOMATA-kvasirakenteen kollokaattina, koska HAVAITA-kvasirakenne on aineistossa moninkertaisesti harvinaisempi kuin HUOMATA-kvasirakenne.

kotekstissa eivät ole merkitseviä, ovat MI-testin mukaan (taulukko 5) TENDENSSI, HÄIVÄHDYS, SIVUSILMÄ, VAHINGONILO, LIIKEHDINTÄ, KIREYS, PILKE, FLIRTTI, VASTAKAIKU, KATSE, ILME, ILKEYS, HALVEKSUNTA ja ELE. Näistä vain kaksi esiintyy HUOMATA-kvasirakenteessa yli 8 kertaa: KATSE saa absoluuttisen frekvenssin 54, ja ILME saa vastaavaksi lukemaksi 17. Matala absoluuttinen frekvenssi ja kyseisten lekseemien harvinaisuus ylipäätään saattaa myös paljolti selittää sen, miksi kyseiset lekseemit eivät ylitä HAVAITA-kvasirakenteen kollokaattina absoluuttisen frekvenssin raja-arvoa 5. HAVAITA-kvasirakenteen kollokaattien absoluuttiset frekvenssit ovat nimittäin HUOMATA-kvasirakenteeseen verrattuna muutenkin keskimäärin huomattavasti matalammat, mikä selittyy sillä, että HUOMATA-kvasirakenne on aineistossa paljon yleisempi. TENDENSSI on MI-testin mukaan kaikista HUOMATA-kvasirakenteen kollokaateista merkitsevin, sillä se saa MI-arvon 7,48. Erittäin korkeat MI-arvot saavat myös HÄIVÄHDYS, SIVUSILMÄ, VAHINGONILO, LIIKEHDINTÄ, KIREYS ja PILKE.

TAULUKKO 5 HUOMATA-kvasirakenteiden 40 merkitsevintä kollokaattia MI-arvon mukaan (HUOMATA-kvasirakenteelle [HAVAITA-kvasirakenteeseen verrattuna] paljon tyypillisemmät kollokaatit lihavoitu).

kollokaatti (HUOMATA-kvasirakenne)	kollokaattiesiintymiä, absoluuttinen frekvenssi <i>f</i>	esiintymiä koko aineistossa, ab- soluuttinen frek- venssi <i>f</i>	MI-arvo	t-arvo
<b>tendenssi</b>	<b>5</b>	<b>1722</b>	<b>7,48</b>	<b>2,22</b>
hienoinen	9	6711	6,37	2,96
<b>häivähdys</b>	<b>5</b>	<b>3834</b>	<b>6,33</b>	<b>2,21</b>
<b>sivusilmä</b>	<b>5</b>	<b>4173</b>	<b>6,21</b>	<b>2,21</b>
ironia	13	15988	5,65	3,53
<b>vahingonilo</b>	<b>5</b>	<b>6881</b>	<b>5,49</b>	<b>2,19</b>
<b>liikehdintä</b>	<b>6</b>	<b>9780</b>	<b>5,24</b>	<b>2,38</b>
<b>kireys</b>	<b>5</b>	<b>8474</b>	<b>5,18</b>	<b>2,17</b>
<b>vilkuilla</b>	<b>13</b>	<b>23748</b>	<b>5,08</b>	<b>3,50</b>
sarkasmi	12	22059	5,07	3,36
<b>pilke</b>	<b>8</b>	<b>15887</b>	<b>4,96</b>	<b>2,74</b>
<b>moikata</b>	<b>13</b>	<b>27536</b>	<b>4,86</b>	<b>3,48</b>
<b>tervehtiä</b>	<b>28</b>	<b>62241</b>	<b>4,79</b>	<b>5,10</b>
välistä	22	68425	4,31	4,45
<b>flirtti</b>	<b>8</b>	<b>25012</b>	<b>4,30</b>	<b>2,68</b>
<b>punastua</b>	<b>5</b>	<b>15829</b>	<b>4,28</b>	<b>2,12</b>
<b>poispäin</b>	<b>5</b>	<b>15876</b>	<b>4,28</b>	<b>2,12</b>
<b>vastakaiku</b>	<b>5</b>	<b>18803</b>	<b>4,03</b>	<b>2,10</b>
<b>katse</b>	<b>54</b>	<b>206513</b>	<b>4,01</b>	<b>6,89</b>
<b>etäinen</b>	<b>5</b>	<b>19702</b>	<b>3,97</b>	<b>2,09</b>
piirre	48	204337	3,86	6,45
<b>ilme</b>	<b>17</b>	<b>82047</b>	<b>3,68</b>	<b>3,80</b>
<b>ilkeys</b>	<b>6</b>	<b>29054</b>	<b>3,67</b>	<b>2,26</b>
<b>kirkastua</b>	<b>5</b>	<b>24758</b>	<b>3,64</b>	<b>2,06</b>
<b>halveksunta</b>	<b>5</b>	<b>25157</b>	<b>3,61</b>	<b>2,05</b>

kiinnostus	39	196475	3,61	5,73
ilmiö	48	244026	3,60	6,36
vastaantulija	6	31101	3,57	2,24
lievä	23	120420	3,56	4,39
<b>teeskennellä</b>	<b>7</b>	<b>36741</b>	<b>3,55</b>	<b>2,42</b>
toisinaan	25	134550	3,52	4,56
<b>jonkunlainen</b>	<b>5</b>	<b>27915</b>	<b>3,46</b>	<b>2,03</b>
<b>flirttailla</b>	<b>5</b>	<b>28175</b>	<b>3,45</b>	<b>2,03</b>
<b>erinäinen</b>	<b>6</b>	<b>34019</b>	<b>3,44</b>	<b>2,22</b>
<b>molemminpuolinen</b>	<b>5</b>	<b>28423</b>	<b>3,44</b>	<b>2,03</b>
<b>vältellä</b>	<b>12</b>	<b>72663</b>	<b>3,35</b>	<b>3,12</b>
jonkinlainen	56	343254	3,33	6,74
sävy	14	87538	3,30	3,36
<b>ele</b>	<b>7</b>	<b>44067</b>	<b>3,29</b>	<b>2,38</b>
olka	8	51405	3,26	2,53

Substantiivikollokaateista kaksi on johdettu adjektiiveista: KIREYS ja ILKEYS. Sanaa KIREYS käytetään HUOMATA-kvasirakenteen kollokaattina tyypillisesti kuvaannollisessa merkityksessä, kuten esimerkissä 34, kun taas sanojen ILME ja ELE on tyypillisesti kirjaimellinen, kuten esimerkeissä 35 ja 36. Kollokaatti PILKE esiintyy lauseissa, joissa kuvaillaan katsomisen tapaa, jolloin sen sijaintiin viitataan sanoilla *silmässä* tai *silmäkulmassa*, kuten esimerkissä 37. Sanaa FLIRTTI voidaan käyttää joko substantiivina tai adjektiivina (KS s.v. *flirtti*). HUOMATA-kvasirakenteen kollokaattina sitä käytetään lähes poikkeuksetta substantiivina, kuten esimerkissä 38.

34) Hannan Leave me alone oli noista kyllä paras mutta olin *huomaavinani* pienoista **kiireyttä** ja jännitystä Hannassa.

35) Hautausmenot tulevat lapsilleni ja kumppanille olemaan liki lottovoitto ja olen ollut *huomaavinani* **ilmeissä** odotuksen makua.

36) Lisäksi aina kun olen yrittänyt pientä lähentymistä, olen ollut *huomaavinani* torjumi-  
seen viittaavan **eleen** ja se on ihan ymmärrettävääkin.

37) Kyllähän ne nuoremmat miehet kiinnostaa mut vaikka olis *huomaavinaan* kiinnostuksen **pilkkeen** toisen silmäkulmassa sitä ei oikein tahdo uskoa että se olis totta.

38) Jos flirttailet, varaudu siihen, että nainen huomauttaa sulle teidän olevan vain ystäviä, nainen ei ole *huomaavinaan* **flirtiäsi** tai sit lähtee siihen mukaan.

Vain HUOMATA-kvasirakenteelle tyypillisistä substantiivikollokaateista kaikki ovat merkitykseltään varsin abstrakteja ja moni viittaa nonverbaaliseen viestintään, erityisesti PILKE, FLIRTTI, KATSE, ILME ja ELE. Suurin substantiivikollokaateista osa on jaollisia; jaottomia ovat HÄIVÄHDYS, SIVUSILMÄ, ELE, KATSE ja ILME sekä joissakin tekstiyhteyksissä PILKE.

MI-testin perusteella HUOMATA-kvasirakenteen adjektiivikollokaatteja, jotka eivät nouse esiin HAVAITA-kvasirakenteen kollokaatteina, ovat ETÄINEN, JONKUNLAINEN, ERINÄINEN ja MOLEMMINPUOLINEN. Näissä mielenkiintoista on se, että kaikki on

muodostettu *-nen*-johtimen avulla ja jokaisen absoluuttinen frekvenssi HUOMATA-kvasirakenteen kollokaattina on varsin matala – korkeintaan 6. Kollokaatti ETÄINEN on näistä ainoa suhteellista ominaisuutta ilmaiseva adjektiivi, ja sen MI-arvo on korkein, 3,97. ETÄINEN viittaa HUOMATA-kvasirakenteen kollokaattina henkiseen suhteeseen tai tunnesuhteeseen (ks. KS s.v. *etäinen*), kuten esimerkissä 39. Sekä JONKUNLAINEN että ERINÄINEN ilmaisevat epämääräisyyttä ja MOLEMMINPUOLINEN puolestaan vastavuoroisuutta. Näiden kolmen kollokaatin MI-arvo on keskenään lähes sama – hieman alle 3,5.

39) Mies katsahti Katariinan ruskeisiin silmiin omilla, syvän metsälammen veden vihreillä silmillään ja Katariina oli *huomaavinaan* niissä oudon, hyvin **etäiseltä** vaikuttavan välähdyn...

Ainoastaan HUOMATA-kvasirakenteen merkitseviä verbikollokaatteja ovat MI-testin tulosten mukaan VILKUILLA, MOIKATA, TERVEHTIÄ, PUNASTUA, KIRKASTUA, TEESKENNELLÄ, FLIRTTAILLA sekä VÄLTELLÄ. Edellä käsiteltyjen substantiivikollokaattien tavoin nämä verbitkin viittaavat merkitykseltään paljolti ihmisten väliseen vuorovaikutukseen, ja nonverbaalinen viestintä korostuu kollokaateissa VILKUILLA, PUNASTUA, FLIRTTAILLA ja TEESKENNELLÄ. Paitsi nämä nonverbaalista viestintää koskevat merkitykset myös verbikollokaattien korkea lukumäärä itsessään osoittautui analyysissäni HAVAITA- ja HUOMATA-kvasirakenteiden synonymiaa erottavaksi tekijäksi, sillä HAVAITA-kvasirakenteen kollokaatteina verbejä ei tutkimusaineistossani juuri esiinny (ks. alaluku 4.2). Merkitsevistä verbikollokaateista suhteellisen moni on frekventatiivimuotoisia: VILKUILLA, TEESKENNELLÄ, FLIRTTAILLA ja VÄLTELLÄ. Frekventatiiviverbit ilmaisevat toistuvaa, ajoittaista tekemistä (VISK § 1735). PUNASTUA ja KIRKASTUA ovat puolestaan adjektiiveista johdettuja intransitiivisia, tapahtumaa ilmaisevia muutosverbejä (ks. VISK § 459). Kuten esimerkeissä 40 ja 41 nähdään, verbikollokaatit VILKUILLA ja VÄLTELLÄ esiintyvät yleensä virkkeissä, joissa ilmaistaan ihmisen nonverbaalista reagointia toisen ihmisen läsnäoloon. Verbit MOIKATA ja TERVEHTIÄ ovat merkitykseltään hyvin lähellä toisiaan, sillä molemmat ovat tervehtimiseen viittaavia kommunikaatioverbejä.

40) Tyttö vaikuttaa ujolta ja hiljaiselta, olen kyllä itekin niin se tekee tästä niin vaikeaa, olin *huomaavinani* että hän olisi **vilkuillut** suuntaani mutta ehkä vain kuvittelin... tai ehkä en.

41) Sitten tuli jostain syystä sellainen tauko, että hän ei ollut *huomaavinaan* minua, lähes **vältteli** minua, syystä jota en tiedä.

Verbeistä HUOMATA-kvasirakenteen merkitsevin kollokaatti on liikkeeseen ja havaintoon viittaava VILKUILLA, jonka MI-arvo on 5,08. Myös verbien MOIKATA ja TERVEHTIÄ MI-arvo on hyvin korkea mutta vähän alle 5. Korkeita MI-arvoja mutta matalia absoluuttisia frekvenssejä saavat verbit PUNASTUA, KIRKASTUA ja FLIRTTAILLA, jotka esiintyvät HUOMATA-kvasirakenteen kollokaattina vain 5 kertaa.

Kollokaatti POISPÄIN on ainoa adverbi, joka on MI-testin perusteella tyypillinen kollokaatti HUOMATA-rakenteelle mutta ei HAVAITA-rakenteelle. Sen merkitys liittyy usein suuntaan, liikkeeseen tai katseeseen, kuten esimerkissä 42.

42) He mollottavat joko suu auki, tai katsovat **poispäin** kun eivät olisi *huomaavinaan*.

### 4.3.2 t-testi

HAVAITA- ja HUOMATA-kvasirakenteiden synonymiaa rikkovia, hyvin merkitseviä HUOMATA-kvasirakenteen kollokaatteja, jotka eivät ole merkitseviä HAVAITA-kvasirakenteen kotekstissa, ovat t-testin tulosten<sup>5</sup> mukaan (taulukko 6) substantiivit NAINEN, KOIRA, KATSE ja MIES, pronominit HÄN, MINÄ, MÄ ja KUKAAN, verbit EI, HUOMATA ja TERVEHTIÄ sekä partikkelit KUN, EDES, TAI, VAIKKA ja MUKA. Substantiivi KATSE ja kommunikaatioverbi TERVEHTIÄ nousivat esiin myös MI-testin tuloksissa kohdassa 4.3.1, vain HUOMATA-kvasirakenteen merkitsevinä kollokaatteina. Huomionarvoista on, että ainostaan HUOMATA-kvasirakenteen merkitsevimpien kollokaattien joukossa ei ole lainkaan adjektiiveja, vaikka kaikissa viidessä edellisessä alaluvun alaluvussa adjektiivikollokaatteja nousi analyysissä esiin. Semanttisesta näkökulmasta tarkasteltuna HUOMATA-kvasirakenteen kollokaateissa korostuvat ihmisiin viittaavat sanat, mikä t-arvojen perusteella on jälleen eräs HAVAITA- ja HUOMATA-kvasirakenteiden välistä synonymiaa rikkova tekijä. Tällaisia nominikollokaatteja ovat substantiivit NAINEN ja MIES, persoonapronominit HÄN, MINÄ ja MÄ sekä kvanttoripronomini KUKAAN.

TAULUKKO 6 HUOMATA-kvasirakenteiden 40 merkitsevintä kollokaattia t-arvon mukaan (HUOMATA-kvasirakenteelle [HAVAITA-kvasirakenteeseen verrattuna] paljon tyypillisemmät kollokaatit lihavoitu; vain HUOMATA-kvasirakenteelle tyypilliset kollokaatit merkitty tähdellä \*).

kollokaatti (HUOMATA-kvasirakenne)	kollokaattiesiintymiä, absoluuttinen frekvenssi <i>f</i>	esiintymiä koko aineistossa, absoluuttinen frekvenssi <i>f</i>	t-arvo	MI-arvo
olla	13286	2,14E+08	85,20	1,94
että	3027	39824984	43,28	2,23
<b>ei*</b>	<b>3662</b>	<b>94962006</b>	<b>35,06</b>	<b>1,25</b>
<b>hän*</b>	<b>666</b>	<b>21698829</b>	<b>12,17</b>	<b>0,92</b>
jokin	406	10054466	12,05	1,32
<b>ettäei</b>	<b>280</b>	<b>5276967</b>	<b>11,62</b>	<b>1,71</b>
joskus	181	2457800	10,49	2,18
mutta	610	22557328	9,88	0,74
<b>minä*</b>	<b>607</b>	<b>22796328</b>	<b>9,63</b>	<b>0,71</b>
pieni	187	3531693	9,49	1,71

<sup>5</sup> On mahdollista, että t-testin arvoissa näkyy se, että HAVAITA-kvasirakenne on aineistossa moninkertaisesti harvinaisempi kuin HUOMATA-kvasirakenne, ja HUOMATA-kvasirakenteen kollokaatit vaikuttavatkin saavan herkemmin korkeita t-arvoja.

itse	415	13982938	9,24	0,87
<b>nainen*</b>	<b>257</b>	<b>7120780</b>	<b>8,83</b>	<b>1,15</b>
<b>huomata*</b>	<b>137</b>	<b>2173305</b>	<b>8,69</b>	<b>1,96</b>
<b>kun*</b>	<b>628</b>	<b>25378736</b>	<b>8,63</b>	<b>0,61</b>
sellainen	185	4319831	8,45	1,40
ainakin	157	3261212	8,31	1,57
kyllä	289	9660595	7,78	0,88
<b>välillä</b>	<b>100</b>	<b>1389907</b>	<b>7,75</b>	<b>2,15</b>
palsta	120	2182503	7,72	1,76
<b>edes*</b>	<b>170</b>	<b>4472771</b>	<b>7,47</b>	<b>1,23</b>
<b>koira*</b>	<b>116</b>	<b>2215799</b>	<b>7,43</b>	<b>1,69</b>
<b>katse*</b>	<b>54</b>	<b>206513</b>	<b>6,89</b>	<b>4,01</b>
jonkinlainen	56	343254	6,74	3,33
<b>tai*</b>	<b>344</b>	<b>13505822</b>	<b>6,73</b>	<b>0,65</b>
kirjoitus	74	1064647	6,59	2,10
jotenkin	70	983516	6,46	2,13
piirre	48	204337	6,45	3,86
ilmiö	48	244026	6,36	3,60
<b>täällä</b>	<b>125</b>	<b>3420099</b>	<b>6,22</b>	<b>1,17</b>
<b>mä*</b>	<b>205</b>	<b>7363086</b>	<b>5,98</b>	<b>0,78</b>
<b>mies*</b>	<b>242</b>	<b>9200069</b>	<b>5,96</b>	<b>0,70</b>
sama	201	7355007	5,76	0,75
kiinnostus	39	196475	5,73	3,61
<b>kukaan*</b>	<b>135</b>	<b>4305773</b>	<b>5,61</b>	<b>0,95</b>
<b>vaikka*</b>	<b>205</b>	<b>7850402</b>	<b>5,42</b>	<b>0,69</b>
muutos	45	563947	5,34	2,30
viesti	67	1535850	5,14	1,43
<b>tervehtiä*</b>	<b>28</b>	<b>62241</b>	<b>5,10</b>	<b>4,79</b>
<b>muka*</b>	<b>43</b>	<b>601973</b>	<b>5,07</b>	<b>2,14</b>
keskustelu	60	1312183	5,00	1,50

Kollokaatit NAINEN ja MIES ovat konkreettisia ja jaottomia substantiiveja, joissa korostuu paitsi ihmisen täysikasvuisuus myös sukupuoli. Näistä kahdesta NAINEN on t-testin tulosten mukaan HUOMATA-rakenteelle tyypillisempi, sillä se on 12:nneksi merkitsevin, kun taas MIES on taulukon 6 kollokaateista 31:nneksi merkitsevin. Ne esiintyvät melko usein samassa lauseessa keskenään, kuten esimerkissä 43.

43) Siis jos **mies** ei ole *huomaavinaan* selvää **naisen** lähentelyä, ei hän todennäköisesti ole kiinnostunut ko. naisesta.

Persoonapronomini HÄN on todella paljon tyypillisempi kollokaatti HUOMATA-kuin HAVAITA-kvasirakenteelle. Se on t-testin tulosten mukaan HUOMATA-kvasirakenteen 4:nneksi merkitsevin kollokaatti ja samalla myös 4:nneksi yleisin kollokaatti. Selvästi suurin osa HÄN-kollokaatin esiintymistä on noodin oikeanpuoleisessa koteks-tissa, kuten esimerkissä 44. Kollokaatti MINÄ esiintyy HUOMATA-kvasirakenteen kollokaattina usein kvasirakenteellisen lauseen objektina, yksikön partitiivimuodossa, kuten esimerkissä 45, mutta melko usein myös subjektina yksikön nominatiivissa.

Pronominin arkikielisempi variantti MÄ puolestaan esiintyy useammin noodin vasemmanpuoleissa kuin oikeassa kotekstissa ja useammin kvasirakenteellisen lauseen subjektina kuin kollokaatti MINÄ. Kieltohakuinen kvanttoripronomini KUKAAN esiintyy selvästi useammin noodin vasemman- kuin oikeanpuoleisessa kotekstissa, hyvin usein osana kielteistä kvasirakenteilmausta *kukaan ei ole huomaavinaan(kaan)*, kuten esimerkissä 46.

44) Tyttö ei ollut *huomaavinaan häntä* ja käyttäytyi kuin miestä ei olisi olemassakaan.

45) Joskus taas hän ei ole *huomaavinaanakaan minua* vaikka kuinka yritän vilkuilla häneen päin.

46) Ja tietenkään **kukaan** ei ole *huomaavinaanakaan* että se kaikkien suussa paskalle maistuva kakku on valmistettu samalla reseptillä.

Kaikkein merkittävin yksittäinen ero t-testin tuloksissa HAVAITA- ja HUOMATA-kvasirakenteiden välillä on se, että kieltosana EI on erittäin merkitsevä HUOMATA-kvasirakenteen kollokaatti, kun taas HAVAITA-kvasirakenteen kotekstissa se on hyvin epätyypillinen. Kollokaatti EI on siis tekijä, joka varsin vahvasti rikkoo HAVAITA- ja HUOMATA-kvasirakenteiden välistä synonymiaa siten, että vain HUOMATA-kvasirakenteella on preferenssi kieltoverbiin. Kielto sanaa EI voidaan käyttää taipumattomana kieltopartikkelina esimerkiksi kieltävän vastauksen alussa tai kiteytyneissä vastauksissa (VISK § 1625), mutta HUOMATA-kvasirakenteen kollokaattina se esiintyy lähes aina kvasirakenteeseen osana toimivan OLLA-verbin apuverbinä, kuten esimerkissä 47. Kieltoverbinä EI kongruoi subjektin kanssa persoonassa ja luvussa (ks. VISK § 1623). Havaintoa ilmaiseva mentaaliverbi HUOMATA, joka on HUOMATA-kvasirakenteen varaloverbinä, esiintyy myös samaisen kvasirakenteen kollokaattina hyvin usein: se on t-testin tulosten mukaan rakenteen 13:nneksi merkitsevin kollokaatti. HUOMATA esiintyy kollokaattina yleensä sellaisissa lauseissa, joissa asetetaan vaihtoehdoiksi HUOMATA-kvasirakenteen tuoma merkitys ja HUOMATA-verbi ilman kvasirakennetta, kuten esimerkissä 48. Tällaisissa lauseissa puhuja ei siis ole varma siitä, eikä joku tietty henkilö huomaa vai esittäkö hän vain, ettei huomaisi. Tällöin kvasirakenne on yleensä yksikön 3. persoonassa eli muodossa *huomaavinaan*.

47) Tai uhkaurhea oikeistohallitus yrittää kiristellä suhteitamme Venäjään, Venäjä **ei** ole *huomaavinaan* sitä.

48) Rouva ei **huomannut tai** ei ollut *huomaavinaan*, kun katsoin suoraan hameensa alle.

Partikkelikollokaateista suurin osa on konjunktioita: adverbiaalikonjunktio KUN ja VAIKKA sekä rinnastuskonjunktio TAI. Adverbiaalikonjunktio KUN on erittäin yleinen HUOMATA-kvasirakenteen kollokaatti, jonka esiintymiä on 5:nneksi eniten – 628. Se on t-testin tulosten mukaan kollokaateista 14:nneksi merkitsevin. Se ilmaisee usein aikaa, kuten esimerkissä 49, ehtoa tai syytä, kuten esimerkissä 50. Adverbiaalikonjunktio VAIKKA esiintyy HUOMATA-kvasirakenteen kollokaattina useimmiten noodin

oikeanpuoleisessa kotekstissa, merkityksessä 'siitä huolimatta että' (KS. s.v. *vaikka*), kuten esimerkissä 51. Rinnastuskonjunktio TAI yhdistää eri vaihtoehtoisia sanoja tai lauseita (ks. KS. s.v. *tai*). HUOMATA-kvasirakenteen kollokaattina yhtenä rinnastetuista vaihtoehdoista on yleensä itse HUOMATA-kvasirakenne, ja kollokaatti TAI esiintyy yleensä noodin vasemmanpuoleisessa kotekstissa eli HUOMATA-kvasirakenteen edellä. Nämä seikat ilmenevät myös edellä esimerkissä 48.

49) Yleensä se kaikki tuntui minusta ihanalta, mutta oli hetkiä, **kun** olin *huomaavinani* jonkin menevän liian pitkälle.

50) Ihan voimavarasyistä, **kun** olin *huomaavinani*, että aloittajan kuvaamat ihmiset vievät suunnattomasti energiaa kun vastavuoroisuus ei toimi.

51) Ei ole sinua *huomaavinaankaan*, **vaikka** on hetki sitten luritellut korvaasi kauniita asioita.

EDES on fokuspartikkeli, jota käytetään tavallisesti muissa kuin määränilmauksissa (VISK § 848). Se muodostaa polaarisen parin fokuspartikkelin JOPA kanssa, joka ei t-testin tulosten mukaan ole tyypillinen HUOMATA- vaan HAVAITA-kvasirakenteelle (ks. kohta 4.2.2). JOPA-sanaa käytetään tavallisesti myönteisissä ja EDES-sanaa kielteisissä yhteyksissä (VISK § 849). Kollokaatti EDES on selvästi yleisempi HUOMATA-kvasirakenteen vasemmassa kuin oikeassa kotekstissa. Partikkelikollokaattia MUKA käytetään kommenttiadverbiaalina, jonka avulla puhuja yleensä ilmaisee pitävänsä referoimaansa toimintaa teeskentelynä. Sen vaikutusalana on yleensä nimenomaan HUOMATA-kvasirakenne. MUKA esiintyy käytännössä aina vasemmassa kotekstissa ja yleensä kielteisissä lauseissa, kuten esimerkissä 52. (Ks. VISK § 1495.)

52) Jotkut tuijottivat saappaitani kiinnostuneina, toiset eivät **muka** olleet *huomaavinaankaan*.

HUOMATA-kvasirakenteelle hyvin tyypillisiä kollokaatteja (kuuluvat 40 merkitsevimmän kollokaatin joukkoon), jotka ovat HAVAITA-kvasirakenteelle jokseenkin tyypillisiä (kuuluvat 80 merkitsevimmän mutta eivät 40 merkitsevimmän kollokaatin joukkoon), ovat adverbit VÄLILLÄ ja TÄÄLLÄ sekä partikkeli ETTÄEI. Adverbi VÄLILLÄ esiintyy HUOMATA-kvasirakenteen kollokaattina yleensä ajanilmauksena, merkityksessä 'silloin tällöin', 'harvoin' tai 'hetkittäin', kuten esimerkissä 53. Kyseessä on siis tällöin toistuvuuden adverbi (ks. VISK § 651). Adverbikollokaatti TÄÄLLÄ esiintyy selvästi useammin HUOMATA-kvasirakenteen oikean- kuin vasemmanpuoleisessa kotekstissa. Se viittaa yleensä Suomi24-keskustelupalstaan tai palstalla käytävään viestiketjuun, kuten esimerkissä 54. Paikka, johon viitataan ei siis ole fyysinen vaan virtuaalinen ympäristö. Kollokaatti ETTÄEI on suomen kielen selvästi yleisin sulaumakonjunktio, jossa yhdistyvät lekseemit ETTÄ ja EI (ks. VISK § 141). Sillä on HUOMATA-kvasirakenteen kollokaattista 6:nneksi korkein t-arvo, eli se on erittäin merkitsevä kollokaatti. Se on hyvin yleinen sekä noodin vasemmalla että oikealla puolella. Noodin oikealla puolella se tavallisesti aloittaa HUOMATA-kvasirakenteen objektina olevan sivulauseen, kuten esimerkissä 54.



53) Olen aina ajatellut että olemme vain ystäviä, vaikka **välillä** olen *huomaavinani* kiinnostusta hänen puheistaan aina välillä.

54) Ja näin sivusta seuranneena olen *huomaavinani* **ettei täällä** varmaan kukaan tosissaan kirjoittele ja solvaa toisia.

## 5 PÄÄTÄNTÖ

### 5.1 Yhteenveto

Tutkimuksellani oli kaksi tavoitetta: Ensinnäkin halusin tutkia, millaisia merkitseviä kollokaatteja HAVAITA- ja HUOMATA-kvasirakenteet saavat suomenkielisissä verkko-keskusteluissa. Toiseksi pyrin selvittämään, mitä näiden kollokaattien avulla voidaan päätellä kyseisten kvasirakenteiden synonyymisyydestä. Kollokaatioanalyysini perusteella voidaan todeta, että verkkokeskusteluissa esiintyvien HAVAITA- ja HUOMATA-kvasirakenteiden merkitsevistä kollokaateista hyvin monet ovat sanaluokaltaan substantiiveja. Niiden lisäksi etenkin HAVAITA-kvasirakenteelle tyypillisistä kollokaateista monet ovat adjektiiveja. HUOMATA-kvasirakenteelle tyypilliset kollokaatit ovat sen sijaan selvästi useammin verbejä, partikkeleja tai pronomineja kuin HAVAITA-kvasirakenteelle tyypilliset kollokaatit.

Tarkastelemieni kvasirakenteiden välistä synonymiaa rikkovaksi tekijäksi osoitautui analyysissäni erityisesti kollokaatti EI, joka on HUOMATA-kvasirakenteelle hyvin tyypillinen mutta HAVAITA-kvasirakenteelle hyvin epätyypillinen. HUOMATA-kvasirakenne on siis tämän havainnon perusteella kieltohakuisempi (ks. VISK § 1634) kuin HAVAITA-kvasirakenne. Verbikollokaattien näkökulmasta kvasirakenteiden synonyymiaa rikkoo myös se, että HUOMATA-kvasirakenne saa merkitseviksi kollokaateikseen verbejä paljon useammin kuin HAVAITA-kvasirakenne. Niiden joukossa on sekä konkreettisia, mentaalisia että kommunikaatioverbejä. Osa HUOMATA-kvasirakenteen saamista verbikollokaateista viittaa merkitykseltään ihmisten väliseen non-verbaaliseen viestintään, esimerkiksi VILKUILLA ja TEESKENNELLÄ. HAVAITA-kvasirakenteen kollokaattien joukossa tällaisia oli aineistossa vähemmän, mikä myös osaltaan rikkoo tarkastelemieni kvasirakenteiden välistä synonyymiaa. Kaikille kvasirakenteille yhteinen apuverbi OLLA on luonnollisesti kaikkein yleisin kollokaatti sekä HAVAITA- että HUOMATA-kvasirakenteen pääverbin kotekstissa – ja t-testin mukaan

myös kaikkein merkitsevin. Merkittävä osa HUOMATA-kvasirakenteen verbikollokaateista on frekventatiiviverbejä.

HAVAITA- ja HUOMATA-kvasirakenteiden partikkelikollokaateista varsin moni on kollokaatioanalyysini mukaan fokuspartikkeleja, kuten AINAKIN, MYÖS ja EDES. Konjunktiot ovat tutkimustulosteni mukaan tyypillisiä etenkin HUOMATA-kvasirakenteelle. Tarkastelemieni kvasirakenteiden merkitseviä konjunktiokollokaatteja ovat esimerkiksi ETTÄ, MUTTA, KUN ja TAI.

Tarkastelemieni kvasirakenteiden substantiivikollokaatit ovat tutkimustulosteni perusteella paljolti abstrakteja, ja niiden joukossa on paljon sekä jaollisia että jaottomia substantiiveja. Kvasirakenteiden välistä synonymiaa kuitenkin hieman rikkoo HUOMATA-kvasirakenteen t-testitulosten mukaan se, että HUOMATA-kvasirakenteen kollokaatteina esiintyy tyypillisesti konkreettisia ja jaottomia ihmisiin viittaavia substantiiveja ja pronomineja, kuten NAINEN, MIES, HÄN ja MINÄ. Metakeskusteluun eli itse verkkokeskustelusta keskustelemiseen viittaavat substantiivikollokaatit, kuten KIRJOITUS, PALSTA ja KOMMENTTI, yhtäältä yhdistävät, mutta jossain määrin myös rikkovat tarkastelemieni kvasirakenteiden välistä synonymiaa: ne ovat erityisen tyypillisiä HAVAITA-kvasirakenteelle, mutta HUOMATA-kvasirakenteen tekstiympäristöissä metakeskustelu ei korostu läheskään yhtä voimakkaasti. Molemmat kvasirakenteet kollokoivat varsin voimakkaasti ivallisuuteen viittaavien sanojen kanssa, joita ovat substantiivit SARKASMI ja IRONIA.

HAVAITA- ja HUOMATA-kvasirakenteiden adjektiivikollokaatteja voi luonnehtia tyypillisemmin absoluuttisiksi kuin suhteellisiksi. Adjektiivikollokaatit esiintyvät varsin tyypillisesti partitiivimuodossa. Kvasirakenteiden synonymiaa yhdistää sekin, että molempien rakenteiden merkitsevimmistä adjektiivikollokaateista suhteellisen merkittävä osa on proadjektiivieja, kuten JONKINLAINEN, SELLAINEN ja TÄLLAINEN, ja osa adjektiivikollokaateista viittaa merkitykseltään 'vähäisyyteen': HIENOINEN, LIEVÄ ja PIENI.

## 5.2 Tutkimuksen arviointi ja pohdinta

Tutkimukseni vahvuutena on, että suuren korpusaineiston ja käyttämieni tilastollisten menetelmien myötä tutkimustuloksiani voidaan pitää varsin luotettavina. Ne ovat empiirisiä, perustuvat todelliseen kielenkäyttöön ja tarjoavat sellaista tietoa, jota yksinomaan kielellisen intuition avulla ei voi saada. Tulosten pohjalta ei voida kuitenkaan luoda sellaisia tieteellisiä yleistyksiä, jotka koskisivat suomen kieltä kokonaisuudessaan, koska käytin aineistonani yksinomaan verkkokeskusteluista muodostettua korpusta enkä esimerkiksi jotakin suomenkielistä yleiskorpusta. Tutkimukseni vahvuutena voidaan pitää sitäkin, että käyttämiäni menetelmiä hyödyntäen vastaavan tutkimuksen voisi toistaa muillakin korpusaineistoilla. Eri korpusaineistoilla tehtyjen tutkimusten tulokset myös lisäisivät tutkimiani kvasirakenteita koskevan tieteellisen

tiedon yleistettävyyttä. Kollokaatioanalyysiani voi pitää kattavana ja syvällisenä, sillä tarkastelin merkitseviä kollokaatteja vertailevalla otteella varsin systemaattisesti useasta eri näkökulmasta. Tutkimukseni laadulliseen puoleen keskittyessäni kollokaatioanalyysini systemaattisuutta ja objektiivisuuteen pyrkimistä luonnollisesti haastoi ajoittain se, että laadullisten tulkintojen tekeminen edellyttää jossain määrin omaan intuitioon nojaamista.

Tutkimustulosteni luotettavuutta laskee hieman se, että käyttämäni Suomi24-korpus sisältää lemmatisointivirheitä, kuten saneiden yhdistämistä väärin lemmoihiin. Niitä on syntynyt automaattisessa lemmatisoinnissa, joka on tehokasta ja taloudellista muttei virheetöntä. Myös kirjoitusvirheet, joita anonyymissä verkkokeskustelussa esiintyy suhteellisen paljon, johtavat automaattisessa lemmatisoinnissa siihen, että sanetta ei tunnisteta, vaan se luokitellaan omaksi lemmakseen, mikä voi vääristää tuloksia. Lemmatisointi- ja kirjoitusvirheiden vaikutus tutkimustuloksiin on kuitenkin lopulta varsin vähäinen, ja korpuksen suuri koko vähentää yksittäisten virheiden painoarvoa.

Tutkimukseni merkitystä korostaa se, että suomenkielisten verbirakenteiden kollokaatteja ja synonymiaa on toistaiseksi tutkittu hyvin vähän. Rakenteisiin keskittyvä näkökulmani fraseologia-, kollokaatio- ja synonymiatutkimukseen on siis huomattavasti harvinaisempi kuin esimerkiksi yksinomaan lekseemeihin keskittyvä näkökulma. Tutkimalla kollokaatteja systemaattisesti korpusaineiston avulla pystytään lisäämään tieteellistä tietoa siitä, minkälaisissa yhteyksissä lekseemejä tai rakenteita käytetään, mikä on hyvin keskeistä niiden syvällisen, kielitieteellisen ymmärtämisen kannalta. Fraseologiseen viitekehukseen perustuvan tutkimukseni tuloksia onkin mahdollista hyödyntää käytännössä esimerkiksi verbien käyttöä ja verbirakenteita opeteltaessa tai sanakirjatyössä (ks. Jantunen, Kumpulainen, Tammimies & Tokola 2013; Hanks 2012), suomi toisena kielenä -opetuksessa (ks. Jantunen 2009a; ks. Granger & Meunier 2008b) ja ylipäätään kielenkäytön kuvauksissa, joissa käsitellään suomen kielen kvasirakennetta tai verbien välisiä synonymiasuhteita. Kielenoppimisen näkökulmasta yksittäistenkin kollokaattien opettaminen on tärkeää ja hyödyllistä (ks. Nesselhauf 2005). Sekä synonyymien välisten erojen että kollokaattien oppiminen saatetaan nimittäin kokea erityisen haasteelliseksi (Tsui 2005: 338–339).

Suomen kielen kvasirakennetta ja synonymiaa olisi kiinnostavaa tutkia tulevaisuudessa tarkastelemalla muistakin keskenään lähisynonyymisistä verbeistä muodostettuja kvasirakenteita, esimerkiksi KÄSITTÄÄ- ja YMMÄRTÄÄ- ja TAJUTA-kvasirakenteita, jotka ovat tyypillisiä ainakin verkkokeskustelujen kielenkäytölle (Kumpulainen 2019: 10). Kiinnostavaa voisi myös olla toteuttaa tämä tutkimukseni muutoin samanaikaisena, mutta erilaisella aineistolla kuin verkkokeskustelukorpus, ja vertailla eri aineistoilla saatuja tuloksia sen selvittämiseksi, minkälainen vaikutus tekstilajilla on tämänkaltaisten tutkimusten tuloksiin. Toisaalta samaa aineistoa tutkimalla, mutta keskittymällä semanttiseen prosodiaan, voisi puolestaan tehdä lisää kiinnostavia havainnot negaatioelementeistä, negatiivisista konteksteista sekä kieltohakuisuudesta, jonka

havaitsin tässä maisterintutkimuksessani rikkovan HAVAITA- ja HUOMATA-kvasirakenteiden välistä synonymiaa.

## LÄHTEET

- Baker, Paul 2006: *Using corpora in discourse analysis*. London: Continuum.
- Barnbrook, Geoff 1996: *Language and computers. A practical introduction to the computer analysis of language*. Edinburgh University Press.
- Biber, Douglas, Conrad, Susan & Reppen, Randi 2006: *Corpus linguistics. Investigating structure and use*. 5. painos. Cambridge: Cambridge University Press.
- Church, Kenneth & Hanks, Patrick 1990: Word association norms, mutual information, and lexicography. – *Computational Linguistics* 16 (1) s. 22–29.
- Cruse, Alan 2000: *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Daley, Robert, Jones, Susan, Sinclair, John & Krishnamurthy, Ramesh. *English Collocation Studies: The OSTI Report*. London: Continuum.
- Forsman-Svensson, Pirkko 2000: Vanhan kirjasuomen kvasirakenne. *Sanajalka* 41 (1) s. 43–58.
- Granger, Sylviane & Meunier, Fanny 2008a: Introduction. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspective* s. XIX–XXVIII. Amsterdam: John Benjamins.
- & Meunier, Fanny 2008b: *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- & Paquot, Magali 2008: Disentangling the phraseological web. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology. An interdisciplinary perspective* s. 27–49. Amsterdam: John Benjamins.
- Gries, Stefan 2008: Phraseology and linguistic theory. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology. An interdisciplinary perspective* s. 3–25. Amsterdam: John Benjamins.
- Hakulinen, Lauri 1979: *Suomen kielen rakenne ja kehitys*. 4. painos. Helsinki: Otava.
- Hanks, Patrick 2012: Corpus evidence and electronic lexicography. – Sylviane Granger & Magali Paquot (toim.), *Electronic Lexicography* luku 4. Oxford: Oxford University Press.
- Harju, Auli 2017: Suomi24-käyttäjäkyselyn tuloksia. *Citizen Mindscapes -hankkeen verkkosivusto*. – <https://blogs.helsinki.fi/citizenmindscapes/2017/08/21/suomi24-kayttajakyselyn-tuloksia/> 5.4.2021.
- 2018. Suomi24-keskustelut kohtaamisten ja törmäysten tilana. *Media & viestintä* 41 (1) s. 51–74. – <https://journal.fi/mediaviestinta/article/view/69952/31051> 5.4.2021.
- Heinonen, Tarja 2013: *Idiomien leksikaalinen kuvaus kielenkäytön ja vaihtelun näkökulmasta*. Väitöskirja. Helsingin yliopiston nykykielten laitos. – [https://helda.helsinki.fi/bitstream/handle/10138/37918/heinonen\\_vaitoskirja.pdf](https://helda.helsinki.fi/bitstream/handle/10138/37918/heinonen_vaitoskirja.pdf) 6.4.2021.
- Hoey, Michael 1991: *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- 2005: *Lexical priming. New theory of words and language*. London: Routledge.
- Ikola, Osmo, Palomäki, Ulla & Koitto, Anna-Kaisa 1989: *Suomen murteiden lauseoppia ja tekstikielioppia*. Helsinki: Suomen Kirjallisuuden Seura.
- Jantunen, Jarmo 2001: ”Tärkeä seikka” ja ”keskeinen kysymys”: Mitä korpuslingvistinen analyysi paljastaa lähisynonymeista?. *Virittäjä* 105 (2) s. 170–192. – <https://journal.fi/virittaja/article/view/40082> 3.2.2021.

- 2004: *Synonymia ja käännössuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin*. Joensuun yliopiston humanistisia julkaisuja 35. Joensuu: Joensuun yliopistopaino.
- 2009a: Minulla on aivan paljon rahaa – Fraseologiset yksiköt suomen kielen opetuksessa. *Virittäjä* 113 (3) s. 356–381.
- 2009b. ”Ei pelkäästään mielikuviuksen puutteen vuoksi – Kieliaineistojen systemaattinen käyttö kielen tutkimuksessa”. *Virittäjä* 113 (1) s. 101–113. – <https://journal.fi/virittaja/article/view/4171> 13.5.2021.
- 2012: Akateemiset ja populaaritekstit korpusvertailussa: sanalistat, avainsanat ja fraseologiset yksiköt. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi: Tekstilajitutkimuksen käytäntöä* s. 186–213. Kotimaisten kielten keskuksen verkkojulkaisuja 29. Helsinki: Kotimaisten kielten keskus.
- & Kumpulainen, Tammimies, & Tokola 2013: Korpuspohjaista oppijansanakirjaa tekemässä: esimerkkiinä ConLexis. *Lähivördlusi. Lähivertailuja* 23 s. 89–118. – <https://jyx.jyu.fi/handle/123456789/45668> 14.4.2021.
- Kangas, Akseli 2018: *Adverbien täysin ja kokonaan kollokaatit ja semanttinen prosodia*. Maisterintutkielma. Jyväskylän yliopiston kieli- ja viestintätieteiden laitos. – <http://urn.fi/URN:NBN:FI:JYU-201808153839> 3.2.2021.
- Kangasniemi, Heikki 1997: *Sana, merkitys, maailma. Katsaus leksikaalisen semantiikan perusteisiin*. Helsinki: Finn Lectura.
- Karlsson, Fred 2004: *Yleinen kielitiede*. Uudistetun laitoksen 2. painos. Helsingin yliopistopaino.
- KS = *Kielitoimiston sanakirja*. Helsinki: Kotimaisten kielten keskus. – <https://www.kielitoimiston-sanakirja.fi/> 29.10.2020.
- Kumpulainen, Aaro 2019: ”Olen aistivoinani jotain sellaista”: Kvasirakenteelle tyypillisimmät verbit Suomi24-keskusteluissa. Kandidaatintutkielma. Jyväskylän yliopiston kieli- ja viestintätieteiden laitos. – <http://www.urn.fi/URN:NBN:FI:JYU-201912165317> 6.10.2020.
- Kumpulainen, Iiro 2020: *Collocateanalysis*. – <https://github.com/MusserO/CollocateAnalysis> 28.12.2020.
- Laakkonen, Maria 2017: *Lähisynonyymien keho, vartalo ja ruumis kontekstuaalista semantiikkaa*. Maisterintutkielma. Oulun yliopisto. – <http://jultika.oulu.fi/files/nbnfioulu-201706012333.pdf> 4.2.2021.
- Lagus, Krista, Ruckenstein, Minna, Pantzar, Mika & Ylisiurua, Marjoriikka 2016: *Suomi24: muodonantoa aineistolle*. Valtiotieteellisen tiedekunnan julkaisuja 10. Helsinki: Helsingin yliopisto. – <http://hdl.handle.net/10138/163190> 27.12.2020.
- Leech, Geoffrey 1990: *Semantics. The Study of Meaning*. 2. painos. Harmondsworth: Penguin Books.
- Lyons, John 1968: *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- 1995: *Linguistic semantics. An introduction*. Cambridge: Cambridge University Press.
- Manning, Christopher & Hinrich Schütze 1999: *Foundations of statistical natural language processing*. Cambridge: MIT press. – [https://www.cs.vassar.edu/~cs366/docs/Manning\\_Schuetze\\_StatisticalNLP.PDF](https://www.cs.vassar.edu/~cs366/docs/Manning_Schuetze_StatisticalNLP.PDF) 2.12.2020.

- Nesselhauf, Nadja 2005: *Collocations in a Learner Corpus*. Philadelphia: John Benjamins.
- Partington, Alan 1998: *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins.
- 2004: "Utterly content in each other's company". Semantic prosody and semantic preference. – *International Journal of Corpus Linguistics* 9 (1) s. 131–156.
- Pirkola, Silja 2016: *Synonyymien EHKÄ ja MAHDOLLISESTI kollokaatit ja semanttiset preferenssit*. Maisterintutkielma. Jyväskylän yliopiston kielten laitos. – <http://urn.fi/URN:NBN:FI:JYU-201602021400> 3.2.2021.
- Salminen, Taru 2000: *Morfologiasta moniäänisyyteen. Suomen kielen kvasirakenteen merkitys, käyttö ja kehitys*. Väitöskirja. SKST 796. Helsinki: Suomen Kirjallisuuden Seura.
- Setälä, Emil 1926: *Suomen kielen lauseoppi*. 11. painos. Helsinki: Kustannusosakeyhtiö Otava.
- Sinclair, John 1991: *Corpus, concordance, collocation. Describing English Language*. Oxford: Oxford University Press.
- & Carter, Ronald 2004. *Trust the text: Language, corpus and discourse*. London: Routledge.
- 2008: Preface. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspective* s. XV–XVIII. Amsterdam: John Benjamins.
- Stubbs, Michael 1996: *Text and corpus analysis. Computer-assisted studies of language and culture*. Oxford: Blackwell.
- 2001: *Words and Phrases. Corpus Studies of Lexical Semantics*. Malden: Blackwell.
- Suomi24 2020: *Suomi24 Keskustelun ohjeet*. – <https://www.suomi24.fi/opastus/keskustelu> 29.10.2020.
- Suoraniemi, Henriikka 2011: *Vuoksi ja takia – korpuspohjainen tutkimus lähisyronyymeistä*. Maisterintutkielma. Oulun yliopisto.
- Tognini-Bonelli, Elena 2001: *Corpus linguistics at work*. Studies in corpus linguistics 6. Amsterdam: John Benjamins.
- Tsui, Amy 2005: ESL teachers' questions and corpus evidence. – *International Journal of Corpus Linguistics* 10 (3) s. 335–356.
- Tuuri, Emilia 2010: *Postpositioiden vuoksi ja takia merkitys, käyttö ja synonymia*. Kandidaatintutkielma. Tampereen yliopiston kieli- ja käännöstieteiden laitos.
- 2014: *Adpositiot kausaalisten suhteiden ilmaisussa. Lähimerkityksisyyden korpuspohjaista analyysia*. Maisterintutkielma. Tampereen yliopiston kieli-, käännös- ja kirjallisuustieteiden yksikkö. – <http://urn.fi/URN:NBN:FI:UTA-201405261475> 3.2.2021.
- Vilkuna, Maria 2003: *Suomen lauseopin perusteet*. 2. painos. Helsinki: Edita.
- VISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. – <http://scripta.kotus.fi/visk> 30.10.2020.

### Aineistolähteet:

- Aller Media Oy 2020: *The Suomi24 Corpus 2001–2017, VRT version 1.1 [tekstikorpus]*. Kielipankki. – <http://urn.fi/urn:nbn:fi:lb-2020021801> 1.12.2020.



# LIITTEET

## LIITE 1

HAVAITA- ja HUOMATA-kvasirakenteiden 140 merkitsevintä kollokaattia MI-arvojen mukaan järjestettyinä.

kollokaatti (HAVAITA-kvasi- rakenne)	kollo- kaatti- esiinty- miä, ab- soluutti- nen frek- venssi <i>f</i>	esiinty- miä koko aineis- tossa, ab- soluutti- nen frek- venssi <i>f</i>	MI- arvo	kollokaatti (HUOMATA-kvasira- kenne)	kollo- kaatti- esiinty- miä, ab- soluutti- nen frek- venssi <i>f</i>	esiintymiä koko ai- neistossa, absoluutti- nen frek- venssi <i>f</i>	MI- arvo
hienoinen	13	6711	9,06	tendenssi	5	1722	7,48
sarkasmi	16	22059	7,65	hienoinen	9	6711	6,37
ironia	11	15988	7,57	häivähdys	5	3834	6,33
jonkinasteinen	6	19136	6,44	sivusilmä	5	4173	6,21
yhtäläisyys	5	19315	6,16	ironia	13	15988	5,65
lievä	30	120420	6,10	vahingonilo	5	6881	5,49
tietynlainen	10	54454	5,66	liikehdintä	6	9780	5,24
piirre	31	204337	5,39	kireys	5	8474	5,18
ristiriita	23	153000	5,37	vilkuilla	13	23748	5,08
katkeruus	16	117307	5,23	sarkasmi	12	22059	5,07
jonkinlainen	43	343254	5,11	pilke	8	15887	4,96
välistä	8	68425	5,01	moikata	13	27536	4,86
epätoivo	5	44462	4,96	tervehtiä	28	62241	4,79
älyllinen	6	53412	4,95	välistä	22	68425	4,31
pyrkimys	5	59851	4,53	flirtti	8	25012	4,30
sävy	7	87538	4,46	punastua	5	15829	4,28
kateus	12	156687	4,40	poispäin	5	15876	4,28
rivi	13	215868	4,06	vastakaiku	5	18803	4,03
kirjoitus	57	1064647	3,89	katse	54	206513	4,01
ilmiö	13	244026	3,88	etäinen	5	19702	3,97
kiinnostus	9	196475	3,66	piirre	48	204337	3,86
toisinaan	6	134550	3,62	ilme	17	82047	3,68
havaita	15	377623	3,45	ilkeys	6	29054	3,67
ajattelu	7	212945	3,18	kirkastua	5	24758	3,64
tietty	29	892025	3,17	halveksunta	5	25157	3,61
asenne	12	370956	3,16	kiinnostus	39	196475	3,61
pieni	108	3531693	3,08	ilmiö	48	244026	3,60
tällainen	41	1370279	3,05	vastaantulija	6	31101	3,57
kommentti	21	710549	3,03	lievä	23	120420	3,56
lause	8	271448	3,02	teeskennellä	7	36741	3,55
kehitys	9	327326	2,92	toisinaan	25	134550	3,52
uskonnollinen	6	258680	2,68	jonkunlainen	5	27915	3,46
palsta	48	2182503	2,60	flirttailla	5	28175	3,45
jotenkin	21	983516	2,56	erinäinen	6	34019	3,44
muutos	12	563947	2,55	molemminpuolinen	5	28423	3,44

kirjoittaja	11	518900	2,55	vältellä	12	72663	3,35
merkki	19	913190	2,52	jonkinlainen	56	343254	3,33
joskus	51	2457800	2,52	sävy	14	87538	3,30
samanlainen	15	726243	2,51	ele	7	44067	3,29
hieman	26	1279822	2,49	olka	8	51405	3,18
nimenomaan	10	517290	2,42	viittaus	5	33910	3,18
viesti	28	1535850	2,33	tietynlainen	8	54454	3,18
uskovainen	9	493681	2,33	tervehdys	6	43663	3,08
tällöin	7	384718	2,33	virkillijä	7	51591	3,06
viime	18	1016523	2,29	mukamas	10	74757	3,04
ketju	11	631116	2,27	tyly	5	37471	3,04
käytös	5	288317	2,26	rivi	28	215868	3,00
jokin	167	10054466	2,20	tuijottaa	14	114222	2,92
outo	8	484934	2,19	vilkaista	10	81939	2,91
että	636	39824984	2,14	ohitse	6	50622	2,87
jälleen	6	377005	2,14	tapainen	8	67807	2,86
semmoinen	7	451535	2,10	ristiriita	18	153000	2,86
teksti	14	910093	2,09	vanhuus	5	43047	2,84
suunta	11	738243	2,04	älyllinen	6	53412	2,79
olla	3131	2,14E+08	2,02	turvotus	5	44860	2,78
nimimerkki	8	559935	1,98	mamma	5	45711	2,75
ainakin	46	3261212	1,96	seurailia	5	46823	2,72
puhe	10	716525	1,95	mieliala	6	56619	2,71
puute	5	358388	1,95	katkeruus	12	117307	2,66
aiemmin	7	520743	1,89	ihastus	11	110247	2,62
pikku	5	373010	1,89	mustasukkaisuus	5	51619	2,58
oire	6	450220	1,88	kirjoittelu	10	104018	2,57
vaikutus	7	529585	1,87	kateus	15	156687	2,56
eilen	6	460357	1,85	käytävä	5	54930	2,49
esiintyä	6	477235	1,80	tarkkailla	8	88202	2,48
ilma	6	483584	1,78	kasvot	25	279279	2,46
ero	18	1461866	1,77	ajoittain	5	56649	2,44
sellainen	53	4319831	1,76	selkä	33	377341	2,43
tuossa	5	414763	1,73	päinvastainen	6	68823	2,43
tosin	18	1498736	1,73	hymy	13	153247	2,39
selvä	12	1001109	1,73	ylimielinen	6	70773	2,39
lisääntyä	5	417214	1,73	pieru	5	59187	2,38
hiukan	5	425543	1,70	muualle	16	195233	2,34
suhteen	8	691328	1,68	epävarmuus	5	61611	2,32
keskustelu	15	1312183	1,66	pystyssä	10	125111	2,30
tuttu	7	626520	1,62	muutos	45	563947	2,30
kuitenkin	40	3610904	1,61	samankaltainen	6	75490	2,29
yhä	5	452388	1,61	feministi	7	91269	2,24
lähinnä	7	636634	1,60	touhuta	6	78430	2,24
selvästi	6	546704	1,60	naureskella	5	65738	2,23
usea	8	736929	1,58	että	3027	39824984	2,23
välillä	15	1389907	1,57	seikka	13	175573	2,19
jopa	22	2050378	1,57	joskus	181	2457800	2,18
sijaan	6	582533	1,51	ohi	33	454393	2,16

nimittäin	5	507294	1,44	välillä	100	1389907	2,15
kirjoitella	5	509345	1,44	muttaei	26	362760	2,14
lisäksi	16	1635241	1,43	muka	43	601973	2,14
kohtaan	7	727525	1,41	kiusaaaja	6	84134	2,14
liike	6	632949	1,39	jotenkin	70	983516	2,13
sama	67	7355007	1,33	myönteinen	5	71498	2,11
eräs	9	988606	1,33	hermostua	5	71499	2,11
myös	85	9832708	1,25	kirjoitus	74	1064647	2,10
moni	40	4763827	1,21	vastapuoli	5	72796	2,08
toisaalta	6	715480	1,21	karkuun	5	73870	2,06
ääni	9	1073841	1,21	pilkka	5	74627	2,05
tuollainen	12	1438015	1,20	väsytys	5	75069	2,04
kyllä	78	9660595	1,16	autoilija	7	107910	2,00
kiinnostua	5	624510	1,14	narttu	5	78354	1,98
usein	12	1507638	1,14	keskuudessa	10	157660	1,97
kulua	6	762845	1,12	huomata	137	2173305	1,96
uskonto	7	894654	1,11	olut	10	160168	1,94
seurata	8	1069117	1,05	olla	13286	2,14E+08	1,94
vaihe	6	834256	0,99	meininki	7	114657	1,91
itse	100	13982938	0,98	hyppiä	5	82126	1,91
melkein	5	699371	0,98	nokka	6	98981	1,90
joukko	6	854647	0,95	työkaveri	7	115685	1,90
täällä	24	3420099	0,95	näet	5	83511	1,88
muutama	15	2210893	0,91	pikemminkin	6	100914	1,87
nuori	12	1798823	0,88	Hesari	5	84211	1,87
näin	23	3516446	0,85	suhtautua	24	409786	1,85
toiminta	7	1092206	0,82	hymyillä	11	188944	1,84
erilainen	6	952029	0,80	kääntää	23	404518	1,81
verran	7	1133715	0,77	esimies	5	88047	1,81
sillä	10	1622264	0,77	toimittaja	13	229445	1,80
mainita	7	1150831	0,75	katsella	41	741034	1,77
ettäei	32	5276967	0,74	palsta	120	2182503	1,76
aiheuttaa	7	1164967	0,73	ettäei	280	5276967	1,71
mutta	135	22557328	0,72	pieni	187	3531693	1,71
seksi	6	1022226	0,70	monesti	16	303370	1,70
peruste	6	1057741	0,65	pysähtyä	10	190348	1,70
pari	12	2116865	0,65	koira	116	2215799	1,69
vastaus	9	1590868	0,64	etenkin	16	306050	1,69
kyseinen	6	1063561	0,64	kävellä	24	466883	1,66
josei	5	933388	0,56	oire	23	450220	1,66
joku	38	7117381	0,56	reaktio	5	98100	1,65
tämä	147	27634109	0,55	merkki	46	913190	1,63
kohta	14	2665108	0,54	isäntä	7	143415	1,59
silmä	7	1339070	0,53	käyttäytyminen	7	143430	1,59
juuri	18	3477521	0,51	tuttava	7	144253	1,58
nainen	36	7120780	0,48	moinen	20	412852	1,58
tuo	64	12924224	0,45	kirjoittaja	25	518900	1,57
muuten	17	3443390	0,45	ainakin	157	3261212	1,57
vika	6	1245595	0,41	silla	8	168163	1,55

vaikea	7	1467859	0,40	motiivi	5	105968	1,54
enemmän	12	2539663	0,38	tietty	42	892025	1,54
mielipide	5	1064462	0,37	anoppi	6	128586	1,52
sinä	63	13454653	0,37	vähettä	8	173738	1,51
puoli	11	2355738	0,37	asenne	17	370956	1,50
johtua	5	1071680	0,36	huolestua	5	109297	1,50
ajatus	6	1287418	0,36	keskustelu	60	1312183	1,50

## LIITE 2

HAVAITA- ja HUOMATA-kvasirakenteiden 140 merkitsevintä kollokaattia t-arvojen mukaan järjestettyinä.

kollokaatti (HAVAITA-kvasi- rakenne)	kollo- kaatti- esiinty- miä, ab- soluutti- nen frek- venssi <i>f</i>	esiinty- miä koko aineis- tossa, ab- soluutti- nen frek- venssi <i>f</i>	t- arvo	kollokaatti (HUOMATA-kvasira- kenne)	kollo- kaatti- esiinty- miä, ab- soluutti- nen frek- venssi <i>f</i>	esiintymiä koko ai- neistossa, absoluutti- nen frek- venssi <i>f</i>	t-arvo
olla	3131	2,14E+08	42,13	olla	13286	2,14E+08	85,20
että	636	39824984	19,50	että	3027	39824984	43,28
jokin	167	10054466	10,10	ei	3662	94962006	35,06
pieni	108	3531693	9,16	hän	666	21698829	12,17
kirjoitus	57	1064647	7,04	jokin	406	10054466	12,05
jonkinlainen	43	343254	6,37	ettäei	280	5276967	11,62
joskus	51	2457800	5,89	joskus	181	2457800	10,49
palsta	48	2182503	5,79	mutta	610	22557328	9,88
tällainen	41	1370279	5,63	minä	607	22796328	9,63
piirre	31	204337	5,43	pieni	187	3531693	9,49
lievä	30	120420	5,40	itse	415	13982938	9,24
myös	85	9832708	5,36	nainen	257	7120780	8,83
sellainen	53	4319831	5,13	huomata	137	2173305	8,69
ainakin	46	3261212	5,04	kun	628	25378736	8,63
itse	100	13982938	4,93	sellainen	185	4319831	8,45
sama	67	7355007	4,93	ainakin	157	3261212	8,31
kyllä	78	9660595	4,87	kyllä	289	9660595	7,78
tietty	29	892025	4,79	välillä	100	1389907	7,75
ristiriita	23	153000	4,68	palsta	120	2182503	7,72
mutta	135	22557328	4,59	edes	170	4472771	7,47
kuitenkin	40	3610904	4,26	koira	116	2215799	7,43
viesti	28	1535850	4,24	katse	54	206513	6,89
hieman	26	1279822	4,19	jonkinlainen	56	343254	6,74
kommentti	21	710549	4,02	tai	344	13505822	6,73
sarkasmi	16	22059	3,98	kirjoitus	74	1064647	6,59
katkeruus	16	117307	3,89	jotenkin	70	983516	6,46
tämä	147	27634109	3,87	piirre	48	204337	6,45
jotenkin	21	983516	3,81	ilmiö	48	244026	6,36

merkki	19	913190	3,60	täällä	125	3420099	6,22
hienoinen	13	6711	3,60	mä	205	7363086	5,98
moni	40	4763827	3,60	mies	242	9200069	5,96
havaita	15	377623	3,52	sama	201	7355007	5,76
rivi	13	215868	3,39	kiinnostus	39	196475	5,73
viime	18	1016523	3,37	kukaan	135	4305773	5,61
ilmiö	13	244026	3,36	vaikka	205	7850402	5,42
kateus	12	156687	3,30	muutos	45	563947	5,34
ironia	11	15988	3,30	viesti	67	1535850	5,14
samanlainen	15	726243	3,19	tervehtiä	28	62241	5,10
jopa	22	2050378	3,11	muka	43	601973	5,07
tietynlainen	10	54454	3,10	keskustelu	60	1312183	5,00
asenne	12	370956	3,08	kuitenkin	110	3610904	4,90
ero	18	1461866	2,99	tosin	62	1498736	4,79
tosin	18	1498736	2,96	myös	231	9832708	4,70
muutos	12	563947	2,87	tällainen	58	1370279	4,70
teksti	14	910093	2,86	selkä	33	377341	4,68
kiinnostus	9	196475	2,76	rivi	28	215868	4,63
kirjoittaja	11	518900	2,75	merkki	46	913190	4,60
välistä	8	68425	2,74	toisinaan	25	134550	4,56
keskustelu	15	1312183	2,65	katsella	41	741034	4,53
ketju	11	631116	2,63	ohi	33	454393	4,46
kehitys	9	327326	2,60	välistä	22	68425	4,45
välillä	15	1389907	2,57	lievä	23	120420	4,39
nimenomaan	10	517290	2,57	tietty	42	892025	4,25
sävy	7	87538	2,53	vaan	233	10372193	4,24
lisäksi	16	1635241	2,52	hieman	50	1279822	4,14
suunta	11	738243	2,51	kasvot	25	279279	4,09
lause	8	271448	2,48	eräs	42	988606	4,01
jonkinasteinen	6	19136	2,42	moni	121	4763827	3,97
selvä	12	1001109	2,42	sinä	285	13454653	3,95
uskovainen	9	493681	2,40	muttaei	26	362760	3,94
älyllinen	6	53412	2,37	usein	53	1507638	3,92
täällä	24	3420099	2,37	viime	41	1016523	3,83
ajattelu	7	212945	2,35	seurata	42	1069117	3,80
puhe	10	716525	2,34	ilme	17	82047	3,80
ettäei	32	5276967	2,28	ristiriita	18	153000	3,66
toisinaan	6	134550	2,25	kommentti	32	710549	3,62
outo	8	484934	2,21	suhtautua	24	409786	3,54
yhtäläisyys	5	19315	2,20	ironia	13	15988	3,53
epätoivo	5	44462	2,16	joku	160	7117381	3,52
tuo	64	12924224	2,15	vilkuilla	13	23748	3,50
näin	23	3516446	2,14	moikata	13	27536	3,48
pyrkimys	5	59851	2,14	kääntää	23	404518	3,43
tällöin	7	384718	2,12	sävy	14	87538	3,36
nimimerkki	8	559935	2,11	sarkasmi	12	22059	3,36
uskonnollinen	6	258680	2,07	kävellä	24	466883	3,35
semmoinen	7	451535	2,03	samanlainen	30	726243	3,33
joku	38	7117381	1,98	kirjoittaja	25	518900	3,32

tuollainen	12	1438015	1,96	varsinkin	31	777460	3,30
suhteen	8	691328	1,94	oire	23	450220	3,27
aiemmin	7	520743	1,93	tuijottaa	14	114222	3,25
vaikutus	7	529585	1,92	toisaalta	29	715480	3,23
jälleen	6	377005	1,89	kateus	15	156687	3,22
usein	12	1507638	1,89	muualle	16	195233	3,21
usea	8	736929	1,88	vältellä	12	72663	3,12
eräs	9	988606	1,81	teksti	32	910093	3,05
muutama	15	2210893	1,80	koko	101	4344323	3,04
sinä	63	13454653	1,80	asiakas	26	663994	2,99
tuttu	7	626520	1,79	muuten	83	3443390	2,98
oire	6	450220	1,78	moinen	20	412852	2,97
lähinnä	7	636634	1,77	hienoinen	9	6711	2,96
käytös	5	288317	1,77	ero	43	1461866	2,94
eilen	6	460357	1,77	hymy	13	153247	2,92
esiintyä	6	477235	1,74	katkeruus	12	117307	2,91
ilma	6	483584	1,73	jatkaa	35	1130548	2,82
ääni	9	1073841	1,70	seikka	13	175573	2,82
nainen	36	7120780	1,70	pentu	19	424557	2,78
puute	5	358388	1,66	mukamas	10	74757	2,78
kohtaan	7	727525	1,65	ihastus	11	110247	2,78
selvästi	6	546704	1,64	katsoa	72	2988094	2,77
pikku	5	373010	1,63	monesti	16	303370	2,77
sijaan	6	582533	1,59	etenkin	16	306050	2,76
nuori	12	1798823	1,58	suunta	26	738243	2,75
tuossa	5	414763	1,56	vilkaista	10	81939	2,74
lisääntyä	5	417214	1,56	pilke	8	15887	2,74
hiukan	5	425543	1,55	jopa	53	2050378	2,71
liike	6	632949	1,51	opettaja	19	442815	2,71
yhä	5	452388	1,50	vaikutus	21	529585	2,71
seurata	8	1069117	1,46	tyyppi	22	573309	2,71
uskonto	7	894654	1,42	flirtti	8	25012	2,68
nimittäin	5	507294	1,41	sijaan	22	582533	2,68
kirjoitella	5	509345	1,41	asenne	17	370956	2,66
toisaalta	6	715480	1,39	ketju	23	631116	2,66
kulua	6	762845	1,32	äläkä	8	31080	2,65
sillä	10	1622264	1,30	eilen	19	460357	2,65
juuri	18	3477521	1,27	kirjoittelu	10	104018	2,63
pari	12	2116865	1,25	nuori	47	1798823	2,60
kiinnostua	5	624510	1,22	toimittaja	13	229445	2,57
vaihe	6	834256	1,22	olka	8	51405	2,53
joukko	6	854647	1,19	jutella	18	449104	2,53
kohta	14	2665108	1,16	pystyssä	10	125111	2,52
toiminta	7	1092206	1,15	tietynlainen	8	54454	2,52
melkein	5	699371	1,10	tapainen	8	67807	2,44
muuten	17	3443390	1,10	nimenomaan	19	517290	2,43
verran	7	1133715	1,09	myyjä	16	388182	2,43
vastaus	9	1590868	1,08	aiemmin	19	520743	2,42
mainita	7	1150831	1,07	teeskennellä	7	36741	2,42

aiheuttaa	7	1164967	1,05	hymyillä	11	188944	2,39
erilainen	6	952029	1,04	outo	18	484934	2,39
seksi	6	1022226	0,94	liikehdintä	6	9780	2,38
peruste	6	1057741	0,89	ele	7	44067	2,38
kyseinen	6	1063561	0,88	tuolla	25	811518	2,37
silmä	7	1339070	0,81	keskuudessa	10	157660	2,35
enemmän	12	2539663	0,81	semmoinen	17	451535	2,35
vähän	16	3594887	0,74	olut	10	160168	2,34
puoli	11	2355738	0,74	virkillijä	7	51591	2,33
josei	5	933388	0,72	tarkkailla	8	88202	2,32
vaikea	7	1467859	0,64	aluksi	13	286077	2,32
eri	10	2218395	0,62	käytös	13	288317	2,31
vika	6	1245595	0,61	ilkeys	6	29054	2,26
kerta	19	4519678	0,60	vastaantulija	6	31101	2,24