

TARTON YLIOPISTO  
HUMANISTISTEN TIETEIDEN JA TAIDEAINEIDEN TIEDEKUNTA  
VIRON KIELEN JA YLEISEN KIELITIETEEN LAITOS  
SUOMALAIS-UGRILAISTEN KIELTEN OSASTO

Ewa Pikkell

KIELENOPPIJA FRASEOLOGISESSA VIIDAKOSSA:  
ASTEMÄÄRITTEIDEN *AIVAN*, *ERITTÄIN*, *IHAN*, *KOVIN*, *OIKEIN* JA *SANGEN*  
KÄYTTÖKONTEKSTIT

Maisterintutkielma

Ohjaaja Tiina Liisa Marjatta Nahkola

Tartto 2020

# SISÄLLYS

JOHDANTO .....	4
1. ASTEMÄÄRITTE	7
1.1. Ominaisuuksia .....	7
1.2. Luokitteluongelmia .....	10
2. FRASEOLOGIAA .....	13
2.1. Brittiläisestä näkökulmasta korpuslingvistiikkaan .....	15
2.2. Syntagmaattisia suhteita .....	19
2.2.1. Kollokaatio .....	19
2.2.2. Klusteri .....	21
2.2.3. Kolligaatio .....	22
2.2.4. Semanttinen preferenssi .....	22
2.2.5. Semanttinen prosodia .....	23
3. KIELENOPPIJA FRASEOLOGISESSA VIIDAKOSSA .....	24
3.1. Ympäristön vaikutus oppijankieleen .....	25
3.2. Astemääritteet oppijankielessä .....	27
4. ASTEMÄÄRITTEIDEN AIVAN, ERITTÄIN, IHAN, KOVIN, OIKEIN JA SANGEN KONTRASTIIVINEN ANALYYSI .....	30
4.1. Aineisto .....	31
4.1.1. Eduskunta- ja ICLFI-korpuksat .....	31
4.1.2. Aineiston muokkaaminen .....	32
4.2. Astemääritteiden käyttäminen natiivi- ja oppijanaineistossa .....	35
4.2.1. Astemääritteiden kontekstuaalisia piirteitä natiivien kielessä .....	38
4.2.1.1. Kolligaatio .....	39
4.2.1.2. Kollokaatio .....	47
4.2.1.3. Klusteri .....	57
4.2.2. Astemääritteiden kontekstuaaliset erot natiivi- ja oppijanaineiston välillä .....	59
4.2.2.1. <i>Aivan</i> vasemmanpuoleinen konteksti .....	59
4.2.2.2. <i>Aivan</i> oikeanpuoleinen konteksti .....	61
4.2.2.3. <i>Erittäin</i> vasemmanpuoleinen konteksti .....	63

4.2.2.4.	<i>Erittäin</i> oikeanpuoleinen koteksti.....	65
4.2.2.5.	<i>Ihan</i> vasemmanpuoleinen koteksti.....	68
4.2.2.6.	<i>Ihan</i> oikeanpuoleinen koteksti .....	70
4.2.2.7.	<i>Kovin</i> vasemmanpuoleinen koteksti .....	72
4.2.2.8.	<i>Kovin</i> oikeanpuoleinen koteksti.....	74
4.2.2.9.	<i>Oikein</i> vasemmanpuoleinen koteksti .....	76
4.2.2.10.	<i>Oikein</i> oikeanpuoleinen koteksti.....	79
4.2.2.11	<i>Sangen</i> vasemmanpuoleinen koteksti .....	81
4.2.2.12.	<i>Sangen</i> oikeanpuoleinen koteksti.....	82
4.3.	Pohdintaa .....	83
	YHTEENVETO .....	89
	LÄHTEET.....	91
	KEELEÖPPIJA FRASEOLOGIA TIHNİKUS: MÄÄRAMÄÄRSÖNADE AIVAN, <i>ERITTÄIN, IHAN, KOVIN, OIKEIN</i> JA <i>SANGEN</i> KASUTUSKOHAD. RESÜMEE. 101	
	Liite 1. Oikeanpuoleiset adjektiiv-, adverbi- ja kvanttorikolligaatit Eduskunta- aineistossa. ....	104

## JOHDANTO

Välillä toisen sanan käyttö tuntuu kontekstissa luontevalle, mutta sen synonyymien puolestaan ei. Natiivi pystyy valitsemaan kontekstiin sopivan sanan intuitiivisesti, mutta kielenoppija joutuu usein hankaluksiin. Tässä tutkimuksessa tarkastellaan intensiteettiä ilmaisevien astemääritteiden käyttöä ensiksi natiivien kielenkäytössä ja sen jälkeen analysoidaan, miten oppijoiden käyttämä kieli vastaa natiivien kielenkäyttöä.

Havaintoyksiköiksi valittiin suomen kielen astemääritteet *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen*. Nämä astemääritteet ovat toisaalta niin samanlaisia, että sanakirjat käsittelevät niitä usein synonyymisinä sanoina (KTS), toisaalta jokaisella niistä on kirjallisuuden perusteella omia erityispiirteitään. Vaikka ne kaikki ilmaisevat ominaisuuden suurta astetta, erikoistuvat *aivan* ja *ihan* ominaisuuden suurimman asteen ilmaisemiseen (Jantunen 2004: 63; 71–73). *Ihan*-määritteen on todettu olevan melko puhekielinen (Jantunen 2015: 112) ja *sangen*-määritteen kirjakielinen (VISK: § 853). Lisäksi näkyy, että muutamat astemääritteet kuten *erittäin* esiintyvät yleensä myönteisessä kontekstissa, mutta muutamat kuten *kovin* tavallisimmin kielteisessä kontekstissa (VISK § 666).

Aiemmin suomen kielen astemääritteitä on tutkittu eri näkökulmista. Esimerkiksi Jarmo Harri Jantunen (2004) vertaili väitöskirjassaan sanojen *hyvin*, *oikein* ja *kovin* käyttöä natiivi- ja käännössuomessa. Myöhemmin hän on tarkastellut myös sanan *aivan* kontekstuaalisia erityispiirteitä Virittäjän sivuilla (2009). Lisäksi Hanna Lappalainen (2001) on tarkastellut pääkaupunkiseudun puhekieltä ja kiinnittänyt huomiota muun muassa astemääritteisiin. Suvi Halme (2017) puolestaan tutki pro gradu -tutkielmassaan sanan *sangen* kehitystä vanhan kirjasuomen kaudelta varhaisnykysuomen kauden loppupuolelle (1540–1870-luku).

Varsinaista oppijansuomeen keskittyvää tutkimusta astemääritteistä ei tietääkseni ole. Niitä on sivunnut oppimiskontekstin vaikutuksen näkökulmasta esimerkiksi Jaana Isohäätä (2014) pro gradu -tutkielmassaan. Oppimiskontekstin vaikutusta astemääritekäyttöön on pohtinut myös Jantunen (2015) Lähivördlusi. Lähivertailuja -sivuilla.

Tässä tutkimuksessa astemääritteitä lähestytään Jantusen (2004, 2009) tavoin tarkastelemalla niiden kontekstuaalisia erityispiirteitä. Tarkemmin keskitytään kolligatioon ja kollokatioon. Tutkimuksen lähtökohtana ovat brittiläisen fraseologisen koulukunnan ideat siitä, että jokaisen sanan käyttöä ohjaavat tietynlaiset säännöt tai preferenssit. Siis jokaisen sanan erityispiirteet ovat tarkisteltavissa nimenomaan toistuvien käyttöyhteyksien perusteella. Nämä toistuvuudet on poimittava todellisesta kielenkäytöstä. Todellisena kielenkäyttönä nähdään myös korpusaineistossa sisältyvät tekstit. (ks. luku 3) Tässä tutkimuksessa kielenoppijoiden tekstit on kerätty Korp-hakuohjelman (Korp) kautta Kansainvälisestä oppijansuomen korpukselta (ICLFI) ja natiivien tekstit Eduskunnan täysistuntoja sisältävästä korpukselta (Eduskunta).

Oman intuitioni ja kokemukseni perusteella pystytän hypoteesin, että suuresti synonyymisten astemääritteiden käyttö on kielenoppijalle haasteellista. Sen tarkastelemiseksi luodaan tutkimuksessa ensiksi katsaus astemääritteiden erityispiirteisiin natiivien kielenkäytössä ja sen jälkeen tutkitaan, miten kielenoppijoiden astemääritekäyttö vastaa natiivien astemääritekäyttöä. Kielenoppijoiden taustamuuttujien on oltava mahdollisimman samanlaista, koska oppijoiden kielenkäyttöön vaikuttavat esimerkiksi äidinkieli, taitotaso ja se, opitaanko kieltä kohdekulttuurissa vai sen ulkopuolella (ks. luku 3.1.). Sen takia analyysiin otetaan B1–B2-tasoisten suomea vieraana kielenä opiskelevien virolaisten astemääritekäyttö.

Tutkimustavoitteiden toteuttamiseksi esitän seuraavat tutkimuskysymykset:

1. Millaisia eroja ja yhtäläisyyksiä astemääritteillä *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* on keskenään natiivien kielenkäytössä?
2. Miten B1–B2-tasoisten virolaisten suomen oppijoiden kielenkäyttö vastaa natiivien astemääritevalintoja?

Tässä tutkimuksessa vertaillaan saman kielen eri variantteja, eli kyseessä on kontrastiivinen intralingvaalinen analyysi (*Contrastive Interlanguage Analysis* eli CIA). Kontrastiivisen intralingvaalisen analyysin yhtenä tutkimussuuntana on nimenomaan natiivien ja kielenoppijoiden tuottaman kielen vertailu (Granger 1996: 43). Sen tavoitteena on tarkastella, miten oppijankieli vastaa natiivien kielenkäyttöä (Granger 1998: 13). Kontrastiivisessa intralingvaalisessa analyysissä kvantitatiivisten menetelmien hyödyntäminen nähdään antoisana, koska frekvenssien avulla voi tutkia erilaisten kielenpiirteiden yli- ja aliedustumia (Granger 1996: 45). Näistä periaatteista lähdetään liikkeelle myös tässä tutkimuksessa, joka perustuu nimenomaan kvantitatiiviseen analyysiin. Analyysissä sanojen toistuvien käyttöyhteyksien löytämiseksi käytettiin WordSmith Tools 8.0 ohjelmaketin Concord-ohjelmaa ja yksityiskohtaisempi kvantitatiivinen analyysi toteutettiin Excel-ohjelman avulla.

Tämä tutkimus koostuu neljästä suuremmasta luvusta. Ensimmäisessä luvussa luodaan katsaus astemääritteisiin ja toisessa esitellään fraseologista näkökulmaa. Kolmannessa luvussa kiinnitetään huomiota siihen, miksi kielenoppija voi joutua vaikeuksiin valitessaan tekstiin sopivaa sanaa. Niiden havaintojen valossa tarkastellaan myös kielenoppijoiden astemääritekäyttöä niiltä osin, mitä siitä tiedetään nykyisten tutkimusten perusteella. Sen jälkeen neljännessä luvussa siirrytään analyysiin. Ensiksi esitellään tarkemmin Eduskunta- ja ICLFI korpuksia, annetaan kuvaus hakuprosessissa esiin tulleista haasteista ja niiden ratkaisemisesta. Sitten vertaillaan keskenään natiivien ja kielenoppijoiden astemääritteiden käyttöfrekvenssejä, tarkastellaan astemääritteiden käyttöä natiivien kielessä ja vertaillaan oppijoiden astemääritevalintoja natiivien astemääritevalintoihin. Luvun lopussa pohditaan tuloksia.

# 1. ASTEMÄÄRITE

Astemäärite-luvun ensimmäisessä alaluvussa annetaan yleiskuva astemääritteiden ominaisuuksista ja käytöstä. Astemääritteiden luokittelu sekä sanaluokkana että sanaluokan sisällä ei ole yhtenäistä ja päivänselvää. Myös terminologian kannalta esiintyy paikoin ristiriitaa. Toisessa alaluvussa käsitellään tarkemmin erilaisia luokitteluun liittyviä näkökulmia ja niiden välisiä vastakohtia. Sen perusteella selitetään tämän tutkimuksen astemääritteisiin liittyvää terminologian valintaa.

## 1.1. Ominaisuuksia

Astemääritteet (*degree modifiers*) ovat puolikieliopillisia sanoja, joilla ei ole suoraa tarkoitetta kielenulkoisessa maailmassa (Jantunen 2015: 110). Ne ilmaisevat toisaalta jonkun ominaisuuden, tavan tai määrän suurta (*erittäin hienosti*), kohtalaista (*melko mukava*) tai vähäistä astetta (*hieman vinossa*) (VISK: § 664), toisaalta puhujan sitoutumista tai suhtautumista aiheeseen (Biber 1988: 106; ks myös Lorenz 1999: 24). Syntaktisesti astemääritteet ovat adjektiivien ja tavan, ajan, paikan, määrän adverbien etumääritteitä (VISK: määritelmät; § 665), mutta lisäksi ne voivat määrittää kvanttoreita (*hyvin vähän, kovin moni*) ja adpositioita/ adpositiolausekkeita (*hyvin lähellä kuolemaa; aivan ilman apua*) (Jantunen 2004: 70; 95; Hakulinen, Karlsson 1988: 155).

Pragmaattisessa mielessä astemääritteet ilmaisevat, missä määrin ominaisuus pätee kielenkäyttäjän mielestä: miten suuressa arvossa jotakin pidetään tai miten merkittäväksi piirteeksi tai määräksi jokin nähdään (Jantunen 2015: 111–112). Normi, johon kielenkäyttäjä tukeutuu, on implisiittinen ja ”mittarin” osoitinta ohjaavat kriteerit voivat

olla yksilölliset (Orpana 1988: 197–198). Siksi astemääritteet ilmaisevat subjektiivisia asenteita (VISK: § 666; Jantunen 2015: 111–112).

Ominaisuuden suurta astetta ilmaisevat astemääritteet voi jakaa kahtia. Ominaisuuden hyvin suurta astetta ilmaisevat ns. *vahventajat* (*boosters*) kuten *hyvin, kovin, erittäin, oikein, sangen* ja ominaisuuden suurinta astetta *maksimoijat* (*maximizers*) kuten *aivan, ihan* (Jantunen 2004: 63, 71–73; Kennedy 2003: 469). Rajat vahventavan ja maksimaalisen asteen välillä voivat olla häilyviä (Jantunen 2004: 72). Joskus jotkut vahventajat ja maksimoijat esitetäänkin synonyymisinä. Esimerkiksi Kielitoimiston sanakirjassa (KTS) *aivan* selitetään muun muassa synonyymien *ihan, hyvin, kovin, erittäin* kautta, joista viimeiset kolme esitetään myös *sangen* synonyymeinä.

Astemäärite on kattokäsite melko heterogeeniselle luokalle. Tyypillisin astemäärite on joko partikkelimainen ja lähes opaakki (*lähes, järin*) tai nominin genetiivimuoto (*suhteellisen, hirveän*). Pienempi määrä on nominatiivimuotoisia sanoja (*tosi, sika*) ja adverbejä (*tarpeeksi, liian*). (VISK: § 615, § 664). Muutamista astemääritteistä käytetään myös komparatiivi- ja superlatiivimuotoja (*vähemmän, eniten*) (VISK: määritelmät).

Astemääritteet eroavat toisistaan myös kombinatoristen mahdollisuuksiensa puolesta. Muutamia astemääritteitä toimivat yleismääritteinä, joiden kombinatoriset suhteet ovat vapaampia (*hyvin, todella, erittäin*), muutamien kombinaatiosuhteet ovat taas tiukempia ja ilmentävät vain vähän variaatiota. (Jantunen 2015: 111; ks. myös Granger 2001: 147) Esimerkiksi *järin* (usein myös *kovin* ja *erityisen*) esiintyy vain negatiivisessa kontekstissa, samalla kun *erittäin, hyvin, todella* ja *tosi* ovat tavallisimmin positiivisessa kontekstissa (VISK § 666).

Jantunen (2015) mukaan astemääritteiden luokka on produktiivinen, muuttuva ja kirjava (ks. myös Lorenz 1999: 25–26). Kun tuntuu, että olemassa olevat astemääritteet ovat menettäneet ilmaisuvoimansa, on mahdollista ketjuttaa niitä (*ihan sairaan sika siisti*). (Jantunen 2015: 112) Merkitystä voi intensifioida myös toistolla. Esimerkiksi *hyvin* voi duplikoitua (*hyvin hyvin*), triplikoitua (*hyvin, hyvin, hyvin*) tai esiintyä jopa äärettömän monta kertaa peräkkäin. (Räsänen 2010: 27, 210) Sen lisäksi luodaan uusia intensiteetiltään voimakkaamman tuntuksia sanoja, esimerkiksi (*sairaana hyvä, über*

*kaunis*). (Jantunen 2015: 112) VISK (§ 664) täsmentää, että produktiivinen on juuri nominien genetiivimuotoisten astemääritteiden ryhmä.

Suurin piirtein samaa merkitseviä astemääritteitä on paljon. Sille luokalle on ominaista laaja synonyymikirjo. Samaa merkitsevät astemääritteet ovat erikoistuneet erilaisiin käyttökonteksteihin eli astemääritteet ovat rekisteri- ja tekstilajisidonnaisia (vrt. *erittäin hyvä – tosi hyvä – sika hyvä – vitun hyvä*). (Jantunen 2015: 111; ks. myös Yaguchi ym. 2010: 586; 590; Orpana 1988: 167). Suomen kielestä on huomautettu, että kirjoitetulle tekstille ominaisia ovat *perin, sangen, varsin ja ylen* (VISK: § 853) ja puhekielelle *ihan* (Jantunen 2015: 112). Jantunen (2004: 69) olettaa, että käsiteltävä aihepiiri ei vaikuta astemääritteiden käyttöön.

Astemääritteiden valinta voi riippua myös käyttötilanteesta (Yaguchi ym. 2010: 590). Muutamia astemääritteet ovat ominaisia informaatiota antavalle ja vähäistä interaktiota edellyttävälle konteksteille (*expository talk*), kuten julkiselle puheelle, jotkut taas sellaiselle, jossa edellytään vahvaa interaktiota ja puhujalla on affektiivinen asenne kuuntelijaa kohtaan (Yaguchi ym. 2010: 587, 590). Kun yritetään olla objektiivisia ja mahdollisimman ”piilossa” (esim. akateemisessa käytössä), valitaan tyypillisesti sellaisia astemääritteitä, jotka eivät korosta kirjoittajan emotionaalista sitoutuneisuutta (*emotional involvement*) tekstiin (Lim ja Hong 2012: 142–143).

Suomen kielestä Terttu Orpana (1988) huomauttaa, että neutraaleja ominaisuuden asteen osoittimia ovat opaakit astemääritteet (*oikein, hyvin*), koska niiden kantasanojen kvalitatiiviset komponentit ovat täydellisesti hävinneet niistä ja jäljellä on vain kvantifioivia komponentteja. Vastaavasti läpinäkyvissä (*tavattoman, mahdottoman*) on jossain määrin säilynyt kantasanojen kvalitatiivisia komponentteja. (Mts. 166) Kvalitatiivisten ominaisuuksien hävitessä myös astemääritteen liittymisvapaus suurenee (vrt. *\*tappavan lyhyet hiukset*) (Jantunen 2004: 70).

Astemääritteiden käyttöön vaikuttavat myös sosiolingvistiset tekijät. Gunter R. Lorenzin (1999: 25) mukaan sukupolvet (etenkin nuoret) luovat uusia määritteitä ja näin on eri sukupolvilla omat suosikkimääritteensä, esimerkiksi 1990-luvun loppupuolella teinikäisillä *ab fab (absolutely fabulous)*, *bloody brill (brilliant)* ja *dead cool*. Yaguchin ym.

(1999: 590) tutkimistulosten mukaan myös sukupuoli vaikuttaa astemääritevalintaan sen mukaan, että naisten kielenkäyttö on miesten kielenkäyttöä formaalimpaa, mutta samalla interaktiivisempaa (ks. *real/really* käyttö).

## 1.2. Luokitteluongelmia

Orpanan (1988: 57) mukaan suomen kielen adverbien tutkimuksessa on ollut esillä lähinnä kaksi peruskysymystä: miten luokitella adverbejä sanaluokkakategoriana ja miten kategorisoida adverbejä luokan sisällä. Tämä kysymys nousi esiin myös tätä tutkimusta kirjoitettaessa silloin kun huomattiin, että käsitteiden sisällöt eivät limity aina.

Jantusen (2015: 110) mukaan astemääritteet ovat sanaluokaltaan adverbejä, tarkemmin intensiteettiadverbejä. Hän toteaa (kuitenkin perustuen VISKiin), että astemääritteiden luokka on muodoltaan heterogeeninen, johon sisältyy toisaalta sekä opaakkeja, läpinäkyviä genetiivimuotoisia, että nominatiivimuotoisia astemääritteitä, toisaalta myös komparatiivi- ja superlatiivimuotoja (Jantunen 2015: 111) Jantunen käyttää siitä monimuotoisesta luokasta (katto)termiä *astemäärite*.

Kattoterminä *astemäärite* toimii myös Ison suomen kielenopin määrittelyssä (VISK: määritelmät). VISKin mukaan tyypillisin astemäärite on intensiteettimäärite, mutta astetta voi määrittää myös kvanttoriadverbi. Intensiteettimääritteinä toimivat intensiteettipartikkeleiden – jotka ovat pisimmälle partikkelistuneet intensiteettisanat (*aika, melko, aivan, erittäin, ihan, kovin, oikein* ja *sangen*) – lisäksi myös nominien (adjektiivien tai voimasanojen) genetiivimuodot (*sairaana, turhana*) ja joukko ”muita” sanoja (*tosi, liika, tarpeeksi*). (VISK: määritelmät, § 664, § 853)

VISKissä todetaan, että intensiteettipartikkelit ovat adverbien ja partikkelien välimaastossa (VISK: § 664), mutta ne luokitellaan syntaktisin perustein partikkeleiksi. Syynä on se, että intensiteettipartikkelit eivät saa määritteitä. (VISK: § 646) Genetiivimuotoiset nominit eivät ole adverbejä eivätkä partikkeleita, mutta toimivat

intensiteettipartikkeleiden tavoin intensiteettimääritteinä. Samoin toimivat luokassa ”muut” olevat sanat. (VISK: § 664)

Jantusen ja VISKin välillä on eroa astemääritteiden luokittelyssä sanaluokkana, mutta osin monitulkintaista on myös käsitteiden käyttäminen. Jantusen (2015) mukaan astemääritteet ovat toisin sanoen intensiteettiadverbejä (ks. mts. 110). VISKin mukaan intensiteettiadverbi-käsite limittyy intensiteettipartikkeli-käsitteeseen (VISK: § 853) eli vain yhteen VISK:in luokkaan kolmesta intensiteettimääritteinä käytettävien sanojen luokasta.

Viimeisen sadan vuoden aikana on tarjottu runsaasti erilaisia näkemyksiä intensiteettiä ilmaiseviin sanoihin. Esimerkiksi Erik Ahlman (1933) jakoi adverbit semanttis-syntaktisperusteisiin kolmeen luokkaan: *materiaaliset adverbit* (johon kuuluvat myös kvanttoriadverbit; ks Ahlman 1933: 147), *intensiteettiadverbit* (se ryhmä vastaa suuressa määrin VISKin intensiteettimääritteitä) (ks. VISK: § 666) ja *modaaliset adverbit*. Ahlman totesi, että sanojen määrittely ei ole aina yksiselitteistä ja esiintyy rajatapauksia. Esimerkiksi *kovin*, *erittäin*, *ihan*, *sangen* ovat hänen mukaansa perinteisesti intensiteettiadverbejä, mutta esiintyvät muidenkin luokkien tehtävissä, samalla kuin modaalinen adverbi *aivan* ja materiaalin adverbi *oikein* esiintyvät myös intensiteettiadverbien tehtävissä. (Mts. 143–144, 157)

Ahlmanin käsittelyä pidettiin pitkään suomen kielen adverbien luokittelun perusluontoisena esityksenä (Koski 1988: 407). Myös VISKissä (§ 793) todetaan, että heidän luokituksensa muistuttaa Ahlmanin klassista jaottelua, mutta samalla intensiteettisanoja käsittelevissä luvuissa hyödynnetään Orpanan väitöskirjaa (ks. VISK: kirjallisuutta pykäliin § 664–§ 666). Orpana (1998) analysoi Ahlmanin luokittelua väitöskirjassaan (ks. esim. mts. 62–74), mutta tuli päätökseen, että Ahlmanin kriteerit soveltuvat yksiselitteisesti ja ristiriidattomasti vain pieneen adverbien joukkoon (mts. 69).

Lisäksi aihetta ovat pohtineet ja erilaisia näkökulmia tarjonneet Auli Hakulinen ja Fred Karlsson (1988: 83–84; 137), Maria Vilkuna (1996: 44, 207–208) ja monet muutkin (ks. myös Räsänen 2010: 58). Tuoreimpien käsitysten, kuten VISKin ja Jantusen (esim. 2015) perusteella nähdään kuitenkin, että täyttä yksinmielisyyttä ei ole vielä löydetty.

Osittain ristiriitaisten käsittelytapojen perusteella voi todeta, että kyseisiä sanoja ei voi yksiselitteisesti määritellä. Tämän tutkimuksen kannalta ei ole ratkaisevaa, onko kyse adverbista vai partikkelista ja millainen on astemääritteiden luokan sisäinen luokittelu. Siksi sitä ei pohditakaan tarkemmin, vaan lähdetään niiden syntaktisesta roolista määrittää sekä adjektiiveja, adverbeja, kvanttoireita että adpositioita/adpositiolausekkeita ja käytetään esimerkiksi Jantusen (2004, 2009, 2015) ja Tuomas Huumon (2019) tavoin kattotermiä *astemäärite*, jota pidetään tässä tutkimuksessa yksiselitteisimpänä.

## 2. FRASEOLOGIAA

Astemääritteitä käsitellään tässä tutkimuksessa fraseologisesta näkökulmasta. Termi *fraseologia* periytyy kreikan kielestä ja merkitsee kirjaimellisesti oppia puheen käännteistä (phrasis = käänne, puhetapa; logos = oppi) (Kuusi 1971: 10).

Fraseologiaa on aikoinaan käsitelty leksikologian alaluokkana (ks. myös Kuusi 1971: 13–14), joka käsittelee yksittäisten sanojen sijaan sanojen myötäesiintymäkokonaisuuksia. (Granger 2005: 1). Näillä monisanaisilla kokonaisuuksilla on yhtymäkohtia erilaisiin kielitieteen aloihin kuten morfologiaan, syntaksiin ja semantiikkaan (Granger 2005: 1), mutta lisäksi myös kirjallisuuteen ja perinnetieteeseen (Kuusi 1971:5). Tiivis sidonnaisuus erilaisiin kielitieteen aloihin saattaa Sylviane Grangerin (2005) mukaan olla syy, minkä takia fraseologia on saanut oman vakiintuneen positionsa kielitieteessä erillisenä tieteenhaarana vasta hiljattain. Lisäksi tutkijat eivät ole aina yksimielisiä siitä, millaisia monisanaisia yksiköitä ottaa fraseologiseen analyysiin. (Mts. 1)

Anna-Leena Kuusi (1971) on aikoinaan sanonut, että fraasiluokitteluja on lähes yhtä monta kuin on luokittelijaa. Luokitusten erot johtuvat siitä, millaista kriteeriä pidetään jaotteluissa silmällä: ovatko ne kielellisiä, tyyllisiä, rakenteellisia, merkitysopillisia, sisällöllisiä, kontekstuaalisia tai muita ominaisuuksia. (Mts. 18–19) Kuusi viittaa fraasiaineksen monimutkaisuuteen ja yhtenäisen systematiikan puuttumiseen sanalla *fraasiviidakko* (mts. 5–6). Tästä hauska vertauskuvasta on inspiroitunut myös tämän tutkimuksen otsikko.

Fraseologiaan alettiin kiinnittämään enemmän huomiota vasta 1900-luvun keskivaiheilla (A. Öim, K. Öim 2019: 6) ja vielä 1970-luvulla tutkittiin sitä maailman eri alueilla aika epätasaisesti (Kuusi 1971: 5). Suomessa fraseologia on Kuusen (1971: 10) mukaan jäänyt kielitieteen, tyyliopin ja perinnetieteen väliseen tyhjiöön, koska fraseologista tutkimusta

omana tieteenhaaranaan ei ollut. Slaavilaisten kielten alueella sekä anglosaksisissa maissa on ollut kiinnostus fraseologiaan ja sen tutkimukseen keskimääräistä mittavampaa (Kuusi 1971: 5). Entisessä Neuvostoliitossa ja itäeurooppalaisissa maissa se sai oman asemansa muuta maailmaa aiemmin, Länsi-Euroopassa ja Yhdysvalloissa vasta viimeisten 40 vuoden aikana (Cowie 2001: 1).

Fraseologian tutkimuksessa on Anthony Paul Cowien (2001) mukaan kolme peruslinjaa. Ensimmäinen on ns. klassinen venäläinen traditio (ks. tarkemmin mts. 4–8), jolla on vanhimpana linjana ja sitä mukaa tietynlaisena perustana laaja-alainen vaikutus muihin koulukuntiin, koska muut käyttävät tavalla tai toisella siinä luotua klassifikaatiota (ks. Howarth 2001: 164). Venäläisen lähestymistavan edustajia ovat esimerkiksi Igor Mel’čuk ja Victor Vladimirovich Vinogradov. (Mts. 2–4)

Toinen peruslinja on suurelta antropologinen (ks. tarkemmin Cowie 2001: 8–9), jonka mukaan fraseologia antaa erityisen hyvän pohjan kielellis-kulttuurillista (*linguo-cultural*) analyysia varten. Kulttuurisidonnaisuutta pidetään oleellisena, jos tavoitteena on fraseologisten ilmiöiden aukoton kuvaileminen ja tulkitseminen. Cowien mukaan antropologinen linja on suurelta venäläisen tradition laajennusta kulttuuridimension suuntaan. Tätä linjaa edustaa esimerkiksi Veronika Teliya. (Cowie 2001: 2, 8)

Kolmas on 1950-luvulla John Rupert Firthin alkuunpanema ja 1960-luvulla Michael Alexander Kirkwood Hallidayn laajentama, tälläkin hetkellä produktiivinen läheistymistapa (Cowie 2001: 3; Carter 2004: 2), jota nimitetään brittiläiseksi tai firthiläis-sinclairilaiseksi koulukunnaksi. Tätä linjaa edustavat esimerkiksi John McHardy Sinclair, Elena Tognini-Bonelli ja Michael Stubbs. (Jantunen 2009: 360) Brittiläisen koulukunnan ideat ovat nykyisin näkyvillä korpuksiin pohjautuvissa tutkimuksissa (Cowie 2001: 3), muun muassa Grangerin alulle panemassa korpuksiin pohjautuvassa oppijankielen tutkimuksessa (Jantunen 2009: 360).

Brittiläisessä koulukunnassa lähestytään sanojen merkityksiä toistuvan kontekstuaalisen käytön kautta (Tognini-Bonelli 2001: 165). Toistuvat ja suhteellisen vakiintuneet myötäesiintymäkokonaisuudet otetaan erilaisten fraasien, sanontojen, idiomien ja keskustelurutiinien lisäksi fraseologiseen tarkasteluun (ks. Jantunen 2009: 357).

Seuraavassa alaluvussa annetaan yleiskuva (korpuslingvistikäsitteistä) kontekstuaalisesta merkityksen tutkimuksesta ja sen perusideoista.

## 2.1. Brittiläisestä näkökulmasta korpuslingvistiikkaan

Brittiläisen koulukunnan tutkimuskohteena on merkityksen kontekstuaalinen luonne (Stubbs 1996: 25; viitattu Jantunen 2004: 7 perusteella). Huomioon otetaan ilmausta ympäröivä lähempi tekstikonteksti tai koko sosiaalinen ympäristö ja ilmausten usuaaliset merkitykset kuvataan toistuvien käyttöyhteyksien perusteella. Toistuvuutta nähdään sanojen myötäesiintymiä ohjaavien sääntöjen tai preferenssien tuloksena ja näin nämä säännöt tai preferenssit ovatkin tutkimuksen keskiössä (Jantunen 2004: 7–8). Lähestymistavan kulmakivenä on todelliseen ja luonnolliseen kielenkäyttöön perustuva analyysi (Stubbs 1993: 1).

Niin kuin edellisessä luvussa todettiin, lähestymistavan alkuunpanija oli Firth. Hänen mukaansa ympäristö vaikuttaa ihmiseen niin suuressa määrin, että ihminen ja ympäristö muodostavat yhtenäisen, erottamattoman kokonaisuuden. Tämä yhtenäisyys heijastuu myös kielenkäytössä, esimerkiksi niin, että ihminen kiinnittää huomiota ympäröiviin esineisiin ja tapahtumiin. Sen takia firthiläinen kielen kontekstuaalinen merkitys ottaa kielensisäisten (syntagmaattisten ja paradigmaattisten) suhteiden lisäksi huomioon kielenulkoiset suhteet, jotka muodostavat yhdessä *tilannekontekstin* (*context of situation*). (Firth 1962: 4–5, 7)

Firthin tilannekonteksti on abstrakti skemaattinen kokonaisuus, jonka osia ovat kielenkäyttötilanteeseen osallistujat (verbaalisine ja nonverbaalisine toimintoineen), kielenkäyttötilanteeseen vaikuttavat oleelliset objektit, tapahtumat ja kielellisen toiminnan vaikutus. Firthin mukaan tilannekontekstille on ominaista toistuvuus. (Mts. 5, 7–9) Jantunen (2004: 11) tarkentaa, että yksikään kielenkäyttötilanne ei voi toistua kahta kertaa identtisenä ja sen takia toistuvuuden vaatimukseen pitää sisältyä likimääräisyyden hyväksyminen.

Firthin käsitys kontekstuaalisesta merkitysteoriasta on kantautunut nykypäivään asti. Vaikka se nähdään liian laajana ja korpuslingvistinen kielentutkimus lähestyy merkitystä suppeamman kontekstin avulla, Firthin käsitteet kuten *kollokaatio* ja *kolligaatio* (esim. Firth 1962) ovat yhä mukana. (Jantunen 2004: 11) Firthin kontekstuaalisen merkityksen teoria keskittyi teksteihin, ei korpuksiin (Tognini-Bonelli 2001: 4). Hän loi teoriansa 1950-luvulla, ennen kuin korpuksia alettiin käyttää tutkimuksen apuna (mts. 157).

Ensimmäisiä elektronisia korpuksia alettiin koostamaan 1960-luvun alkupuolella (Kennedy 2010: 23) ja oppijankielen korpuksia 1990-luvun alussa (Jantunen 2009: 360). Korpuslingvistiikan kehittämisessä keskeinen rooli on ollut Sinclairilla (Cowie 2001: 3). Silloin kun hän alkoi käyttämään korpuksia, ne olivat tämän päivän näkökulmasta katsoen pieniä ja niiden käyttäminen oli monimutkaista. Alussa niitä käytti vain pieni määrä tutkijoita. (Johansson 2011: 17–18) Nyt korpusten myötä on sekä oppijankielen että laajemmin kielen tutkimuksessa meneillään fraseologinen buumi (Jantunen 2009: 360). Myös kollokaation eli leksikaalisen myötäesiintymisen tutkiminen pääsi vauhtiin vasta korpusten laajenemisen ansioista. Ennen 80- ja 90-luvua ilmestyi vain muutama niihin keskittyvä tutkimus. (Jantunen 2004: 15)

Korpusten käyttöönottoa kielitieteessä on vertailtu teleskoopin käyttöönoton vaikutukseen astronomiassa (Ranalli 2003: 2). Ennen korpusten käyttöä kielitiede perustui suuressa määrin natiivien kielenkäyttäjien intuitioon ja introspektioon, kuvaillen sitä, mitä kielestä tiedettiin tai mitä sen uskottiin olevan (Tsui 2004: 39). Intuition perusteella ei ole mahdollista arvioida sanojen ja erilaisten kielessä löytyvien rakenteiden todellista esiintymistajuutta (Kitsnik 2006: 94). Lisäksi ihmiset voivat tajuta erilaisten sanojen yhteisesiintymisen mahdollisuudet eri tavoin, vaikka kyse on heidän omasta äidinkielestään (Nesselhauf 2003: 228; ks. myös Vanhatalo 2003: 356). Korpuksissa esitettyä kieltä pidetään autenttisenä ja luonnollisena (Tognini-Bonelli 2001: 2). Eli niiden perusteella on mahdollista kielen todellisen käytön tutkiminen, joka on yksi brittiläisen koulukunnan korostama keskeinen seikka (Jantunen 2009: 360).

Tärkein korpusten perusteella tehty havainto Mare Kitsnikin (2006: 94–95) mukaan on se, että eri ihmisten kielenkäyttö ei olekaan kovin omaleimaista, koska tekstit koostuvat

yksittäisten sanojen sijaan pikemmin 2–7 sanan pituisista yksiköistä. Britt Ermanin ja Beatrice Warrenin (2000: 38) mukaan sellaisen yksikön pituus puhutussa englannin kielessä on keskimäärin ~2,6 sanaa ja kirjoitetussa 2,8 sanaa. Eli kun käytetään Göran Kjellmerin (1991: 124) metaforaa, kieli koostuu yksittäisten tiilien (*individual bricks*) sijaan pikemmin valmiista elementeistä (*prefabricated sections*).

Tämän uudenlaisen näkökulman tarjosi korpuksiin perustuvan Collins Cobuild English Language Dictionary -sanakirjan (COBUILD) koostaminen (ks. Sinclair 1987a), joka käänsi koko tähänastisen fraseologia-alaisen käsityksen ylösalaisin (Granger 2005: 2). COBUILDin koostamisen aikana huomattiin korpusaineistosta, että kielessä on suuri määrä toistuvia rakenteita. Ne näkyivät jopa arkisimpien sanojen (esim. *see* 'nähdä' ja *give* 'antaa') käytössä. Marginaalisen käytön sijaan ne olivat kielessä hallitsevia. (Carter 2004: 3; Sinclair 1987a: vii)

COBUILD-projektia ohjannut (ks. Collins Dictionary) Sinclair (1987b) tuo esimerkin fraasiverbin *set off* esiintymisestä Cobuild Main Corpus -korpuksessa, jossa on yhteensä ~7,3 miljoonaa sanaa. Sinclair sanoo, että jos sanat esiintyisivät kielessä satunnaisesti, tämän kokoisessa korpuksessa voisi olla vain yksi *set off*-esiintymä, mutta sen sijaan niitä löytyy 70. (Mts. 152) Sellaisten havaintojen valossa ilmauksia alettiin luokittelemaan aineistosta nousevasti (*bottom-up approach*), eikä enää ennalta formuloituihin luokkiin (*top-down approach*). Näin fraseologian tutkimuskohteeseen nousivat sellaisetkin yhteisesiintymät, joita aiemmin ei pidetty fraaseina. (Granger 2005: 2)

Sinclairin (1991: 110) mukaan sanat eivät esiinny tekstissä satunnaisesti eli tekstejä ei luoda vapaan valinnan perusteella (*open-choice principle*), vaan sanojen käyttöä ohjaavat idiomaattiset esiintymisrajoitukset ja -preferenssit (*idiom principle*). Hänen mukaansa sekä täysin vapaa valinta ja että erittäin tiukat rajoitukset ovat harvinaisia ja ominaisia vain yksittäisille sanoille. Suurimmassa osassa rakenteista on mahdollista jossain määrin variaatiota. (Sinclair 1987b: 158; Sinclair 2004a: 29–30). Silloin sanan käyttöä ohjaavat tiukkoja rajoituksia (*mahdollinen valinta*) lievemmat *myötäesiintymäpreferenssit* (*tyypillinen valinta*) (Jantunen 2004: 15). Jantunen (2009: 360) viittaa näin muodostuneisiin rakenteisiin käsitteillä *elementti-* tai *valmisrakenne*.

Sinclairin (esim. 2004a: 29–30) mukaan fraseologiset suhteet ovat kielessä keskeisiä ja ohjaavat leksikaalisten yksiköiden käyttöä. Joskus tätä lähestymistapaa pidetään liian radikaalina ja pikemmin vältetään tämänkaltaista liiallista yleistystä (Granger 2005: 2). Esimerkiksi Ermanin ja Warrenin (2000: 50) mukaan tekstistä (vain) hieman yli puolet muodostetaan idiomaattisen valinnan perusteella. Idiomaattista valintaa noudetaan puhutussa kielessä (~59 %) hieman enemmän kuin kirjoitetussa kielessä (~52 %) (mts. 37). Jantunen (2009: 373) puolestaan arvioi, että erittäin taajakäyttöisiä kieliopillisia sanoja ei voi esittää tarkkarajaisina fraseologisina yksikköinä (esim. konjunktioita).

Lisäksi korpuksia hyödyntävät tutkimukset ovat kyseenalaistaneet perinteisen näkökulman kielen jakautumisesta sanastoon ja kielioppiin, johon kielenopetus on pitkään perustunut, ja näyttävät, että ne kaksi ovat tiiviisti sidoksissa toisiinsa (Kitsnik 2006: 94; Ranalli 2003: 2). Korpusten perusteella huomattiin, että vaikka saman lemmän eri muotojen merkityksissä olisikin osittaista limittymistä (esim. *face* lemmasta muodustetut *faced* ja *facing*), ne voi yhdistää tyypilliseen käyttöympäristöön ja konkreettisiin leksikaalisiin, semanttisiin ja kieliopillisiin valintoihin. Sen mukaan voi väittää, että kielioppi ja merkitys ovat saman ilmiön (kielenkäytön) kaksi eri aspektia. (Tognini-Bonelli 2001: 92–96, 99 )

Vaikka korpuksia ovat tarjonneet uusia näkökulmia sekä kielen rakenteisiin että käyttöön (Tsui 2004: 40), niistä juontunut uudenlainen näkökulma jäi Sinclairin mukaan huomiotta melkein neljännesvuosisadaksi ja on saanut enemmän huomiota vasta kuluvalla vuosikymmenellä. (Sinclair 2004b: 1; ks myös Ellis 2008: 3–4) Seuraavassa alaluvussa esitellään, millaisina syntagmaattisina suhteina idiomaattinen valinta ilmenee.

## 2.2. Syntagmaattisia suhteita

Sanojen käyttöön vaikuttavat monenlaiset seikat: esimerkiksi paradigmaattiset (kuten *synonymia*, *meronymia*, *hyponymia*) ja syntagmaattiset (kuten *kollokaatio*, *klusteri*, *kolligaatio*, *semanttinen preferenssi* ja *semanttinen prosodia*) suhteet, kieliopilliset rajoitukset, tyyllilaji-, rekisteri- ja tekstilajipiirteet ja kielenulkoisen todellisuuden (Pajusalu 2009: 131, 158–159; Jantunen 2009: 358–360).

Firth (1962: 11) on sanonut: ”You shall know a word by the company it keeps!” Seuraavaksi annetaan yleiskuva sanojen käyttöä ohjaavista syntagmaattisista suhteista, jotka realisoituvat joko väljempinä valintapreferensseinä tai tiukkoina lähitekstiympäristöön suuntautuvina valintarajoituksina (Jantunen 2009: 358). Syntagmaattisten suhteiden perusteella nähdään, mitkä lekseemit esiintyvät tyypillisesti yhdessä ja mitkä eivät (vrt. idiomaattinen valinta). Konteksteittain saattaa näkyä eroja (esim. yllämainitut rekisteri ja tekstilaji).

Kontekstia koskevia käsitteitä käytetään eri tavoin. Tässä tutkimuksessa hyödynnetään suomen kielen astemääritteitä tutkineen Jantusen (2004) määrittelyä, jonka mukaan *koteksti* on tekstuaalinen lähiympäristö, laajempaan tekstikontekstiin tai tekstinulkoiseen maailmaan viittaa *tilannekonteksti*. Kotekstin ja tilannekontekstin kattokäsitteenä toimii *konteksti*, johon sisältyvät molempien sisältämät piirteet. (Mts. 12).

### 2.2.1. Kollokaatio

Kollokaatio on kahden sanan leksikaalinen myötäesiintymä (Jantunen 2009: 358). Se voi olla arbitraari, mutta saattaa myös johtua kielenulkoisesta todellisuudesta (mts. 357). Kielenulkoisessa todellisuudessa *lihan paistaminen on jäätelön paistamista* yleisempi ja näin on se myös kielessä (Pajusalu 2009: 159; ks. myös Tognini Bonelli 2001: 87–88). Myös onomatopoeetiikka voi vaikuttaa sanojen yhteisesiintymisiin, esimerkiksi *kana kotkottaa*, mutta *kukko kiekuu* (Jantunen 2009: 358).

Kollokaatio on toistuva sanojen yhteisesiintymä. Yleensä taajemmin esiintyvää kollokaattia pidetään analyysin kannalta merkitsevämpänä. (Jantunen 2004: 16, 17). Kollokaation merkittävyyttä voi tarkastella tilastollisten testien avulla (esim. MI-testi, ks. Kennedy 2003: 472), joiden perusteella nähdään, esiintykö sana kontekstissa enemmän kun on satunnaisesti mahdollista. (Jantunen 2004: 17)

Yhdessä tekstissä eniten esiintyvät kollokaatit eivät ole välttämättä hyväksyttäviä kaikissa tilannekonteksteissa, koska kollokationaalinen valinta on sidoksissa sekä rekisteriin, tekstilajiin että tyyliin. Vaikka kollokationaalisen valinnan hallitsemista pidetään kielitaidon osana, natiivitkaan kielenkäyttäjät eivät ole sopivuudesta aina yksimielisiä. (Partington 1998: 17–19)

Toisen sanan kanssa kollokaationaalisessa suhteessa oleva sana on *kollokaatti*. Ennen tarkasteltavaa sanaa esiintyy *vasemmanpuoleinen kollokaatti* (V1 tai 1L, joka on ensimmäinen sana vasemmalla, V2 tai 2L toinen jne.) ja tarkasteltavan sanan jälkeen *oikeanpuoleinen kollokaatti* (O1 tai 1R, toinen sana oikealla O2 tai 2R jne.). (Pajusalu 2009: 159; ks. Jantunen 2001; Jantunen 2009) Sanat varioivat kombinatorisilta suhteiltaan. Löytyy sanoja, jotka sallivat vain muutaman harvan kollokaatin (*bitterly cold*) ja niitä, jotka liittyvät toisiinsa suhteellisen vapaasti (*completely different/new/free* ym.). (Granger 2001: 147; ks. myös Partington 1998: 25–26)

Vaikka yhteisesiintyminen nähdään tietynlaisena asteikkona, jonka yhdessä päässä sanojen valintaa ohjaavat preferenssit ja toisessa päässä säännöt, tutkijat ovat eri mieltä siitä, pitäisikö sellaisia yhteisesiintymiä erotella toisistaan. Esimerkiksi Nadja Nesselhauf (2003: 226) luokittelee yhteisesiintymät vapaisiin yhteisesiintymiin, kollokaatioihin ja idiomeihin (esim. Nesselhauf: 2003: 226; ks myös Howarth 2001: 164), muutamat tutkijat taas näkevät ne saman ilmiön (fraseologisen kielenkäytön) ääripäinä (Ellis 2008: 5). Grangerin (2005: 2) mukaan yhteisesiintymiä jaotellaan pikemminkin klassisessa koulukunnassa, samalla kun Sinclairin kannustajat eivät kiinnitä siihen kovin paljon huomiota. Tässä työssä yhteisesiintymistä käytetään yhteistä nimitystä *kollokaatio* eikä pyritä erottelemaan niitä toisistaan.

Tutkijoidenväliset erot tulevat esiin siinäkin, kuinka monen sanan yhteisesiintymää pidetään kollokaationa (vain kahden vai useamman) (ks. esim. Jantunen 2004: 15–21; Partington 1998: 15–17). Tässä tutkimuksessa kollokationaaliseen tarkasteluun otetaan Jantusen (esim. 2009) tavoin vain kaksisanaiset yhteisesiintymät, joka on myös enemmän levinnyt tapa (Partington 1998: 16). Kolmen ja useamman sanan yhdistelmät lasketaan klusteriksi (tätä käsitellään tarkemmin alaluvussa 2.2.2.).

Kun useampisanaisissa kombinaatioissa katsotaan yleensä vain peräkkäisiä sanoja, kaksisanaisissa kombinaatioissa sanojen välillä voi olla pieni välimatka (Jantunen 2004: 18). Välimatkan pituus määrätään tarkastelualueen rajoittamisella ja se varioi tutkijoittain ja jopa tutkimuksittain (ks. tarkemmin Jantunen 2004: 19, 21). Esimerkiksi Jantunen tarkastelee väitöskirjassaan astemääritteiden kollokaatioita V3–O3-positioista (2004), mutta myöhemmin on tyytynyt pelkästään vieruskollokaatteihin eli V1- ja O1-positioihin. Kalliorannan (2009) tarkastelualue *paljon*-sanan kollokaateista on V2–O2 eli kaksi sanaa *noodin* molemmilla puolilla. Noodilla tarkoitetaan tarkasteltavana olevaa lekseemiä eli hakusanaa (Jantunen 2004: 15).

Kollokaattien merkitysisällöstä ilmenevät sanojen semanttiset ominaisuudet kuten semanttinen preferenssi ja semanttinen prosodia (Jantunen 2004: 24). Niistä tarkemmin luvuissa 2.2.4 ja 2.2.5.

### 2.2.2. Klusteri

Klusteri eli n-grammi on kolmen tai useamman peräkkäisen sanan muodostama toistuva sanaketju (Jantunen 2009: 359). Esimerkiksi Jantunen (2009) Natiivisuomen korpuksesta kerätyssä aineistossa 28 kertaa esiintyvä *aivan varma että* on 3-grammi, mutta 7 kertaa esiintyvä *oli aivan varma että* 4-grammi (ks. mts. 368, 380).

Klusterit eroavat kollokaatioista sekä rakenteeltaan että merkitsevyydeltään. Kollokaatioissa sanojen välillä voi olla pieni välimatka, mutta klusterissa katsotaan vain peräkkäisiä sanoja (Jantunen 2004: 18–19). Lisäksi Kjellmer (1984: 167; viitattu Jantunen 2004: 18 perusteella) on huomauttanut, että toistuvan sanaketjun pitenemisen

myötä laskee sen satunnaisen muodostumisen todennäköisyys. Sitä mukaa pitemmän sanaketjun merkitsevyys nousee.

### 2.2.3. Kolligaatio

Lekseemit voivat esiintyä vakiintuneiden leksikaalisten rakenteiden lisäksi myös vakiintuneissa kieliopillisissa rakenteissa (Jantunen 2004: 21). Sellaista leksikaalis-kieliopillista myötäesiintymää nimitetään kolligaatioksi (Jantunen 2009: 359). Kieliopilliset kategoriat, esimerkiksi sanaluokat ja lauseenjäsenluokat esitetään kolligaatteina (Jantunen 2004: 22).

Sinclair (2004a) tuo esimerkin *naked eye* -kollokaatiosta (yhteensä 151 esiintymää 211 miljoonan sanan kokoisesta korpuksesta), jonka vasemmanpuoleisena vieruskolligaattina (V1) on melkein aina artikkeli *the* (95 %). Sitä ennen (V2) prepositio (yli 90 %) ja V3-positiossa verbi tai adjektiivi. Eli tyypillisin rakenne on *verbi/adjektiivi + prepositio + the + naked eye*. (Mts. 31–32) Useilla lekseemeillä onkin taipumus esiintyä vain tietyn kielioppikategorian (esim. sanaluokan, lauseenjäsenen) yhteydessä (Jantunen 2001: 174).

Lisäksi kolligationaalisen suhteen osana voi pitää sanan syntagmaattista morfologista preferenssiä (Jantunen 2009: 359). Se tarkoittaa, että sanan erilaisilla taivutusmuodoilla voi olla oma kontekstuaalinen valintansa (Jantunen 2004: 27) ja kielioppikategorioihin liittyvät piirteet voivat löytyä sekä kolligaateista että hakusanasta. Esimerkiksi *kauniiksi*-sanon kollokaatti on usein puhumista tai muuttumista ilmaiseva verbi (kuten *sanoa*, *remontoida*), mutta *kauniilta*-sanon kollokaatit joltakin näyttämistä ilmaisevat verbit (kuten *näyttää*, *vaikuttaa*) (Jantunen 2001: 174–175)

### 2.2.4. Semanttinen preferenssi

Semanttinen preferenssi on lekseemin ja kontekstissa esiintyvän merkityspiirteiden syntagmaattinen suhde eli leksikaalis-semanttinen myötäesiintymä (Jantunen 2009: 359). Semanttista preferenssiä tarkastellaan noodin ja sen kollokaattien merkityssisältöjen perusteella. Jos sanan kollokaatit ja kolligaatit ovat helposti aineistosta poimittavia, semanttisen preferenssin tarkastelussa myös tutkijan intuitiolla on roolinsa, koska hän joutuu luokittelemaan kollokaattien merkityssisältöjä (Jantunen 2004: 24).

Sinclair (2004a) tuo esimerkin *naked eye* -kollokaation semanttisesta preferenssistä. Se ilmenee vasemmanpuoleisesta kotekstista, yleensä V3-positiosta tai sen läheltä. Semanttinen preferenssi on näkyvyyteen liittyvillä sanoilla (joista suuri osa on lekseemien *see* ja *visible* variaatioita), esimerkiksi *you can see with the naked eye* ja *just visible to the naked eye*. (Mts. 32–33) Kuten tästä Sinclairin esimerkistäkin ilmenee, merkityspiirteen on oltava säännönmukaisesti kotekstissa, mutta se voi ilmetä eri sanaluokkien sanoissa (verbi *see* vrt. adjektiivi *visible*) ja vaihtelevissa positioissa (Jantunen 2009: 359). Sanoilla voi olla myös useita preferenssejä (Jantunen 2004: 25).

Amy B. M. Tsui (2004) puolestaan erottelee semanttisen preferenssin perusteella synonyymisia adjektiiveja *high* ('korkea') ja *tall* ('pitkä'). Hänen mukaansa *high*-sanan kanssa käytetään pikemmin abstraktimpia merkityksiä kantavia substantiiveja (kuten *inflation rate* 'inflaatio', *growth potential* 'kasvupotentiaali'), mutta *tall*-adjektiivin kanssa konkreettisempia (kuten ihmiset, puut, rakennukset). (Mts. 45–46)

### 2.2.5. Semanttinen prosodia

Semanttisen preferenssin lisäksi leksikaalis-semanttista myötäesiintymää on myös vielä abstraktimpi semanttinen prosodia eli ns. merkitysaura (Jantunen 2009: 359; Louw 1993: 157). Semanttinen prosodia voi olla joko positiivinen, negatiivinen tai neutraali. (Jantunen 2009: 359)

Esimerkiksi ominaisuuden suurta astetta ilmaisevalla *heavily*-sanalla on pikemmin negatiivinen semanttinen prosodia, koska sitä käytetään enemmän negatiivisissa yhteyksissä kuten *heavily taxed* ('verotettu'), *polluted* ('saastunut'), *biased* ('ennakkoluuloinen') ym. Samalla suurta astetta ilmaisevalla *very*-sanalla on positiivinen semanttinen prosodia, koska se esiintyy pikemmin positiivisessa yhteydessä kuten *very fond* ('kiintynyt', 'hellä'), *nice* ('mukava'), *tasty* ('herkullinen') ym. (ks. Kennedy 2003: 477, 480).

### **3. KIELENOPPIJA FRASEOLOGISESSA VIIDAKOSSA**

Edellisestä luvusta tuli ilmi, että sanojen käyttöä ohjaavat lukuisat syntagmaattiset suhteet. Natiivi oppii nämä säännöt tai preferenssit kielenomaksumisen aikana (Jantunen 2009: 356). Pyrkiminen mahdollisimman lähelle äidinkielen kielenkäyttäjän kielitaitoa saattaa olla myös kielenoppijan tavoitteena, mutta jää yleensä saavuttamattomiin (Pällin, Kaivapalu 2012: 288).

Vapaan valinnan avulla kielen tuottaminen erottaa suuren osan kielenoppijoista natiiveista. Kristi Pällin ja Annekatrin Kaivapalu (2012: 288) ovat sanoneet, että äidinkielysitys ilmenee ennen kaikkea idiomaattisena kielenkäyttönä. Se on näin sekä suullisessa että kirjoitetussa kielessä (Howarth 2001: 186). Sanaston kapeuden sijaan oppijan epätyypillisen kielenkäytön syynä pidetään epäidiomaattista kielenkäyttöä, ainakin kun kyse on edistyneestä kielenoppijasta. (Meunier 1998: 32). Alkeistason oppijoilla se ei olekaan niin suuri ongelma ilmeisesti yksipuolisemman sanavaraston takia (Kitsnik 2006: 102).

Kielenoppijan ongelmat johtuvat siitä, että häneltä puuttuu tieto kielen tyypillisyyksistä. Näin oppija joutuu kokeilemaan, mikä sana sopii mihinkin ko(n)tekstiin ja tuottaa sen mukaan kieltä vapaan valinnan periaatteen avulla (ks. Howarth 2001: 177; Granger 2001: 151). Se johtaa helposti epäkonventionaalisiin valintoihin. (Jantunen 2008: 82).

Puhutun ja kirjoitetun kielen sujuvuus on yhteydessä fraseologiseen pätevyteen ja kykyyn muodostaa kohdekielen mukaisia ilmaisuja (Jantunen, Brunni 2012: 75). Vapaan valinnan perusteella muodostetut rakenteet saattavat vaikuttaa natiiveista kummalliselle (Granger 2001: 150) ja sen seurauksena on mahdollista, että viestin sisällön sijaan kiinnitetään huomiota sen kielelliseen asuun (Howarth 2001: 185–186). Kyse on kielen

käyttäjän puutteellisesta fraseologisesta pätevyydestä (mts. 161–163), joka voi heikentää viestinnän tehokkuutta. (Ks. mts. 186).

### 3.1. Ympäristön vaikutus oppijankieleen

Kielenoppija on vaikeassa tilanteessa monesta syystä. Tässä ei pyritä antamaan tyhjentävää selitystä, mutta esitellään muutamia mahdollisia vaikuttajia.

Ensiksi, oppikirjojen kieli poikkeaa todellisesta kielenkäytöstä (ks. tarkemmin Tsui 2004: 40–41; Tanner 2008). Se saattaa olla sekä sanastoltaan että kieliopillisilta struktuureiltaan jopa niin omituista, ettei sellaista kieltä käytetäkään kielituntien ulkopuolella. Oppimateriaalit perustuvat yleensä traditioon ja oppikirjan tekijän intuitioon. (Kitsnik 2006: 95) Vastaavia havaintoja on tuotu esiin myös suomen kielen oppikirjoista, joissa Johanna Tannerin (2008: 163) mukaan kommunikatiivisten taitojen kehittämisen kustannuksella korostetaan kielioppia.

Toiseksi, sanakirjatkaan eivät tarjoa paljon apua sanojen käyttösääntöihin, koska niissä esitetään sanat yleensä kontekstistaan irrallisina, mikä on vastoin kielen olemusta. Poikkeuksena ovat kollokaatiosanakirjat, joita ei ole kovin paljon. (Jantunen 2009: 356) Suomen kielen yleisimmän sanaston kollokaateista saa tietoa ConLexis-verkkosanakirjasta (ConLexis), jossa on hakemiston mukaan 306 sanaa (katsottu 5. toukokuuta 2020).

Kolmas ongelma piileekin siinä, miten sanojen merkityksiä yleensä selitetään opiskelijoille. Yleensä sitä tehdään synonyymien kautta (Tognini-Bonelli 2001: 33–34), mutta synonyymiset sanat eroavat toisistaan kontekstuaalisilta suhteiltaan (esim. Tsui 2004: 45). Jantunen (2004) mukaan epätodennäköistä on, että kielestä löytyy aivan samankäyttöisiä sanoja. Yleensä kyse on pikemmin lähisynonyymeista. Synonyymisten sanojen oletetaan pysyvän hengissä vain niin, että ne erikoistuvat erilaisiin käyttötapoihin ja -ympäristöihin. Sanat, joilla ei ole erityistä käyttöä tai joiden käyttö on identtistä,

karsiutuvat pois kielestä. (Mt. 1; ks tarkemmin Jantunen 2004: 53–62). Sen takia synonyymiset sanat (ainakaan) yleensä eivät ole keskenään vapaasti vaihdettavissa.

Osa oppijan ongelmista johtuu myös äidinkielen vaikutuksesta. Tämä koskee Pällinin ja Kaivapalun (2012) mukaan etenkin alkuvaiheen oppijaa, joka turvautuu usein lähde- ja kohdekielen yksi-yksi-vastaavuussuhteeseen, mutta kieli on yleensä monimutkaisempaa. Siinä esiintyy yksi-moni-, moni-yksi- ja moni-moni-suhteita. (Mts. 290–291) Esimerkiksi Tsuin (2004: 45) artikkelissa kiinalaiset opiskelijat kokivat haasteelliseksi englannin *high* ('korkea') ja *tall* ('pitkä') sanojen erottelemisen, koska kiinan kielessä niitä molempia vastaa sama sana.

Kielten väliseen vastaavuussuhteeseen turvautuminen koskee myös elementti- tai valmisrakenteiden muodostamista. Äidinkielen vaikutus voi olla sekä positiivinen että negatiivinen. Grangerin (2001: 150–151, 158) mukaan ranskalaiset edistyneet englannin oppijat käyttävät stereotyyppisiä englannin astemäärite-kollokaatioita silloin, kun niillä on suora vastine ranskassa (ranskankielinen *sévèrement puni* vrt. englanninkielinen *severely punished*). Toisaalta äidinkieleen perustuminen voi johtaa myös kohdekielelle epätyypillisiin kollokaatteihin. Nesselhaufin (2003: 234–235) mukaan edistyneillä saksalaisilla englanninoppijoilla jopa puolet virheistä on heidän äidinkieltensä vaikutuksen tulosta (*Hausaufgaben machen* vrt. *\*make homework; do homework* sijaan). Äidinkielen lisäksi myös aiemmin opitut kielet voivat vaikuttaa kielenoppijaan (ks. tarkemmin Jantunen 2008: 73).

Kielenoppijaan voi vaikuttaa myös oppimiskonteksti. Yleensä oletetaan, että kohdekielisessä ympäristössä kielen oppiminen johtaa äidinkielistä ympäristöä parempiin tuloksiin, koska silloin kielenoppijalla on enemmän kontakteja natiiveihin kielenpuhujiin. Vaikka kohdekielinen ympäristö tarjoaakin enemmän mahdollisuuksia kielen käyttämiseen, oleellista on, missä määrin mahdollisuuksia hyödynnetään ja millaisia ovat opiskelijan kielelliset ja kognitiiviset taidot. (Segalowitz, Freed 2004: 174, 196). Mahdollisuuksien hyödyntämisen kannalta todennäköisesti ovat oleellisia oppijan omat henkilökohtaiset ominaisuudet, koska hyvin ujo ja hiljainen ihminen ei käytä välttämättä kohdekielisen ympäristön tarjoamia mahdollisuuksia (Palmanto 2004: 411, 412). Vaikka

oppimiskontekstin vaikutuksesta ei olla aina yksimielisiä (ks. tarkemmin Jantunen 2015: 108–110), se näyttäneen vaikuttavan ainakin Isohätälän (2014) pro gradu -tutkielmassa ja Jantunen (2015) suomenoppijoiden astemääritteiden käyttöä koskevassa tutkimuksessa (niistä tarkemmin seuraavassa luvussa 3.2.).

Puutteellisen fraseologisen kompetenssin olennaisin syy piilee ilmeisesti siinä, että usein ei olla edes tietoisia kielessä esiintyvistä fraseologisista suhteista. Kielenoppijoille ongelmallisin osa on se, mikä jää vapaan ja idiomaattisen valinnan väliin. Vaikka opettajat huomaavatkin opiskelijan kielenkäytössä ongelmia, he eivät osaa niitä välttämättä ratkaista. (Howarth 2001: 161–162, 186) Voi olettaa, että silloin, kun opettaja joutuu selittämään kielenoppijalle, miksi yksi sana sopii kontekstiin, mutta sen synonyymi puolestaan ei, nykyistä laajempi tietoisuus kielessä esiintyvistä fraseologiasta tarjoaisi opettajalle selityksissään tukea.

## 3.2. Astemääritteet oppijankielessä

Tässä alaluvussa esitellään astemääritteiden käyttöön liittyviä ongelmia oppijankielessä. Huolimatta astemääritteiden taajasta käytöstä, varsinaista niihin keskittyvää tutkimusta oppijansuomesta ei ole ja niitä on vain sivuttu muutamissa korpustutkimuksissa (Jantunen 2015: 115) Sen takia tässä otetaan esimerkkeihin myös oppijanenglannin esimerkit.

Granger (2001) tutki edistyneiden ranskalaisten oppijoiden englannin teksteissä *-ly* loppuisten suurta astetta ilmaisevien astemääritteiden kollokaatioita. Aineisto paljasti, että oppijat käyttävät astemääritteitä natiiveja vähemmän: erot näkyivät sekä frekvenssissä (oppijoilla 230, natiiveilla 313) että astemääritekirjossa (oppijoilla erilaisia 41, natiiveilla 75). Oppijat ylikäyttivät *completely-*, *totally-*sanoja ja alikäyttivät *highly-*sanaa. (Mts. 146–148)

Vaikka oppijat käyttävät astemäärite-kollokaatioita, he alikäyttävät natiiveille ominaisia ja muodostavat itse vapaan valinnan avulla epätyypillisiä kollokaatioita kuten

*dangerously threatened* ja *irretrievably different*. Natiiveille ja oppijoille yhteisiä siinä tutkimuksessa olivat vain sellaiset stereotyyppiset kollokaatiot, joilla oli suora vastine ranskan kielessä, esimerkiksi *severely punished* (*sévèrement puni*). (Granger 2001: 150–151)

Muun aineiston ohessa Granger (2001: 151) tarkasteli suurta astetta ilmaisevan *very*-sanan käyttöä. Ilmeni, että *very* toimii eräänlaisena yleismääritteenä, jonka ylikäytöllä kompensoidaan *-ly*-loppuisten astemääritteiden alikäyttöä. Samanlaiset havainnot *-ly*-loppuisten astemääritteiden alikäytöstä tulivat esiin Sylviane Grangerin ja Paul Raysonin (1998) tutkimuksessa, jossa keskityttiin ranskalaisten edistyneiden englanninoppijoiden akateemisiin teksteihin.

Pascual Pérez-Paredes ja María Belén Díez-Bedmar (2012) tutkivat esseiden perusteella nuorten, 11–16-vuotiaiden espanjalaisten englanninoppijoiden astemääritekäyttöä. Informantteja oli 499 ja esseitä sama määrä (28 152 sanaa). Aineistossa esiintyi 1375 adjektiivia, joista vain 1,2 % sai astemääritteen. Yhteensä käytettiin neljää erilaista astemääritettä (*very, too, so, really*) eli astemääritekirjo oli melko kapea. Pérez-Paredes ja Díez-Bedmar nostavat esiin, että astemääritettä aletaan enemmän hyödyntämään 8. luokalla eli taitotason kehittymisen myötä. Tutkituissa esseissä esiintyi myös astemääritteiden epäkonventionaalista käyttöä, esimerkiksi kolligaatioissa: 11,4 % kaikista astemääritteistä esiintyi ennen substantiiveja (*very shoes* 'kengät', *very clothes* 'vaatteet'). (Mts. 110–117)

Isohätälän (2014) tutkimuksessa tulee esiin oppimiskontekstin vaikutus B1–B2-taitotason oppijoiden kieleen. Ilmeni, että suomea vieraana kielenä eli suomea ulkomailla oppijat käyttävät astemääritettä suomea toisena kielenä eli suomea Suomessa oppijoita enemmän. Vieraana kielenä -oppijat käyttävät enemmän *oikein*- ja *aika*-astemääritettä, toisena kielenä -oppijat puolestaan *ihan*-astemääritettä. (Mts. 2–3, 90)

Isohätälän (2014) mukaan vieraana kielenä -oppijoiden kielenkäytössä esiintyy enemmän epäkonventionaalisia suhteita kuin toisena kielenä -oppijoiden teksteissä. Esimerkiksi *oikein* esiintyy sielläkin, jossa monen muun astemääritteen (esim. *erittäin*) käyttö olisi luentevampaa. (Mts. 71) Jantusen (2015) mukaan sanojen ylikäyttö saattaa olla tuttujien

ja turvallisten sanojen käyttämisen tulosta. Oletetaan, että niissä ei tehdä virheitä, mutta tuttujen sanojen ylikäytöllä päädytään helposti tyyllillisiin ja fraseologisiin outouksiin. (Mts. 115–116)

Myös Jantunen (2015: 116–117) vertaili suomi vieraana ja suomi toisena kielenä B1-tason oppijoiden kieltä natiivien kielenkäyttöön, mutta hänen aineistossaan näkyy verrattuna Isohäätälään (2014) muutamia eroja. Jantunen aineistossaan suomi toisena kielenä -oppijat käyttivät astemääritteitä vieraana kielenä -oppijoita lähes kaksi kertaa enemmän (vastaavasti 221 ja 102 kertaa 10 000 saneeseen) (Jantunen 2015: 120).

Toisaalta tuntuu, että ulkomailla oppineiden astemääritekirjo on laajempi ja erilaisia astemääritteitä käytetään tasapuolisemmin, mutta erityisesti suositaan kirjakielisiä astemääritteitä. Suomea Suomessa oppineet puolestaan suosivat vain muutamia harvoja ja pikemmin puhekielisiä astemääritteitä (*tosi, aika, liian*). Jantunen olettaa, että toisena kielenä -oppijoihin vaikuttaa kohdekulttuurinen ympäristö siellä tavatun puhekielen ja epävirallisen kirjoitetun kielen kanssa. (Jantunen 2015: 120–121, 125–126; ks. myös Jantunen 2008: 80)

Näiden tutkimusten perusteella voi todeta, että kielenoppijalle astemääritteiden käyttö on haasteellista ja niin kuin Jantunen (2015: 113) sanoo, astemääritteiden käyttöön liittyviä ongelmia esiintyy oppijoilla yhtä lailla riippumatta kielitaustasta ja -taidosta. Ongelmat tulevat esiin sekä epäkonventionaalisten valintojen, äidinkielen vaikutuksen että natiivien kielenkäytöstä poikkeavien frekvenssien kautta. Joidenkin astemääritteiden alikäyttö ja toisten ylikäyttö on yleistä.

## **4. ASTEMÄÄRITTEIDEN AIVAN, ERITTÄIN, IHAN, KOVIN, OIKEIN JA SANGEN KONTRASTIIVINEN ANALYYSI**

Luvusta 1 tuli ilmi, että astemääritteet ovat suurelta osin synonyymisiä sanoja (Jantunen 2004: 70–73). Synonyymisten sanojen oletetaan pysyvän hengessä vain erikoistumalla erilaisiin käyttötapoihin ja -ympäristöihin (mts. 1). Sen takia astemääritteiden käyttöä tarkastellaan tässä työssä niiden kotekstin perusteella. Natiivit valitsevat kotekstiin sopivan sanan intuitiivisesti (Jantunen 2009: 356), mutta kielenoppijoille synonyymisten sanojen (ks. Tsui 2004: 44–45) ja sitä myötä myös astemääritteiden (ks. Granger 2001) käyttö on haasteellista.

Seuraavaksi tarkastellaan astemääritteiden *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* käyttöä kielikorpusten avulla. Aluksi esitellään aineistoja (luku 4.1.1), hakuprosessissa esiin tulleita haasteita ja niiden ratkaisemista (luku 4.1.2.). Sen jälkeen siirrytään yksityiskohtaisempaan analyysiin (luku 4.2). Analyysiosassa ensiksi tarkastellaan astemääritteiden käyttöfrekvenssejä. Sen jälkeen luodaan katsaus astemääritteiden erityispiirteisiin natiiviaineistossa (4.2.1.) ja vertaillaan siihen oppijanaaineistoa (4.2.2.). Luvun lopussa (4.3) pohditaan tuloksia.

## 4.1. Aineisto

Tässä tutkimuksessa natiiviaineistona käytettiin Eduskunta-korpusta (Eduskunta) ja oppijanaaineistona Kansainvälistä oppijansuomen korpusta (ICLFI).

### 4.1.1. Eduskunta- ja ICLFI-korpuks

Eduskunta-korpus sisältää täysistuntojen transkriptioita. Nämä täysistunnot pidettiin 10.9.2008–1.7.2016. Korpuksessa on yhteensä 1 499 627 virkettä, joiden kokonaissanemäärä on 22 458 581. (Korp, katsottu maaliskuussa 2020) Täysistuntojen keskustelua on monenlaisia. Niissä esiintyy lakiesityksiä, lakialoitteita, ajankohtaiskeskustelua, budjettikeskustelua, välikysymyskeskustelua, kyselytuntien keskustelua ym. Esiintyy useita puheenvuorotyyppisiä, esimerkiksi eduskuntaryhmän linjan julkistavia ryhmäpuheenvuoroja ja esillä ollutta asiaa kommentoivia vastauspuheenvuoroja. Osa puheista on kirjoitettu etukäteen paperista luettaviksi ja osa on spontaaneja kannanottoja. (Vuotilainen 2019)

Täysistuntokeskustelut ovat hyvin rajoitettua ja valvottua institutionaalista keskustelua, joiden osallistujilla on näkyviä institutionaalisia rooleja (esim. ministeri, hallituspuolueen kansanedustaja ym.). Edustajien puhetyyli on oltava vakaa ja arvokas. Käytännössä puheenvuoroille asetettuja rajoituksia kierretään varsin taitavasti ja joskus puheita pidetään jopa teatraalisina. (Vuotilainen 2019)

Oppijankielen aineisto koottiin Kansainvälisestä oppijansuomen korpuksesta (International Corpus of Learner Finnish eli ICLFI), joka sisältää suomi vieraana kielenä -aineistoa. ICLFI:iin on kerätty suomen kielen opiskelijoiden kirjoittamalla tuottamia tekstejä enemmän kuin kahdestakymmenestä ulkomaisesta yliopistosta vuodesta 2007 lähtien. Korpuksen sisältö on monipuolinen. Sieltä löytyy kertomuksia, kuvauksia, esseitä, päiväkirjoja, arvostelua, referaatteja, mielipidekirjoituksia, vastineita, uutisia, sähköpostiviestejä, kirjeitä, satiireja ja työhakemuksia. (Jantunen ja Pirkola 2015: 89, 91–92) ICLFI:ssä on yhteensä 92 883 virkettä, joiden kokonaissanemäärä on 959 840. Virolaisten tekstejä on niistä 242 196 sanetta, joista B1-tason tekstien sanemäärä on

134 746 ja B2-tason tekstien puolestaan 49 592 sanetta (yhteensä B1-B2 tason tekstejä 184 338 sanetta). (Korp, katsottu maaliskuussa 2020)

Granger (1996: 45) varoittaa, että jos tutkimuksessa käytettävät korpukset edustavat eri tekstilajeja, tulokset eivät ole keskenään suoraan vertailtavissa ja voivat johtaa virheellisiin päätelmiin. Natiivien kielestä ICLFI-korpukselle analogista korpusta ei tietääkseni löydy. Eduskunta-korpus valittiin natiiviaiaineistoksi monesta syystä. Kitsnikin (2006: 102) mukaan ammattimaisten ja ei-ammattimaisten (esim. kielenoppijat) kirjoittajien kieltä ei pidetä keskenään vertailtavina. Sen perusteella oletan, että tämä korpus tarjoaa paremman kuvan natiivien kielenkäytöstä, kuin esimerkiksi kaunokirjallisuutta tai journalistiikkaa sisältävät korpukset. Lisäksi Eduskunta-tekstit ovat todellista kielenkäyttöä, jota pidetään kontrastiivisessa kontekstuaalisessa kielentutkimuksessa erittäin tärkeänä (Stubbs 1993: 1). Eduskunta-aineiston etuina ovat myös aineiston laajuus (enemmän kuin 22 miljoonaa sanetta) ja puhujien monipuolinen tausta (esim. suku, ikä, murretausta).

Niin kuin astemäärätteiden käyttö on tavallista kielenoppijoilla (esim. Granger 2001: 148–149, 151), on se Douglas Biberin (1988) mukaan näin myös interaktiivisissa ja henkilökohtaista sitoutumista osoittavissa teksteissä. Sellaisia ovat pääasiassa kasvokkais- ja puhelinkeskustelut, mutta myös henkilökohtaiset kirjeet, spontaanit puheet ja haastattelut. (Mts. 129–131) Oletetaan, että erilaiset, joskus jopa henkilökohtaisetkin asenteet tulevat Eduskunta-aineistossa hyvin esiin.

Vaikka aineistot eroavat aihepiireiltään ja (osittain) rekistereiltään, oletan, että niiden molempien aika heterogeeninen sisältö pienentää eroja ainakin osittain. ICLFI-tekstit ovat kirjoitettua kieltä. Eduskunta-aineistosta löytyy sekä puhuttua että kirjoitettua kieltä, koska keskustelujen lisäksi se sisältää etukäteen paperista luettaviksi kirjoitettuja puheita (Vuotilainen 2019).

#### 4.1.2. Aineiston muokkaaminen

Korpin hakutuloksissa esiintyy *aivan-*, *erittäin-*, *ihan-*, *kovin-*, *oikein-* ja *sangen-*sanojen monenlaista käyttöä. Osa esiintymistä oli tämän työn kannalta irrelevantteja. Sellaisia

olivat esimerkiksi *aivan*-sanon dialogipartikkelimaiset käytöt (ks. VISK: § 863), *oikein* ja *ihan* -sanojen esiintymiset sävyartikkeleina (ks. VISK § 821), mutta myös *oikein*-sanon esiintymät tavan adverbina (ks. VISK: § 653) ja *kova*-adjektiivin instruktiivimuodot.

ICLFI-aineiston pienyyden vuoksi oli tutkittavien sanojen ei-astemääritteisen käytön manuaalinen poistaminen ongelmattonta. Aineistossa esiintyi myös vähäinen määrä duplikaatteja. Ne poistettiin Excel-ohjelman avulla (apuna käytettiin kirjoittajien taustatietoa). Taulukossa 1 rivi ”Kaikki” kuvaileekin alkuperäistä sanojen esiintymiä ICLFI-aineistossa ja rivi ”Astemääritteet” karsimisprosessin jälkeen analyysiin jäävien esiintymien lukua.

Taulukko 1. *Aivan, erittäin, ihan, kovin, oikein ja sangen* frekvenssit ICLFI-aineistossa.

	<i>aivan</i>	<i>erittäin</i>	<i>ihan</i>	<i>kovin</i>	<i>oikein</i>	<i>sangen</i>	<b>Yhteensä</b>
<b>Kaikki</b>	17	163	99	24	300	-	603
<b>Astemääritteet</b>	17	131	85	19	277	-	529

Taulukosta näkyy, että karsimisen tuloksena tutkittavien aineisto pieneni 74 konkordanssirivin verran.

Eduskunta-aineistossa tulosten määrä oli lähes 96 000 virkettä ja siitä syystä manuaalisen karsimisen sijaan tuli etsiä muita mahdollisuuksia tutkimuksen kannalta epäolennaisen materiaalin poistamiseksi aineistosta. Relevanttien tulosten löytämiseksi lähdettiin liikkeelle astemääritteiden syntaktisesta roolista määrittää adjektiiveja, adverbeja ja adpositio(lausekke)ita. Siitä johtuen *aivan, erittäin, ihan* ja *sangen* -sanoista tehtiin jokaisesta kaksi hakua, joista ensimmäisessä määrättiin O1-positioon adjektiivi, adverbi tai adpositio. Toisessa etsinnässä jätettiin O1-positio avoimeksi ja O2-positioon määrättiin adjektiivi, adverbi tai adpositio.

Toisella haulla *oikein* ja *kovin* -tulosten joukkoon liittyi lukuisien duplikaattien lisäksi paljon sanojen ei-astemääritteistä käyttöä. Siitä syystä *oikein* ja *kovin* ensimmäinen etsintä jätettiin samaksi, mutta toisessa määrättiin O1-positioon substantiivi ja O2-positioon adpositio, koska oletettiin, että suurin osa määritteen saavista adjektiiveista ja

adverbeista on nimenomaan O1-positiossa ja nyt O2-position täsmentämisen myötä tulosjoukkoon lisääntyä ennen kaikkea postpositioita.

Kaikki tulokset ladattiin muodossa ”Virke per rivi” Excelliin ja jokaisen konkordanssirivin unikaalin koodin (utterance\_id) avulla poistettiin tuloksista duplikaatit. Näin *aivan*, *erittäin*, *kovin* ja *sangen* -tulosten joukkoon jäi vain pieni määrä epätyypillisyyksiä. Erityisesti *oikein* ja *ihan* -esiintymissä huomattiin useita lauseita, joissa sanoilla oli asteen määrittämisen sijaan muu funktio. Sellaisia lauseita ei lähdetty poistamaan manuaalisesti, vaan niistä etsittiin sellaisia toistuvia yhteisesiintymiä, joissa olisi kyse muusta kuin asteen määrittämisestä ja ne poistettiin kerralla. Jos tämän tutkimuksen kannalta irrelevantti yhteisesiintymä esiintyi vain muutaman kerran, siihen ei kiinnitetty huomiota, koska se ei ole myöhemmän analyysin kannalta oleellista (ks. luku 4.2.1)

*Oikein*-sanan ei-astemääritteisiä yhteisesiintymiä olivat esimerkiksi sellaiset, joissa *oikein*-sanan vasemmanpuoleisena vieruuskollokaattina oli *huomioida*, *kuulla*, *kuunnella*, *kuvaila*, *muistaa*, *nostaa*, *ottaa*, *sanoa*, *tuoda*, *ymmärtää*. Silloin *oikein* oli tavan adverbinä (*jos olen ymmärtänyt oikein*). Muutamissa sellaisissa esiintymissä oli *oikein*-sanan oikeanpuoleisena vieruuskollokaattina adverbi (esim. *ymmärsin oikein hyvin*), silloin esiintymä oli silti relevantti ja jätettiin aineistoon.

Lisäksi *oikein*-sanan ei-astemääritteinen käyttö tuli esiin sellaisissa kollokaattiorakenteissa kuten *oikein niin* (*kun se tehdään oikein niin sillä jopa*), *oikein nyt* (*puheenvuorosta ei oikein nyt avautunut se*), *oikein sitten* (*kumpaa nyt oikein sitten halutaan*), *oikein edes* (*en oikein edes itse ymmärrä*), *oikein myös* (*eriarvoisuuden torjuminen on paitsi inhimillisesti oikein myös talouspoliittisesti viisasta*), *oikein vielä* (*en tämän keskustelun perusteella oikein vielä ymmärrä*), *oikein enää* (*sillä vanhalla tuotesortimentilla me emme oikein enää pärjää*) ja *oikein itse* (*nyt poliisit eivät oikein itse löydä käyttötarkoitusta*).

*Ihan*-sanan ei-astemääritteisiä yhteisesiintymiä olivat esimerkiksi sellaiset, joissa *ihan*-sanan oikeanpuoleisena vieruuskollokaattina oli *että*, *itse*, *hetkinen*, *jo*, *jopa*, *juuri*, *kuka*, *kuten*, *kyllä*, *mikä*, *miten*, *myös*, *nyt*, *olla*, *vaan*, *vaikka*, *vain*, *vielä*. Lisäksi poistettiin

sellaiset *ihan näin* myötäesiintymät, joissa *näin*-sanon jäljessä ei ollut adjektiivia, adverbiä tai adpositiota. Irrelevantteja olivat myös *aivan kuten*, *aivan juuri* -kollokaatioita sisältävät lauseet. *Aivan niin* ja *aivan näin* -myötäesiintymistä poistettiin sellaiset, joissa *niin* ja *näin* -sanaa ei seurannut adjektiivi, adverbi tai adpositio.

Taulukossa 2 tuodaan ”Kaikki” rivillä esiin sanojen esiintymismäärä Korpissa ja ”Astemääritteet” rivillä se, mikä jäi etsintöjen ja karsimisten jälkeen analyysiin.

Taulukko 2. *Aivan, erittäin, ihan, kovin, oikein ja sangen* frekvenssit Eduskunta-aineistossa.

	<i>aivan</i>	<i>erittäin</i>	<i>ihan</i>	<i>kovin</i>	<i>oikein</i>	<i>sangen</i>	Yhteensä
<b>Kaikki</b>	23 382	31 543	24 652	5517	10 673	154	95 924
<b>Astemääritteet</b>	18 090	27 921	17 397	4548	1746	141	69 843

Taulukosta näkyy, että aineisto pieneni 26 081 konkordanssirivin verran. On mahdollista, että käytetyn etsimistavan seurauksena muutama sanan astemääritteinen käyttötapaus jäi aineiston ulkopuolelle. Karsimisesta huolimatta aineistossa esiintyy todennäköisesti hieman muutakin kuin astemääritteistä käyttöä, mutta aineiston laajuuden vuoksi sillä ei ole suurta vaikutusta tuloksiin.

## 4.2. Astemääritteiden käyttäminen natiivi- ja oppijanaineistossa

Astemääritteiden käyttölukuja natiivien ja kielenoppijoiden välillä ei voi kuvata pelkästään käyttölukujen perusteella. Eduskunta- ja ICLFI-aineistot ovat selvästi erikokoisia (22 458 581 vrt. 959 840 sanetta, joista vironkielisten B1-B2 tason tekstejä on 184 338 sanetta). Siitä näkyy, että Eduskunta-aineiston sanemäärä on ~121,8 kertaa ICLFI:n vironkielisten B1–B2 aineistoa suurempi. Sen takia astemääritteiden todelliset frekvenssit eivät ole keskenään vertailtavissa. Aineistojen keskeisen vertailtavuuden vuoksi astemääritteiden esiintymismäärät esitetään todellisten frekvenssien lisäksi (f)

myös suhteellisina frekvensseinä 10 000 sanaa kohti ( $f/10\ 000$ ), joka löydetään seuraavalla toimituksella: (astemääritteen todellinen frekvenssi  $\times$  10 000)  $\div$  korpuksen sanemäärä (ks. esim. Biber 1988: 14).

Taulukossa 3 vasemmalta oikealle esitetään ensiksi Eduskunta- aineiston astemääritteiden todelliset frekvenssit ( $f$ ) ja suhteelliset frekvenssit 10 000 sanaa kohti ( $f/10\ 000$ ) ja se jälkeen ICLFI-aineiston astemääritteiden todelliset frekvenssit ( $f$ ) ja suhteelliset frekvenssit 10 000 sanaa kohti ( $f/10\ 000$ ).

Taulukko 3. *Aivan, erittäin, ihan, kovin, oikein ja sangen* käyttö Eduskunta- ja ICLFI-aineistossa.

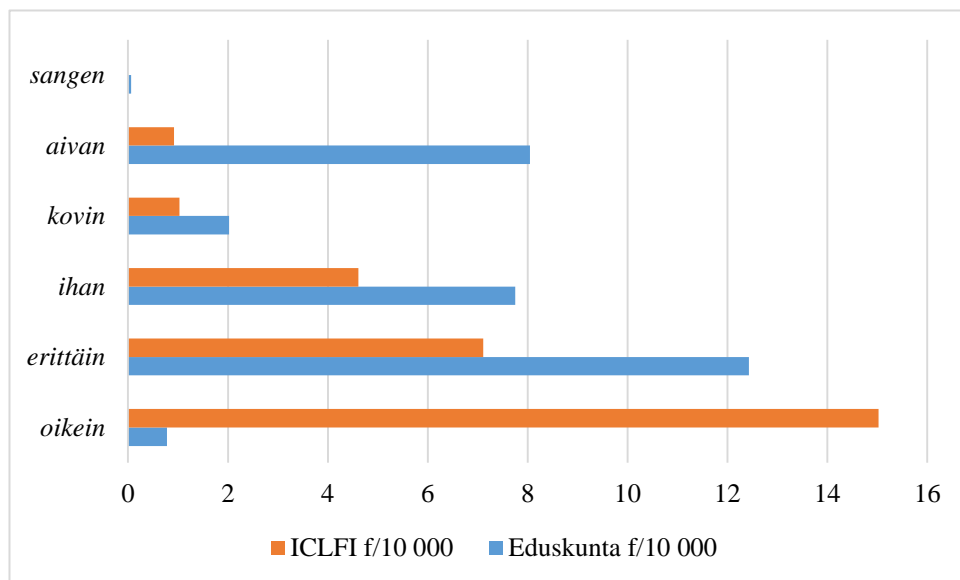
	Eduskunta		ICLFI	
	f	f/10 000	f	f/10 000
<i>aivan</i>	18 090	8,05	17	0,92
<i>erittäin</i>	27 921	12,43	131	7,11
<i>ihan</i>	17 397	7,75	85	4,61
<i>kovin</i>	4548	2,03	19	1,03
<i>oikein</i>	1746	0,78	277	15,03
<i>sangen</i>	141	0,06	-	-
<b>Yhteensä</b>	69 843	31,10	534	28,70

Taulukosta näkyy, että Eduskunta- ja ICLFI-aineistoissa astemääritekäyttö on aika tasaista. Natiiveilla se on 31,1 esiintymää 10 000 sanetta kohti ja kielenoppijoilla 28,7 sanetta 10 000 saneeseen verrattuna. Jos ottaisi lukuun muutkin (esim. *hyvin, tosi*) astemääritteet, frekvenssit saattaisivat olla tosin erilaisia.

Samanlaisista kokonaisfrekvensseistä huolimatta yksittäisten astemääritteiden käyttöluvuissa näkyy eroja. Suurimmat erot koskevat sanan *oikein* valintaa, joka on kielenoppijoiden suosituin astemäärite (~15 esiintymää 10 000 sanetta kohti), mutta natiiveilla toiseksi vähiten käytetty (~0,8 esiintymää 10 000 sanetta kohti). Käyttölukujen perusteella näkyy, että oppijankielessä sille on ominaista jopa suuri ylikäyttö.

Kaikkia muita astemääritteitä kielenoppijat alikäyttävät. Eniten alikäytettyjä ovat *aivan* (~0,9 vrt. ~8,1 esiintymää) ja *erittäin* (~7,1 vrt ~12,4 esiintymää), joista jälkimmäinen on natiivien kielessä taajimmin käytetty astemääritys. Natiivien kielessä vähiten esiintyvä on *sangen* (0,06 esiintymää 10 000 sanetta kohti), jota kielenoppijat eivät käytä ollenkaan. Edellä mainittuja tasaisemmin käytetään astemääritteitä *kovin* (~1 vrt. ~2 esiintymää) ja *ihan* (~4,6 vrt. ~7,8 esiintymää).

Astemääritteiden käytön keskinäistä vertailtavuutta varten havainnollistetaan edellisen taulukon sisältöä kuviolla 1. Pystyakselilla kuvataan astemääritys ja horisontaalisella akselilla astemääritteiden suhteelliset frekvenssit 10 000 sanaa kohti.



**Kuvio 1. *Aivan, erittäin, ihan, kovin, oikein* ja *sangen* frekvenssit 10 000 sanaa kohti Eduskunta- ja ICLFI-aineistossa.**

Jantunen (2015: 127) korostaa, että vaikka sanojen käyttö voi olla frekvensseiltään samanlaista, fraseologisesta näkökulmasta käytössä saattaa olla eroja. Sen takia katsotaan tarkemmin, eroavatko natiivien ja kielenoppijoiden astemääritteiden kontekstuaaliset valinnat.

#### 4.2.1. Astemääritteiden kotekstuaalisia piirteitä natiivien kielessä

Peter Howarthin (2001: 163) tavoin tässä kielenoppijoiden fraseologisen pätevyyden tarkastelu aloitetaan tarkastelemalla natiivien ”normeja”. Sitä varten kuvaillaan astemääritteiden *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* kotekstuaalista ympäristöä Eduskunta-aineistossa. Tarkastelu aloitetaan abstraktimmalta kolligationaaliselta tasolta ja sen jälkeen lähdetään konkreettisempiin kotekstuaalisiin valintoihin – kollokaatioon. Nämä tavoitteet toteutettiin WordSmith Tools 8.0 ohjelmaketin Concord-ohjelmalla, jonka avulla voi etsiä sanan kollokaatteja ennalta rajatulta tarkastelualueelta.

Tässä tutkimuksessa tyydyttiin V2–O2 tarkastelualueeseen. Ensiksi poimittiin tarkastelualueelta kollokaatit ja kolligaatioanalyysia varten määriteltiin manuaalisesti niiden sanaluokat. Oikeanpuoleinen hakualue määritettiin astemääritteiden syntaktisen tehtävän perusteella. O1-positiossa oletettiin esiintyvän ennen kaikkea adjektiiveja, adverbejä ja prepositioita, O2-positioissa puolestaan postpositioita. Tarkastelualuetta laajennettiin vasemmalle oikeanpuoleisen kanssa symmetriseksi (V1–V2). Kollokaattien etsinnässä ei ylitetty virkkeenrajoja. Yksityiskohtaisempi kvantitatiivinen analyysi tehtiin Excel-ohjelmalla.

Vasemman- ja oikeanpuoleista kotekstia käsitellään erikseen. Näin on *aivan*-sanalla sekä vasemman- että oikeanpuoleisen tarkastelualueen koko sanemäärin ilmaistuna 36 180 (2x18 090) sanetta, *erittäin*-sanalla on tarkastelualue 55 842 (2x27 921) sanetta, *ihan*-sanalla 34 794 (2x17 397) sanetta, *kovin*-sanalla 9096 (2x4548) sanetta, *oikein*-sanalla 3492 (2x1746) sanetta ja *sangen*-sanalla 282 (2x141) sanetta.

Eduskunta-aineistossa astemääritteiden käyttöluvut ovat selvästi toisistaan eroavia. Eniten esiintyvää *erittäin*-sanaa (yhteensä 27 921 esiintymää) käytetään 198 kertaa enemmän kuin vähiten esiintyvää *sangen*-sanaa (yhteensä 141 esiintymää). Näin *erittäin*-sanalla 5 kertaa esiintyvää kollokaatti olisi marginaalinen, mutta *sangen*-sanalla jopa merkittävä. Siitä syystä sanojen tarkastelemista niiden todellisten frekvenssien perusteella ei voi pitää proportionaalisenä.

Proportionaalisen vertailtavuuden laskemiseksi lähdettiin vähiten esiintyvistä astemääritteestä *sangen*. Voi olettaa, että yhteisesiintymä ei esiinny kaksi kertaa satunnaisena, samalla kuin kerran esiintyvistä myötäesiintymästä ei voi olla varma. *Sangen*-sanank tarkastelualueella 2 kertaa esiintyvä kollokaatti muodostaa tarkastelualueesta 0,7 % ja sanan kokonaisfrekvenssistä 1,4 %. Samaa vähimmäiskynnystä sovellettiin tästä eteenpäin kaikkiin tutkittaviin astemääritteisiin sekä kolligaatio- että kollokaatioanalyysissä. Näin *aivan*-sanank kollokaattien vähimmäisfrekvenssiksi tuli 253, *erittäin*-sanasta vastaava luku on 391, *ihan*-sanasta 244, *kovin*-sanasta 64 ja *oikein*-sanasta 24. Siitä ei voi päätellä, että vähemmän esiintyvät yhteisesiintymät olisivat satunnaisia, ne eivät vain ole yhtä merkittäviä kuin enemmän esiintyvät. Vähimmäiskynnuksesta johtuen sanaluokkakolligaatit eivät kata 100 % sanank tarkastelualueesta.

Luvun lopussa esitetään jokaisesta astemääritteestä viisi eniten käytettyä klusteria. Nekin on poimittu WordSmith Tools 8.0 ohjelmapaketin Concord-ohjelmalla.

#### 4.2.1.1. Kolligaatio

Kolligationaalisissa analyysissä tarkastellaan sanaluokkakolligaatteja. Astemääritteiden tarkastelualueella esiintyy verbejä, substantiiveja, pronomineja, adpositioita, adverbejä, kvanttoriadverbejä, adjektiiveja, konjunktioita ja partikkeleita. Vaikka konjunktioitkin ovat partikkeleita (VISK: § 795) ja tämän tutkimuksen kvanttorit adverbejä (VISK: § 657), käsitellään niitä muista saman sanaluokan sanoista erikseen, koska ne muodostavat suhteellisen itsenäiset ryhmät. Joskus myös rajat sanaluokkien välillä voivat olla häilyviä (VISK: § 438). Tässä tutkimuksessa *nyt*, *jo*, *vielä* ja *niin* luokiteltiin adverbeiksi, *tosi* adjektiiviksi ja *paljo* (*paljon*-sanank tavoin) kvanttoriksi.

##### 4.2.1.1.1. Vasemmanpuoleiset sanaluokkakolligaatit

Sanaluokkakolligaattien analyysi perustuu niiden frekvensseihin. Taulukossa 4 esitellään astemääritteiden *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* vasemmanpuoleisten (V2–V1) sanaluokkakolligaattien todelliset frekvenssit Eduskunta-aineistossa. Tilan säästämiseksi tästä eteenpäin käytetään taulukoissa ja kuvioissa lyhenteitä.

Taulukossa vasemmalta oikealle esitetään astemääritteen tarkastelualueen kokonaissanemäärä (T. alue), sen jälkeen verbi, (Ver.), substantiivi- (Sub.), pronomini- (Pro.), partikkeli- (Par.), konjunktio- (Kon.), adpositio- (Adp.) ja adverbikolligaattien (Adv.) frekvenssit.

Taulukko 4. **Astemääritteiden vasemmanpuoleiset sanaluokkakolligaatit Eduskunta-aineistossa.**

	<b>T.alue</b>	<b>Ver.</b>	<b>Sub.</b>	<b>Pro.</b>	<b>Par.</b>	<b>Kon.</b>	<b>Adp.</b>	<b>Adv.</b>
<i>aivan</i>	36 180	10 469	-	3784	254	1605	321	390
<i>erittäin</i>	55 842	20 019	-	4370	409	570	1136	-
<i>ihan</i>	34 794	9938	264	4553	524	1395	304	1074
<i>kovin</i>	9096	3756	-	674	123	290	-	266
<i>oikein</i>	3492	1085	34	380	41	193	30	59
<i>sangen</i>	282	68	6	22	2	4	-	8
<b>Yhteensä</b>	139 686	45 335	304	13 783	1353	4057	1791	1797

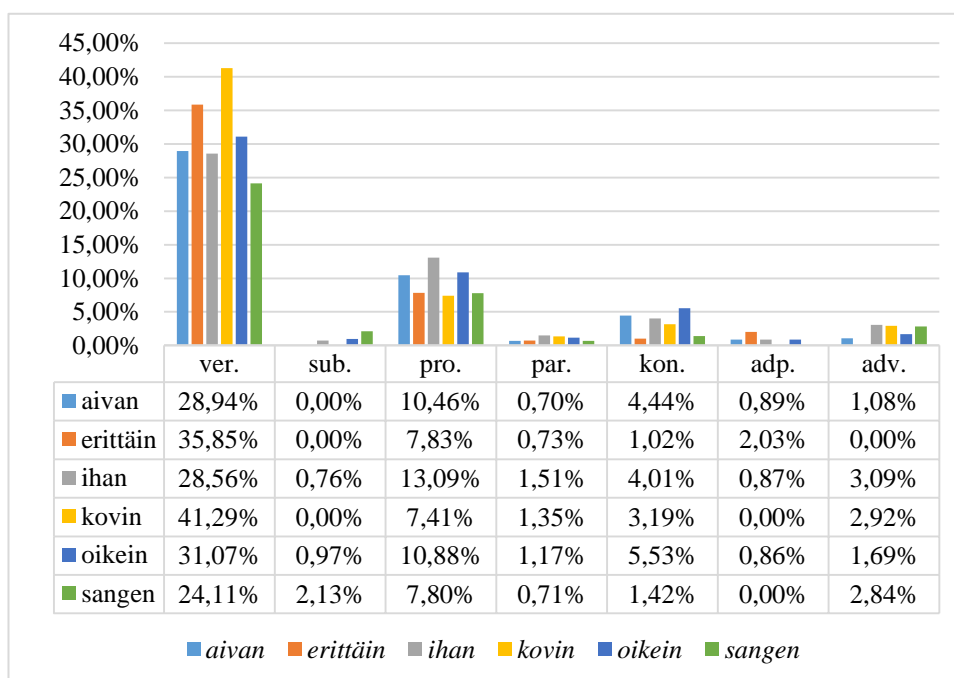
Taulukosta 4 tulee esiin, että *aivan*-sanana vasemmanpuoleiset sanaluokkakolligaatit ovat yleisyysjärjestyksessä: verbi, pronomini, konjunktio, adverbi, adpositio ja partikkeli. *Erittäin*-sanana kolligaattien yleisyysjärjestys on: verbi, pronomini, adpositio, konjunktio ja partikkeli. *Ihan*-sanalla tämä on: verbi, pronomini, konjunktio, adverbi, partikkeli, adpositio ja substantiivi. *Kovin*-sanalla: verbi, pronomini, konjunktio, adverbi ja partikkeli. *Oikein*-sanalla: verbi, pronomini, konjunktio, adverbi, partikkeli, substantiivi ja adpositio. *Sangen*-sanalla: verbi, pronomini, adverbi, substantiivi, konjunktio ja partikkeli.

Siitä näkyy, että kaikkien astemääritteiden yleisimpänä vasemmanpuoleisena sanaluokkakolligaattina on verbi (esimerkkilauseet 1–6) ja toiseksi eniten esiintyy pronomineja. Samanlaista on sekin, että partikkelikolligaattien osuus on kaikilla astemääritteillä suhteellisen matala, mutta muiden sanaluokkien osuuksissa näkyy pieniä eroja.

- (1) **Voin** *aivan* vilpittömästi sanoa että niillä alueilla näitten suurpetojen lukumäärä aiheuttaa todella suurta huolta. (Eduskunta)

- (2) Se ei kovin suuri määräraha olisi mutta se olisi tärkeä korvaus kilometrikorvauksena heille jotka vapaaehtoisesti **tekevät** tätä *erittäin* arvokasta työtä. (Eduskunta)
- (3) Kyllä he **tulivat** *ihan* normaalisti töihin kun pani ilmoituksen lehteen. (Eduskunta)
- (4) Ajoneuvo oikeuden yläraja Euroopassa **tuntuu** *kovin* harvinaiselta. (Eduskunta)
- (5) Toistan kaiken sen minkä olen niissä puheissani sanonut ja totean että vielä*kään ei mene oikein* hyvin. (Eduskunta)
- (6) Ihmeelliset ovat sinun tekosi sieluni **tietää** sen *sangen* hyvin. (Eduskunta)

Astemääritteiden keskinäistä vertailtavuutta varten havainnollistetaan taulukon 4 sisältöä kuviolla 2, jossa esitetään sanaluokkakolligaattien suhteelliset osuudet. Pystyakselilla kuvataan prosenttimäärät ja horisontaalisella akselilla tarkastelualueella esiintyvät sanaluokat. Kuvion alaosassa esitetään jokaisen astemääritteen sanaluokkakolligaattien suhteelliset osuudet koko tarkastelualueella.



Kuvio 2. Eduskunta-aineiston vasemmanpuoleiset sanaluokkakolligaatit.

Kuviosta näkyy, että yleisempien sanaluokkakolligaattien eli verbien valinnassa on astemääritteittäin eroja. Muita enemmän esiintyy verbejä *kovin* (41,29 %) ja *erittäin* (35,85 %) -sanojen yhteydessä, samalla kuin *sangen*-sanon kolligaateista verbit muodostavat vain neljäsosan (24,11 %). Toiseksi yleisemmän sanaluokan, pronomien käytöltä eroavat keskenään eniten *ihan* (13,09 %) ja *kovin* (7,41 %). Myös konjunktioitten käytössä on astemääritteittäin pieniä eroja. Niitä esiintyy muita enemmän *aivan*, *ihan*, *kovin* ja *oikein* -sanojen yhteydessä.

Muiden sanaluokkakolligaattien osuudet ovat pienempiä, mutta niissäkin näkyy muutamia eroja. Adverbejä on *ihan*, *kovin* ja *sangen* -tarkastelualueella noin 3 % tapauksista, mutta *erittäin*-sanalla ei yhtäkään. Adpositioita esiintyy yleisesti astemääritteitten vasemmanpuoleisessa kontekstissa vähän. Niitä ei käytetä kertaakaan *kovin*- ja *sangen*-tarkastelualueilla mutta myös *aivan*, *ihan* ja *oikein* -tarkastelualueella niiden käyttö jää alle prosentin. Muista erottuu hieman *erittäin*, jonka tarkastelualueella niitä on 2,03 %.

Muutamien sanaluokkakolligaattien puuttumisen voi toisaalta käsitellä kontekstuaalisena rajoituksena, toisaalta todennäköisemmin siihen vaikuttaa tässä tutkimuksessa esitetty 1,4 % vähimmäiskynnys, koska toisten astemääritteitten tarkastelualueelta puuttuvat sanat esiintyvät suhteellisen vähän myös toisten tarkastelualueilla. Muutamien sanaluokkakolligaattien puuttuminen tulee esiin enimmäkseen *erittäin* ja *kovin* vasemmanpuoleisessa kontekstissa. *Erittäin*-sana ei saa kolligaatiksi substantiiveja eikä adverbejä ja *kovin*-sanon kolligaattien joukosta puuttuvat substantiivit ja adpositiot. Lisäksi näkyy, että *aivan*-sanon tarkastelualueella ei esiinny substantiiveja ja *sangen*-sanon tarkastelualueella adpositioita. Siitä ei voi kuitenkin tehdä yleistyyksiä sanojen käyttöä koskevista rajoituksista, koska vähimmäiskynnyksestä johtuen suhteellisen suuri osa aineistosta jää analyysin ulkopuolelle.

Substantiivien vähäinen tai puutteellinen osuus voi selittyä sillä, että erilaisiin tarkoitteisiin viitataan substantiivien sijaan pikemmin pronomineilla. Eduskunta-aineistossa kyse on usein tiettyyn teemaan keskittyvistä puheenvuoroista. Kun teema on jo kerran esitelty substantiivilla, siihen viitataan jatkossa pronominilla. Lisäksi voi syynä olla se, että

substantiivien kirjo on laaja (ks. esimerkkilauseet 4, 10), mutta ei ole sellaisia ns. perussubstantiiveja, jotka esiintyisivät niin usein, että ylittäisivät vähimmäyskynnyksen.

Kaikilla astemääritteillä on vasemmanpuoleisella tarkastelualueella vahva preferenssi verbeihin ja pronomineihin. *Aivan*-sanana tarkastelualueesta nämä kattavat jopa 39,39 %, *erittäin*-sanana tarkastelualueesta tämä osuus on 43,68 %, *ihan*-sanasta vastaavasti 41,65 %, *kovin*-sanasta 48,70 %, *oikein*-sanasta 41,95 % ja *sangen*-sanasta 31,91 %. Muiden sanaluokkakolligaattien osuudet ovat huomattavasti pienempiä.

#### 4.2.1.1.2. Oikeanpuoleiset sanaluokkakolligaatit

Seuraavaksi tarkastellaan *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* oikeanpuoleisia sanaluokkakolligaatteja. Toisin kuin vasemmanpuoleisella tarkastelualueella, oikealla ei ole adpositio- ja partikkelikolligaatteja, mutta siellä esiintyy adjektiivi- ja kvanttorikolligaatteja.

Taulukossa 5 esitellään astemääritteiden *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* oikeanpuoleisten (O1–O2) sanaluokkakolligaattien todelliset frekvenssit. Vasemmalta oikealle esitetään tarkastelualueen kokonaissanemäärä (T. alue), sen jälkeen verbi- (Ver.), substantiivi- (Sub.), pronomini- (Pro.), konjunktio- (Kon.), adverbi- (Adv.), adjektiivi- (Adj.) ja kvanttorikolligaattien (Kva.) frekvenssit.

**Taulukko 5. Astemääritteiden oikeanpuoleiset sanaluokkakolligaatit Eduskunta-aineistossa.**

	T. alue	Ver.	Sub.	Pro.	Kon.	Adv.	Adj.	Kva.
<i>aivan</i>	36 180	455	1037	1630	3005	3709	4945	-
<i>erittäin</i>	55 842	508	1903	937	2896	1013	14 683	1208
<i>ihan</i>	34 794	484	672	1805	2252	2092	3656	-
<i>kovin</i>	9096	351	74	433	437	90	966	895
<i>oikein</i>	3492	46	39	181	265	306	438	44
<i>sangen</i>	282	2	9	13	23	12	16	5
<b>Yhteensä</b>	139 686	1846	3734	4999	8878	7222	24 704	2152

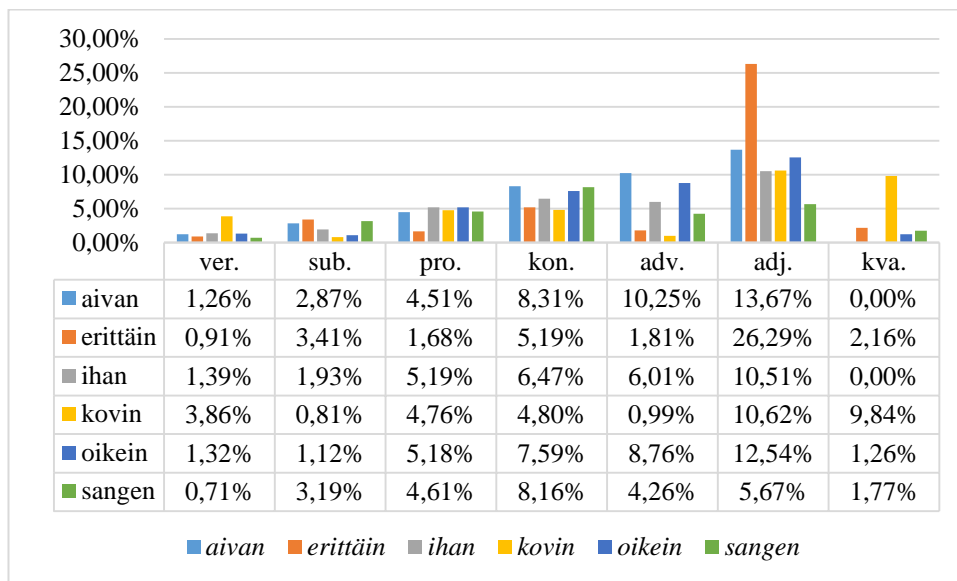
Taulukosta näkyy, että *aivan*-sanana sanaluokkakolligaatit ovat yleisyysjärjestyksessä: adjektiivi, adverbi, konjunktio, pronomini, substantiivi, verbi. *Erittäin*-sanana

sanaluokkakolligaattien yleisyysjärjestus on: adjektiivi, konjunktio, substantiivi, kvanttori, adverbi, pronomini, verbi. *Ihan*-sanalla tämä: adjektiivi, konjunktio, adverbi, pronomini, substantiivi, verbi. *Kovin*-sanalla tämä on: adjektiivi, kvanttori, konjunktio, pronomini, verbi, adverbi, substantiivi. *Oikein*-sanalla tämä on: adjektiivi, adverbi, konjunktio, pronomini, verbi, kvanttori, substantiivi. *Sangen*-sanalla tämä on: konjunktio, adjektiivi, pronomini, adverbi, substantiivi, kvanttori ja verbi.

Siitä näkyy, että astemääritteiden oikeanpuoleinen tarkastelualue on vasemmanpuoleista heterogeenisempi. Kaikilla astemääritteillä on suhteellisen paljon adjektiivi- (esimerkkilauseet 7–12) ja konjunktio-kolligaatteja, mutta sellaisia yleistyksiä kuin vasemmanpuoleisesta tarkastelualueesta ei voi tehdä. Suuret erot tulevat esiin astemääritteiden adverbi- ja kvanttorikolligaattien valinnassa.

- (7) Tämä koskee kaikkia kuntia *aivan isoja* ja pieniä. (Eduskunta)
- (8) Suun terveyden vaikutukset ihmisen terveyteen laajemminkin ovat *erittäin merkittäviä*. (Eduskunta)
- (9) Te elätte kyllä *ihan eri* maailmassa kuin nämä tavalliset suomalaiset ihmiset elävät. (Eduskunta)
- (10) Ja työpätkiä ei *kovin pitkiä* pätkiä välttämättä kaikille ole tarjolla. (Eduskunta)
- (11) Olisi syytä miettiä myös se mikä on piilotyöttömyyden rakenne tällä hetkellä ja tehdä siitä *oikein kunnan* selvitys. (Eduskunta)
- (12) Tuolloin olimme *sangen yksimielisiä* siitä että Suomi nousee vain työhön ja työllisyyteen panostamalla. (Eduskunta)

Astemääritteiden keskinäistä vertailtavuutta varten havainnollistetaan taulukon 5 sisältöä kuviolla 3, jossa esitetään sanaluokkakolligaattien suhteelliset osuudet. Pystyakselilla kuvataan prosenttimäärät ja horisontaalisella akselilla tarkastelualueella esiintyvät sanaluokat. Kuvion alaosassa esitetään jokaisen astemääritteen sanaluokkakolligaattien suhteelliset osuudet koko tarkastelualueella.



Kuvio 3. Eduskunta-aineiston oikeanpuoleiset sanaluokkakolligaatit.

Kuviosta näkyy, että astemääritteet eroavat keskenään ensisijaisesti adjektiivi-, adverbi- ja kvanttorikolligaattien valinnassa. Muita huomattavasti enemmän adjektiiveja esiintyy *erittäin*-sanalla (26,29 %) oikeanpuoleisella tarkastelualueella ja muita vähemmän *sangen* (5,67 %) tarkastelualueella. Ilmenee, että muiden astemääritteiden oikeanpuoleisella tarkastelualueella kvanttorikolligaatteja esiintyy suhteellisen vähän (*erittäin*, *oikein*, *sangen*) tai ne puuttuvat kokonaan (*aivan*, *ihan*), mutta *kovin*-sanalla kvanttorit ovat toiseksi yleisin (9,84 %) sanaluokkakolligaatti. Adverbikolligaatteja on suhteellisen paljon *aivan* (10,25 %) ja *oikein* (8,76 %) yhteydessä, mutta *kovin*-sanalla kvanttorit ovat marginaalisen osuuden (0,99 %). Adverbikolligaattien osuus on vähäisehkö myös *erittäin*- (1,81 %) ja *sangen*-sanoilla (4,26 %).

Astemääritteillä konjunktio-kolligaatteja on aika tasaisesti ja niiden osuus jää kaikkien astemääritteiden oikeanpuoleisilla tarkastelualueilla 4,80 %:sta 8,31 %:in rajoihin. Myös pronominkolligaattien osuus varioidaan melkein kaikilla noin 5 % ympärillä, mutta *erittäin* eroaa siinä muista astemääritteistä. Sen kolligaateista pronominit muodostavat vain 1,68 %. Substantiivikolligaattien osuus on kaikilla sanoilla suhteellisen matala (0,81 %:sta 3,41 %:iin). Huomiota herättää, että *aivan*, *ihan*, *kovin*, *oikein* ja *sangen* tarkastelualueella pronomineja esiintyy substantiiveja hieman enemmän, mutta *erittäin*-

sanasta osuudet ovat päinvastoin. Suhteellisen vähän esiintyy myös verbikolligaatteja (0,71 %:sta 1,39 %:iin), mutta *kovin* eroaa siinä muista, koska sen yhteydessä verbejä on jopa 3,86 %.

Oikeanpuoleinen koteksti näyttää erottavan synonyymisina pidetyt astemääritteet toisistaan. Adverbi-, adjektiivi- ja kvanttorikolligaattien osuuksien perusteella voi päätellä, että *aivan* määrittää aika tasaisesti sekä adjektiiveja (13,67 %) että adverbejä (10,25 %), mutta ei ollenkaan kvanttoreita. *Erittäin* määrittää pääasiassa adjektiiveja (26,29 %), samalla kun kvanttorien (2,16 %) ja adverbien (1,81 %) osuudet ovat suhteellisen pieniä. *Ihan* määrittää tyypillisesti adjektiiveja (10,51 %) ja vähemmän adverbejä (6,01 %), mutta ei ollenkaan kvanttoreita. *Kovin* määrittää tasaisesti sekä adjektiiveja (10,62 %) että kvanttoreita (9,84 %), mutta adverbejä suhteellisen vähän (0,99 %). *Oikein* toimii enimmäkseen adjektiivien (12,54 %) ja adverbien (8,76 %) etumääritteenä, kvanttoreiden osa on marginaalinen (1,26 %). *Sangen* määrittää muita tasaisemmin adjektiiveja (5,67 %), adverbejä (4,26 %), mutta myös kvanttoreita (1,77 %), vaikka niiden osuus on hieman vähäisempi. Muutamien sanaluokkakolligaattien puuttumiseen voi vaikuttaa myös vähimmäiskynnys.

Jokaisen astemääritteen oikeanpuoleisella tarkastelualueella pitää olla joko adjektiivi, adverbi, kvanttoriadverbi tai adpositio. Nämä esiintyvät joko O1- tai O2-positiossa eli vain yhdessä positiossa kerrallaan. Näin näiden sanaluokkien osuuden pitää olla yhteenlaskettuna oikeanpuoleisella tarkastelualueella 50 %. Tässä *aivan*-sanasta osuus on 23,92 %, *erittäin*-sanasta 30,27 %, *ihan*-sanasta 16,52 %, *kovin*-sanasta 21,45 %, *oikein*-sanasta 22,57 % ja *sangen*-sanasta 11,70 %. Syy vähemmän kuin puolen tarkastelualan kattamiseen piilee taas vähimmäiskynnöksessä.

Edellä esitetyn perusteella voi päätellä, että mitä matalampi on kolligaattien kattavuus tarkastelualueella, sitä enemmän astemäärite määrittää alle 0,7 % esiintyviä sanoja. Sitä mukaa kasvaa myös määritettävien sanojen kirjo. Kun määritettävien sanojen kirjo on tasaisesti matala, astemäärite toimii todennäköisesti yleismääritteenä, koska se ei suosi vain muutamia harvoja sanoja määritettäväiksi. Kolligaatioanalyysin perusteella muita

todennäköisemmin yleismääritteinä toimivat *sangen* ja *ihan*, jotka määrittävät suuresti myös harvinaisempia sanoja.

Kolligaatioanalyysistä ilmeni, että astemääritteiden vasemmanpuoleinen konteksti on melko homogeeninen ja yleisempien sanaluokkakolligaattien suhteelliset osuudet limittyvät, mutta oikeanpuoleisella tarkastelualueella näkyy yleisempien sanaluokkakolligaattien eli adjektiivien, adverbien ja kvanttoreiden valinnassa isoja eroja. Myös konjunktio-kolligaatteja esiintyy paljon, mutta niiden käyttö on tasaista. Astemääritteiden yksityiskohtaisemmat valinnat selviävät kollokaatioanalyysissä.

#### 4.2.1.2. Kollokaatio

Jantusen (2004: 90–96) tavoin kollokaatiota tarkastellaan kunkin sanaluokan osalta erikseen. Tällainen tarkastelu tarjoaa hyvän katsauksen astemääritteiden keskinäisiin samanlaisuuksiin ja eroihin. Seuraavaksi esitellään sekä vähimmäiskynnyksen ylittäviä kollokaatteja että niiden suhteellisia osuuksia kaikista saman sanaluokan kollokaateista vähimmäiskynnyksen yläpuolella. Kollokaattien todelliset frekvenssit ovat tarkasteltavissa 4.2.2. luvun taulukoissa.

##### 4.2.1.2.1. Vasemmanpuoleiset kollokaatit

**Verbikollokaatit.** Edellisessä luvussa kävi ilmi, että kaikkien astemääritteiden vasemmanpuoleisella tarkastelualueella esiintyy eniten verbejä ja pronomineja. Kun tarkastellaan vähimmäiskynnyksen yläpuolella olevia kollokaatteja lähemmin, näkyy, että kaikkien astemääritteiden yleisin verbikollokaatti on OLLA.

- *Aivan* <OLLA 93,00 %, EI 4,25 %, VOIDA 2,75 %> (yhteensä 10 469).
- *Erittäin* <OLLA 89,00 %, PITÄÄ 5,83 %, TEHDÄ 3,18 %, TULLA 1,99 %> (yhteensä 20 019).
- *Ihan* <OLLA 78,93 %, EI 11,35 %, VOIDA 4,34 %, TULLA 2,88 %, PITÄÄ 2,51 %> (yhteensä 9938).
- *Kovin* <OLLA 56,44 %, EI 37,51 %, VOIDA 2,24 %, TUNTUA 1,94 %, TULLA 1,86 %> (yhteensä 3756).

- *Oikein* <OLLA 56,96 %, EI 23,13 %, YMMÄRTÄÄ 4,24 %, PITÄÄ 3,13 %, TEHDÄ 2,76 %, TULLA 2,67 %, VOIDA 2,49 %, MENNÄ 2,30 %, SAADA 2,30 %> (yhteensä 1085).
- *Sangen* <OLLA 97,06 %, TIETÄÄ 2,94 %> (yhteensä 68).

OLLA-kollokaatin suhteellinen osuus on korkein *sangen* (97,06 %), *aivan* (93 %), *erittäin* (89 %) vasemmanpuoleisella tarkastelualueella, mutta myös muilla tämä kattaa yli puolen kaikista verbikollokaateista. *Ihan*-sanana verbikollokaateista OLLA muodostaa 78,93 %, *oikein*-sanasta vastaava osuus on 56,96 % ja *kovin*-sanasta 56,44 %.

Jos sanan verbikollokaattina on EI, se on aina toiseksi yleinen verbikollokaatti. EI-kollokaatin osuus on korkein *kovin*-sanana tarkastelualueella, jossa se on jopa 37,51 % kaikista verbikollokaateista (esimerkkilause 13). EI-kollokaattia esiintyy myös *oikein* (21,13 %), *ihan* (11,35 %) ja *aivan* (4,25 %) ympäristössä (esimerkkilauseet 14–16), mutta ei kertaakaan *sangen* ja *erittäin* -tarkastelualueen vähimmäiskynnyksen yläpuolella.

- (13) Kansalla **ei** ole *kovin* hauskaa mutta on hyvä että ainakin täällä edustajilla tuntuu olevan. (Eduskunta)
- (14) **Ei** ole *oikein* hyvä että kansalle välittyy television suorassa lähetyksessä näinkin merkittävästä asiasta näin radikaalisti väärä tieto. (Eduskunta)
- (15) Voitte kuvitella että tämä **ei** ole *ihan* päivänselvä asia kaikille Euroopan maille mutta meille se olisi mahdollisuus koska meidän puolustusteollisuus on teknologiakehittäjä mutta myös hyvin tuottava. (Eduskunta)
- (16) Tältä osin täytyy sanoa että kysymys **ei** ole *aivan* ongelmaton. (Eduskunta)

Kaikkien muiden aineistossa esiintyvien verbien osuus jää alle 6 %:in. Muita laajempi verbikirjo on *oikein* tarkastelualueella, jossa erilaisia vähimmäiskynnyksen ylittäviä verbejä on 9, suppeampi verbikirjo on *sangen* tarkastelualueella, jossa erilaisia verbejä on 2.

OLLA-kollokaatin yleisyys johtuu suuressa määrin todennäköisesti siitä, että astemääritteitä käytetään usein predikaatiivilauseissa. Lisäksi Jantunen (2004: 90) nostaa

esiin, että suomen kielen yleisempänä sanana OLLA (ks. Heikkinen ym. 2001) löytyykin yleensä kollokaattilistojen kärkipäästä. Samalla kaikki muutkin vasemmanpuoleiset verbikollokaatit (paitsi TUNTUA ja YMMÄRTÄÄ) ovat suomen kielen sadan taajimmin esiintyvän sanan seassa (ks. Saukkonen ym. 1979: 41–42).

**Substantiivikollokaatit.** Vähimmäiskynnyksen ylittäviä substantiivikollokaatteja esiintyy vain *ihan*, *oikein* ja *sangen* oikeanpuoleisella tarkastelualueella. *Ihan* ja *oikein* saavat substantiivikollokaatikseen ASIA ja *sangen* sekä ESITYS (66,67 %) että HALLITUS (33,33 %).

- *Ihan* (yhteensä 264), *oikein* (yhteensä 34) <ASIA 100 %>.
- *Sangen* <ESITYS 66,67 %, HALLITUS 33,33 %> (yhteensä 6).

**Pronominikollokaatit.** Kaikille astemääritteille yhteiset vähimmäiskynnykset ylittävät pronominikollokaatit ovat MINÄ ja TÄMÄ, joista jälkimmäinen on *erittäin*- (61,69 %) ja *sangen*- (36,36 %) sanojen yleisin kollokaatti. SE puolestaan on *aivan* (44,64 %), *ihan* (46,87 %), *kovin* (54,60 %) ja *oikein* -sanojen yleisimpänä kollokaattina.

- *Aivan* <SE 44,64 %, TÄMÄ 31,69 %, MINÄ 15,43 %, JOKA 8,25 %> (yhteensä 3784).
- *Erittäin* <TÄMÄ 61,69 %, MINÄ 26,54 %, JOKA 11,76 %> (yhteensä 4370).
- *Ihan* <SE 46,87 %, TÄMÄ 30,92 %, MINÄ 16,30 %, JOKA 5,91 %> (yhteensä 4553).
- *Kovin* <SE 54,60 %, TÄMÄ 32,64 %, MINÄ 12,76 %> (yhteensä 674).
- *Oikein* <SE 39,21 %, TÄMÄ 36,32 %, MINÄ 24,47 %> (yhteensä 380).
- *Sangen* <TÄMÄ 36,36 %, SE 31,82 %, MINÄ 18,18 %, JOKA 13,64 %> (yhteensä 22).

Erilaisia pronominikollokaatteja on aineistossa vähimmäiskynnyksen yläpuolella yhteensä neljä. Nämä ovat SE, TÄMÄ, MINÄ, JOKA (esimerkkilauseet 17–20). SE puuttuu *erittäin*-sanan tarkastelualueelta vähimmäiskynnyksen yläpuolelta ja JOKA sekä *kovin*- että *oikein*-sanojen tarkastelualueelta vähimmäiskynnyksen yläpuolelta.

- (17) Tuollaisenkin sairauden ollessa kyseessä voi olla viikoista kysymys jos **se** on *kovin* aggressiivista laatua. (Eduskunta)
- (18) **Tämä** on *ihan* ok. (Eduskunta)
- (19) **Minulla** on *aivan* erityinen suhde nyt käsiteltävänä olevaan asiaan. (Eduskunta)
- (20) Tölli kävi läpi myös niitä lakimuutoksiin tehtyjä rajoituksia **jotka** ovat *sangen* tarpeellisia ja juuri niitä vastaanhan täällä ed. (Eduskunta)

**Partikkelikollokaatit.** Vähimmäiskynnyksen yläpuolelle yltää vain KYLLÄ, joka esiintyy kaikkien astemääritteiden tarkastelualueilla (esimerkkilauseet 21–26).

- *Aivan* (yhteensä 254), *erittäin* (yhteensä 409), *ihan* (yhteensä 524), *kovin* (yhteensä 123), *oikein* (yhteensä 41), *sangen* (yhteensä 2) <KYLLÄ 100 %>.

- (21) Siinä suhteessa minulla on **kyllä** *aivan* eri arvio kuin pääministerillä tästä asiasta. (Eduskunta)
- (22) Tuli **kyllä** *erittäin* erikoinen olo kun kuuntelin ulkoministerin avauspuheenvuoron. (Eduskunta)
- (23) Nyt syntyy **kyllä** *ihan* väärä kuva siitä. (Eduskunta)
- (24) Täytyy sanoa että jos pitäisi tätä järjestelmää selvittää niin se ei **kyllä** *kovin* helposti onnistu peruskoulussa eikä myöskään välttämättä tässä salissa. (Eduskunta)
- (25) Muuten te vaikutatte **kyllä** olevan *oikein* hyvä hallituksen apupuolue. (Eduskunta)
- (26) Ja niin kuin äsken sanoin jos on 42000 ihmistä niin jos sieltä ei muutamaan metsästysseuraan löydy jäseniä niin **kyllä** minä *sangen* suuresti ihmettelen. (Eduskunta)

**Konjunktio-kollokaatit.** JA, MUTTA, ETTÄ ovat *aivan*, *ihan* ja *kovin* -sanojen tarkastelualueilla yhteisiä vähimmäiskynnyksen ylittäviä konjunktio-kollokaatteja. *Oikein*-sanon konjunktio-kollokaattien kirjo on muita astemääritteitä laajempi ja sen tarkastelualueella esiintyy vähimmäiskynnyksen yläpuolella yhteensä neljä

konjunktiokollokaattia. Niiden seassa myös JOS ja KUN, joita muiden astemääritteiden tarkastelualueilla vähimmäiskynnyksen yläpuolella ei ole. *Erittäin*- ja *sangen*-sanoilla on kummallakin vähimmäiskynnyksen yläpuolella vain yksi kollokaatti. *Erittäin*-sanalla tämä on MYÖS ja *sangen*-sanalla JA.

- *Aivan* <JA 56,57 %, MUTTA 23,18 %, ETTÄ 20,25 %> (yhteensä 1605).
- *Erittäin* <MYÖS 100 %> (yhteensä 570).
- *Ihan* <JA 46,81 %, ETTÄ 32,97 %, MUTTA 20,22 %> (yhteensä 1395).
- *Kovin* <ETTÄ 50,00 %, MUTTA 25,86 %, JA 24,14 %> (yhteensä 290).
- *Oikein* <JA 37,31 %, JOS 27,98 %, ETTÄ 18,13 %, KUN 16,58 %> (yhteensä 193).
- *Sangen* <JA 100 %> (yhteensä 4).

**Adpositiokollokaatit.** Aineistossa esiintyy kaksi vähimmäiskynnyksen ylittävää adpositiokollokaattia. Ne ovat *aivan*, *erittäin*, *ihan* ja *oikein* tarkastelualueella esiintyvä MIELESTÄ ja *erittäin*-sanalla KANNALTA.

- *Aivan* (yhteensä 321), *ihan* (yhteensä 304), *oikein* (yhteensä 30) <MIELESTÄ 100 %> (yhteensä 321).
- *Erittäin* <MIELESTÄ 57,31 %, KANNALTA 42,69 %> (yhteensä 1136).

**Adverbikollokaatit.** *Ihan*, *koivin*, *aivan* ja *oikein* jakavat vähimmäiskynnyksen yläpuolella vain NYT-kollokaatin. Lisäksi esiintyy *ihan*-sanalla SITTEN ja *koivin* tarkastelualueella NIIN. *Erittäin*-sanalla vähimmäiskynnyksen ylittäviä adverbikollokaatteja ei ole.

- *Aivan* (yhteensä 390), *oikein* (yhteensä 59) <NYT 100 %> (yhteensä 390).
- *Ihan* <NYT 67,41 %, SITTEN 32,59 %> (yhteensä 1074).
- *Kovin* <NIIN 62,03 %, NYT 37,97 %> (yhteensä 266).
- *Sangen* <JO 75 %, VIELÄ 25 %> (yhteensä 8).

*Sangen*-sanalla ei ole vähimmäiskynnyksen yläpuolella muiden kanssa yhteisiä kollokaatteja. Sen kollokaatit ovat JO ja VIELÄ (esimerkkilauseet 27–28).

(27) Valtiontalouden tarkastusviraston vuosikertomus nostaa esiin myös lainsäädäntötyössä **jo sangen** yleiseksi kasvaneen ongelman. (Eduskunta)

(28) Pikaluotto on meidän luottomarkkinoillamme **vielä sangen** tuore ilmiö. (Eduskunta)

#### 4.2.1.2.2. Oikeanpuoleiset kollokaatit

Kolligaatioanalyysistä kävi ilmi, että astemääritteet eroavat keskenään ensisijaisesti adjektiivi-, adverbi- ja kvanttorikolligaattien valinnassa.

**Verbikollokaatit.** OLLA on kaikille astemääritteille yhteinen kollokaatti. *Aivan-*, *erittäin-*, *ihan-*, *oikein-* ja *sangen-*sanoilla ei olekaan muita vähimmäiskynnyksen ylittäviä kollokaatteja.

- *Aivan* (yhteensä 455), *erittäin* (yhteensä 508), *ihan* (yhteensä 484), *oikein* (yhteensä 46), *sangen* (yhteensä 2) <OLLA 100 %>.
- *Kovin* <OLLA 69,52 %, EI 30,48 %> (yhteensä 351).

Astemäärite *kovin* eroaa muista astemääritteistä EI kollokaatin puolesta (esimerkkilause 29), joka muodostaa jopa 30,45 % kaikista verbikollokaateista.

(29) Eli tätä valvontaa lisätään ja lisätään ja sekään ei ole mitenkään *kovin* edullista **eikä** halpaa. (Eduskunta)

**Substantiivikollokaatit.** Kaikille astemääritteille yhteinen kollokaatti on ASIA, joka on suurimmalla osalla astemääritteitä ainoa vähimmäiskynnyksen ylittävä kollokaatti. *Aivan*-sanana kollokaattina esiintyy lisäksi TAPA ja *sangen*-sanana kollokaattina RATKAISU, ESITYS ja HALLITUS.

- *Erittäin* (yhteensä 1903), *ihan* (yhteensä 672), *kovin* (yhteensä 74), *oikein* (yhteensä 39) <ASIA 100 %>
- *Aivan* <ASIA 72,13 %, TAPA 27,87 %> (yhteensä 1037).
- *Sangen* <RATKAISU 33,33 %, ESITYS 22,22 %, HALLITUS 22,22 %, ASIA 22,22 %> (yhteensä 9).

**Pronominikollokaatit.** Kaikille astemääritteille on yhteinen kollokaatti TÄMÄ, joka on *erittäin*-sanan ainoa vähimmäiskynnyksen ylittävä pronominkollokaatti. Kaikkien muiden tarkastelualueella löytyy myös SE. Lisäksi *ihan* ja *aivan* saavat kollokaatikseen myös MUU ja *sangen* TUO.

- *Aivan* <SE 48,34 %, TÄMÄ 34,60 %, MUU 17,06 %> (yhteensä 1630).
- *Erittäin* <TÄMÄ 100 %> (yhteensä 937).
- *Ihan* <SE 47,65 %, TÄMÄ 34,13 %, MUU 18,23 %> (yhteensä 1805).
- *Kovin* <SE 51,50 %, TÄMÄ 48,50 %> (yhteensä 433).
- *Oikein* <SE 55,80 %, TÄMÄ 44,20 %> (yhteensä 181).
- *Sangen* <TÄMÄ 61,54 %, SE 23,08 %, TUO 15,38 %> (yhteensä 13).

**Konjunktio-kollokaatit.** Ainoa kaikille astemääritteille yhteinen vähimmäiskynnyksen ylittävä konjunktio-kollokaatti on ETTÄ. Sen lisäksi JA esiintyy *aivan*-, *ihan*-, *kovin*-, *oikein*- ja *sangen*-sanojen kollokaattina, MUTTA *ihan*- ja *sangen*-sanojen kollokaattina, KUIN *aivan*-sanan kollokaattina ja KUN *oikein*-sanan kollokaattina.

- *Aivan* <ETTÄ 68,72 %, JA 22,76 %, KUIN 8,52 %> (yhteensä 3005).
- *Erittäin* <ETTÄ 100 %> (yhteensä 2896).
- *Ihan* <ETTÄ 55,82 %, JA 27,66 %, MUTTA 16,52 %> (yhteensä 2252).
- *Kovin* <JA 53,32 %, 30,66 %, ETTÄ 16,02 %> (yhteensä 437).
- *Oikein* <JA 46,04 %, ETTÄ 42,64 %, KUN 11,32 %> (yhteensä 265).
- *Sangen* <JA 73,91 %, ETTÄ 17,39 %, MUTTA 8,70 %> (yhteensä 23).

**Adverbikollokaatit.** Kaikille astemääritteille on yhteinen HYVIN, joka on *erittäin*- ja *kovin*-sanojen ainoa vähimmäiskynnyksen ylittävä adverbikollokaatti. Muiden astemääritteiden adverbikollokaattien kirjo on laajempi, muista eroaa erityisesti *aivan*, joka saa vähimmäiskynnyksen yläpuolella viisi erilaista kollokaattia, *ihan* ja *sangen* puolestaan saavat neljä ja *oikein* kaksi kollokaattia.

- *Erittäin* (yhteensä 1013), *kovin* (yhteensä 90) <HYVIN 100 %>.

- *Aivan* <OIKEIN 42,19 %, LIIAN 29,20 %, VARMASTI 12,67 %, HYVIN 9,09 %, ERITYISESTI 6,85 %> (yhteensä 3709).
- *Ihan* <OIKEASTI 41,16 %, OIKEIN 31,41 %, HYVIN 15,58 %, NIIN 11,85 %> (yhteensä 2092).
- *Oikein* <HYVIN 75,82 %, KUNNOLLA 24,18 %> (yhteensä 306).
- *Sangen* <HYVIN 41,67 %, SUURESTI 25,00 %, NIIN 16,67 %, SIKSI 16,67 %> (yhteensä 12).

OIKEIN esiintyy sekä *aivan-* että *ihan-*sanojen kollokaattina (esimerkkilauseet 30–31). NIIN esiintyy sekä *ihan* että *sangen* kollokaattina. Vähimmäiskynnyksen yläpuolella LIIAN, VARMASTI (esimerkkilauseet 32–33) ja ERITYISESTI ovat vain *aivan* kollokaatteina, OIKEASTI vain *ihan* kollokaattina (esimerkkilause 34), KUNNOLLA vain *oikein* kollokaattina (esimerkkilause 35) ja sekä SUURESTI (esimerkkilause 36) että SIKSI vain *sangen* kollokaatteina. Suhteellisten osuuksien perusteella näkyy, että *aivan* suosii erityisesti OIKEIN-, *ihan* OIKEASTI- ja kaikki muut HYVIN-kollokaattia.

- (30) Ammatillista koulutusta lisätään mikä onkin *aivan oikein*. (Eduskunta)
- (31) Toisaalta Ben Zyskowicz toi *ihan oikein* esiin sen että panostuksia pitää tehdä myös palveluihin ja näin on tehty. (Eduskunta)
- (32) Vuosihan on nyt aivan käsillä mutta tehty on *aivan liian* vähän. (Eduskunta)
- (33) Ja aivan totta on että nämä matkasynnytykset tulevat *aivan varmasti* lisääntymään. (Eduskunta)
- (34) Sitten sanotaan että tällä pelastetaan se juttu niin onhan tämä *ihan huuhaapuhetta ihan oikeasti*. (Eduskunta)
- (35) Erityistä huolta on kannettava siitä että maakunnallinen yhdenvertaisuus ei *oikein kunnolla* toteudu. (Eduskunta)
- (36) Ihmettelen *sangen suuresti* ketä tällainen näennäispuuhastelu hyödyttää. (Eduskunta)

**Adjektiivikollokaatit.** Vähimmäiskynnyksen yläpuolella kaikille astemääritteille yhteisiä kollokaatteja ei ole. *Erittäin-*, *ihan-*, *kovin-*, *oikein-* ja *sangen-*sanoille yhteinen

adjektiivikollokaatti on HYVÄ, *erittäin-*, *kovin-* ja *sangen-*sanoille SUURI, *erittäin-* ja *kovin-* sanoille VAIKEA. Eniten vähimmäiskynnyksen yläpuolella olevia yhteisiä kollokaatteja on *aivan-* ja *ihan-* sanoilla, nämä ovat: ERI, TOSI, SELVÄ, OIKEA ja äsken mainittu HYVÄ.

- *Aivan* <SELVÄ 22,06 %, OIKEA 14,48 %, KESKEINEN 9,28 %, UUSI 9,14 %, ERI 9,14 %, VÄLTTÄMÄTÖN 8,53 %, TOSI 8,03 %, VARMA 7,68 %, KÄSITTÄMÄTÖN 6,11 %, ERINOMAINEN 5,54 %> (yhteensä 4945).
- *Erittäin* <TÄRKEÄ 40,78 %, HYVÄ 31,85 %, SUURI 9,87 %, MERKITTÄVÄ 6,54 %, VAIKEA 5,05 %, VAKAVA 3,19 %, ISO 2,73 %> (yhteensä 14 683).
- *Ihan* <HYVÄ 42,15 %, OIKEA 17,18 %, SELVÄ 15,92 %, OMA 9,57 %, TOSI 8,18 %, ERI 7,00 %> (yhteensä 3656).
- *Kovin* <SUURI 39,75 %, HYVÄ 21,33 %, HELPPO 13,04 %, VAIKEA 9,52 %, ERILAINEN 8,49 %, PITKÄ 7,87 %> (yhteensä 966).
- *Oikein* <HYVÄ 93,84 %, KUNNON 6,16 %> (yhteensä 438).
- *Sangen* <YKSIMIELINEN 25,00 %, MIELENKIINTOINEN 18,75 %, SUURI 18,75 %, HYVÄ 18,75 %, LYHYT 18,75 %> (yhteensä 16).

Vähimmäiskynnyksen yläpuolella *aivan-*sanalle uniikkeja kollokaatteja ovat KESKEINEN, UUSI, VÄLTTÄMÄTÖN, VARMA, KÄSITTÄMÄTÖN ja ERINOMAINEN. *Erittäin-*sanalle vähimmäiskynnyksen yläpuolella ainutlaatuisia ovat TÄRKEÄ, MERKITTÄVÄ, VAKAVA ja ISO, *ihan-*sanalle OMA, *kovin-*sanalle HELPPO, ERILAINEN ja PITKÄ, *oikein-*sanalle KUNNON ja *sangen-*sanalle YKSIMIELINEN, MIELENKIINTOINEN ja LYHYT.

Muista astemääritteistä *oikein* eroaa siinä mielessä, että sen kollokaattina on melkein aina HYVÄ, jonka osuus on muista vähimmäiskynnyksen yläpuolella olevista adjektiivikollokaateista 93,84 % (esimerkkilause 37). *Ihan-*sanin suosituin kollokaatti on HYVÄ (42,15 %), *erittäin-*sanalla TÄRKEÄ (40,78 %), *kovin-*sanalla SUURI (39,75 %), *sangen-*sanalla YKSIMIELINEN (25 %) ja *aivan-*sanalla SELVÄ (22,06 %) (esimerkkilauseet 38–42).

- (37) Tämä on *oikein* **hyvä** keskustelu ja tältä pohjalta tätä on hyvä jatkaa. (Eduskunta)
- (38) Tämähän tuntuu *ihan* **hyvältä** näin kerrottuna mutta tuloksethan eivät ole olleet sellaisia kuin olisi käytännössä toivottu. (Eduskunta)
- (39) Mutta ennen kaikkea tämä on inhimillisesti *erittäin* **tärkeää** ja omaishoito *erittäin* arvokasta työtä. (Eduskunta)
- (40) Summa ei edes ole *kovin* **suuri** suhteutettuna kaikkiin muihin menoihin. (Eduskunta)
- (41) Tuolloin olimme *sangen* **yksimielisiä** siitä että Suomi nousee vain työhön ja työllisyyteen panostamalla. (Eduskunta)
- (42) Se on *aivan* **selvää** että tätä on yrityksissä *aivan* mahdoton valvoa.

**Kvanttorikollokaatit.** Vähimmäiskynnyksen yläpuolella kvanttoireita esiintyy vain *erittäin-*, *kovin-*, *oikein-* ja *sangen-*sanojen tarkastelualueella. Vähimmäiskynnyksen yläpuolella *erittäin-* ja *oikein-* sanat saavat kvanttorikollokaatikseen vain PALJON- ja *sangen* VÄHÄN-kvanttorin, mutta *kovin* saa molemmat niistä ja lisäksi myös PALJO (esimerkkilauseet 43–48). Yleisin *kovin*-sanon kvanttorikollokaatti on kuitenkin PALJON (82,23 %; esimerkkilause 49).

- *Erittäin* (yhteensä 1208), *oikein* (yhteensä 44) <PALJON 100 %>.
- *Kovin* <PALJON 82,23 %, PALJO 8,94 %, VÄHÄN 8,83 %> (yhteensä 895).
- *Sangen* <VÄHÄN 100 %> (yhteensä 5).

- (43) Se on *erittäin erittäin* **paljon** rahaa. (Eduskunta)
- (44) Kiitoksia *oikein* **paljon**. (Eduskunta)
- (45) Kotimaista kalaa on ruokapöydässä *sangen* **vähän** ja tuemme norjalaista kalateollisuutta ja kalataloutta. (Eduskunta)
- (46) Mutta sähkötupakasta keskustelu on *kovin* **paljon** tutkimustietoa. (Eduskunta)
- (47) Lauslahti nosti esille palveluiden laadun ja se tietysti lämmitti sydäntä koska *kovin* **vähän** on täällä tänään kuultu tästä palveluiden laadusta. (Eduskunta)

(48) Ei ihme jos kansalaiset välillä sanovat etteivät ymmärrä politiikasta *kovin paljoa*. (Eduskunta)

(49) Se on paikka missä humanitäärisiä organisaatioita ei *kovin paljon* vielä ole liikkeellä. (Eduskunta)

Suuren määrän erojen lisäksi erityisesti adjektiivi-, adverbi- ja kvanttorikollokaattien valinnassa kollokaatioanalyysistä ilmeni sekin, että sanoilla on lisäksi yhteisiä kollokaatteja. Kuviossa 4 esitetään *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* yhteiset vähimmäiskynnyksen ylittävät kollokaatit position mukaan. Vasemmalta oikealle esitetään yhteiset kollokaatit ensiksi noodin vasemmalla puolella (V2–V1-positioit) ja sen jälkeen noodin oikealla puolella (O1–O2-positiot).

V2	V1	noodi	O1	O2
tämä	tämä		hyvin	tämä
olla	olla			asia
minä				että
kyllä				

Kuvio 4. Astemääritteiden *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* yhteiset sanaluokkakollokaatit.

#### 4.2.1.3. Klusteri

Sanojen taipumuksista esiintyä tietynlaisissa toistuvissa käyttöyhteyksissä antavat hyvän todisteen myös kielessä löytyvät klusterit. Vaikka tässä tutkimuksessa ei analysoida niitä tarkemmin, esitetään jokaisen astemääritteen viisi yleisempää klusteria.

Eduskunta-aineistossa astemäärite *aivan* esiintyy eniten seuraavissa klusterissa:

1. *Se on aivan* (919 kertaa).
2. *On aivan selvää* (639 kertaa).
3. *Aivan selvää että* (551 kertaa).
4. *On aivan selvää että* (513 kertaa).
5. *Tämä on aivan* (479 kertaa).

Astemääräte *erittäin* esiintyy eniten seuraavissa klustereissa:

1. *On erittäin hyvä* (1969 kertaa).
2. *On erittäin tärkeää* (1260 kertaa).
3. *Tämä on erittäin* (1194 kertaa).
4. *On erittäin tärkeä* (1123 kertaa).
5. *Se on erittäin* (1075 kertaa).

Astemääräte *ihan* esiintyy eniten seuraavissa klustereissa:

1. *Se on ihan* (958 kertaa).
2. *On ihan hyvä* (532 kertaa).
3. *Tämä on ihan* (385 kertaa).
4. *On ihan selvä* (248 kertaa).
5. *On ihan totta* (243 kertaa).

Astemääräte *kovin* esiintyy eniten seuraavissa klustereissa:

1. *Ei ole kovin* (466 kertaa).
2. *Ei kovin paljon* (105 kertaa).
3. *Ole kovin paljon* (102 kertaa).
4. *Se ei ole* (99 kertaa).
5. *Se ei ole kovin* (65 kertaa).

Astemääräte *oikein* esiintyy eniten seuraavissa klustereissa:

1. *On oikein hyvä* (151 kertaa).
2. *Ei ole oikein* (67 kertaa).
3. *Oikein hyvä että* (51 kertaa).
4. *Ymmärrän oikein hyvin* (38 kertaa).
5. *On oikein hyvä että* (35 kertaa).

Astemääritteestä *sangen* esiintyy eniten seuraavissa klustereissa:

1. *On ollut sangen* (4 kertaa).

2. *Sangen hyvin ymmärrän* (2 kertaa).
3. *Sangen epäselväksi tässä* (2 kertaa).
4. *Sangen epäreilusti pani* (2 kertaa).
5. *Pitää sangen pienenä* (2 kertaa).

#### 4.2.2. Astemääritteiden kontekstuaaliset erot natiivi- ja oppijanaineiston välillä

Natiiviaineiston tavoin oppijankielenanalyysiinkin sovelletaan astemääritteen kokonaisfrekvenssistä 1,4 %:in vähimmäiskynnystä. Näin otetaan tarkasteluun ainakin neljä kertaa esiintyvät *oikein*-sanana kollokaatit ja kaksi kertaa esiintyvät *erittäin*-sanana kollokaatit. Muista astemääritteistä vähimmäiskynnys vastaa vähemmän kuin kahta esiintymää. Tarkastelemalla vain kerran esiintyviä kollokaatteja ei voi tehdä yleistyksiä oppijankielen piirteistä. Sen takia otetaankin analyysiin vain kaksi tai useampia kertoja esiintyvät kollokaatit.

##### 4.2.2.1. *Aivan* vasemmanpuoleinen konteksti

Natiiviaineistossa *aivan* saa vasemmanpuoleisella tarkastelualueella 13 vähimmäiskynnystä ylittävää kollokaattia. Niistä muodostuneet kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- OLLA/SE/TÄMÄ/JA/MINÄ/EI/NYT/MUTTA/ETTÄ/MIELI/JOKA/VOIDA/KYLLÄ <*aivan*>.

Oppijanaineiston pienuudesta johtuen vähimmäiskynnystä (ja vähintään kahden esiintymän rajan) ylittivät vain 2 kollokaattia. Kielenoppijoiden käyttämät kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- OLLA/MINÄ <*aivan*>.

Taulukossa 6 esitetään kaikki vähimmäiskynnystä ylittävät *aivan*-sanana vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa. Vasemmalta oikealle esitetään kollokaatit, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %). Natiivi- ja oppijanaineiston yhteiset kollokaatit on lihavoitu.

Taulukko 6. *Aivan* vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.

Eduskunta			ICLFI		
Kollokaatti (t. alue 36 180)	f	f %	Kollokaatti (t. alue 34)	f	f %
<b>OLLA</b>	9 736	26,91 %	<b>OLLA</b>	13	38,24 %
SE	1 689	4,67 %	<b>MINÄ</b>	2	5,88 %
TÄMÄ	1 199	3,31 %			
JA	908	2,51 %			
<b>MINÄ</b>	584	1,61 %			
EI	445	1,23 %			
NYT	390	1,08 %			
MUTTA	372	1,03 %			
ETTÄ	325	0,90 %			
MIELESTÄ	321	0,89 %			
JOKA	312	0,86 %			
VOIDA	288	0,80 %			
KYLLÄ	254	0,70 %			

Natiivi- ja oppijanaineistolle on vähimmäiskynnyksen yläpuolella 2 yhteistä kollokaattia. Nämä ovat OLLA ja MINÄ. Oppijanaineistossa kaikki vähimmäiskynnyksen ylittävät kollokaatit ovat tavallisia myös natiiviaineistossa. Yhteiset kollokaatit muodostavat natiiviaineiston kollokaattikirjosta 15,38 % ja oppijanaineiston vähimmäiskynnyksen ylittävästä käytöstä 100 %. Verrattuna natiiviaineistoon kielenoppijat ylikäyttävät hieman OLLA- ja MINÄ-kollokaatteja (esimerkkilauseet 50–51).

(50) Vältän ulos menemistä, jos on lämpi päivä, koska tavallisesti **on** lämpeä *aivan* kestämatön. (ICLFI)

(51) En ole aina halunut opiskella suomen kieltä, se idea oli **minulle** *aivan* odottamaton. (ICLFI)

Ylikäytöstä huolimatta natiivi- ja oppijanaineistot ovat kuitenkin sen verran samanlaisia, että niissä molemmissa vähimmäiskynnyksen yläpuolella on eniten verbikollokaatteja, jotka muodostavat koko tarkastelualueesta Eduskunta-aineistossa 28,94 % ja ICLFI-aineistossa 38,24 %. Toiseksi eniten on kummassakin aineistossa vähimmäiskynnyksen yläpuolella pronominikollokaatteja. Eduskunta-aineistossa nämä muodostavat 10,46 % tarkastelualueesta ja ICLFI-aineistossa 5,88 % tarkastelualueesta.

Eduskunta-aineistossa löytyy vielä konjunktio- (JA, MUTTA, ETTÄ), adverbi- (NYT), adpositio- (MIELESTÄ) ja partikkelikollokaatteja (KYLLÄ). Siis Eduskunta-aineistossa on sekä kollokaattikirjo että erilaisten sanaluokkakollokaattien määrä suurempi kuin ICLFI-aineistossa.

#### 4.2.2.2. *Aivan* oikeanpuoleinen konteksti

Natiiviaineistossa *aivan* saa oikeanpuoleisella tarkastelualueella 24 vähimmäiskynnyksen ylittävää kollokaattia. Nämä kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- <*aivan*>  
ETTÄ/OIKEIN/SELVÄ/LIIAN/SE/ASIA/OIKEA/JA/TÄMÄ/VARMASTI/  
KESKEINEN/OLLA/UUSI/ERI/VÄLTTÄMÄTÖN/TOSI/VARMA/HYVIN/  
KÄSITTÄMÄTÖN/TAPA/MUU/ERINOMAINEN/KUIN/ERITYISESTI.

Oppijanaineistossa on vain 2 vähimmäiskynnyksen ylittävää kollokaattia. Kielenoppijoiden käyttämät kollokaattiorakenteet näyttävät tältä:

- <*aivan*> EI/TAVALLINEN.

Taulukossa 7 esitetään kaikki vähimmäiskynnyksen ylittävät *aivan*-sanin oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %).

Taulukko 7. *Aivan* oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.

Eduskunta			ICLFI		
Kollokaatti (t. alue 36 180)	f	f %	Kollokaatti (t. alue 34)	f	f %
ETTÄ	2 065	5,71 %	EI	2	5,88 %
OIKEIN	1 565	4,33 %	TAVALLINEN	2	5,88 %
SELVÄ	1 091	3,02 %			
LIIAN	1 083	2,99 %			
SE	788	2,18 %			
ASIA	748	2,07 %			
OIKEA	716	1,98 %			
JA	684	1,89 %			
TÄMÄ	564	1,56 %			
VARMASTI	470	1,30 %			
KESKEINEN	459	1,27 %			
OLLA	455	1,26 %			
UUSI	452	1,25 %			
ERI	452	1,25 %			
VÄLTTÄMÄTÖN	422	1,17 %			
TOSI	397	1,10 %			
VARMA	380	1,05 %			
HYVIN	337	0,93 %			
KÄSITTÄMÄTÖN	302	0,83 %			
TAPA	289	0,80 %			
MUU	278	0,77 %			
ERINOMAINEN	274	0,76 %			
KUIN	256	0,71 %			
ERITYISESTI	254	0,70 %			

Natiivi- ja oppijanaineistolle vähimmäiskynnyksen yläpuolella yhteisiä oikeanpuoleisia kollokaatteja ei ole. Kielenoppijoiden yleisemmät kollokaatit EI ja TAVALLINEN (esimerkkilauseet 52–53) eivät ylitä vähimmäiskynnystä natiivianeistossa.

- (52) Pidän tästä, että torilta ostettu asiat ei ole koskaan *aivan* samanlaiset **eivätkä** maistuu samanlaiselta, kuin viime kertana. (ICLFI)
- (53) Omasta mielestäni olen *aivan* **tavallinen** virolainen tyttö. (ICLFI)

Natiivianeistossa *aivan* ei saa tavallisesti kollokaatikseen EI-kieltoverbiä. EI-kollokaatin osuus on pieni myös oikeanpuoleisella tarkastelualueella (1,23 %), mutta vasemmalta

puolelta tämä puuttuu vähimmäiskynnyksen yläpuolella kokonaan. Vaikka Eduskunta-aineistossa *aivan* ei saa kollokaatikseen TAVALLINEN-sanaa, ovat adjektiivikollokaatit *aivan*-sanana oikeanpuoleisella tarkastelualueella kuitenkin yleisempiä sanaluokkakollokaatteja (13,67 %). Pieniä eroja on verbikollokaattien käytössä. Natiiviaineistossa verbikollokaatteja löytyi niukalti (1,26 %), mutta oppijanaaineistossa vähimmäiskynnyksen ylittäneet verbikollokaatit muodostavat tarkastelualueesta 5,88 %.

Eduskunta-aineistossa löytyy vielä adverbi- (OIKEIN, LIIAN, VARMASTI, HYVIN, ERITYISESTI), pronomini- (SE, TÄMÄ, MUU), substantiivi- (ASIA, TAPA) ja konjunktio-kollokaatteja (ETTÄ, JA, KUIN). Eduskunta-aineistossa on sekä kollokaattikirjo että erilaisten sanaluokkakollokaattien määrä suurempi kuin ICLFI-aineistossa.

#### 4.2.2.3. *Erittäin* vasemmanpuoleinen koteksti

Natiiviaineistossa *erittäin* saa vasemmanpuoleisella tarkastelualueella 11 vähimmäiskynnyksen ylittävää kollokaattia. Niistä muodostuneet kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- OLLA/TÄMÄ/PITÄÄ/MINÄ/MIELESTÄ/TEHDÄ/MYÖS/JOKA/  
KANNALTA/KYLLÄ/TULLA <*erittäin*>.

Oppijanaaineistossa *erittäin* vasemmanpuolisia vähimmäiskynnyksen ylittäviä kollokaatteja on 20. Niistä muodostuneet kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- OLLA/PITÄÄ/EI/HÄN/MINÄ/SE/JA/HUONE/ÄITI/OSATA/PALJON/  
SIELLÄ/TYKÄTÄ/KOTI/KEITTIÖ/KESÄ/ENÄÄ/OMMELLA/MIKÄ/  
MIELESTÄ <*erittäin*>.

Taulukossa 8 esitetään kaikki vähimmäiskynnyksen ylittävät *erittäin*-sanana vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %). Natiivi- ja oppijanaaineiston yhteiset kollokaatit on lihavoitu.

Taulukko 8. *Erittäin* vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.

Eduskunta			ICLFI		
Kollokaatti (t. alue 55 842)	f	f %	Kollokaatti (t. alue 262)	f	f %
OLLA	17 816	31,90 %	OLLA	80	30,53 %
TÄMÄ	2 696	4,83 %	PITÄÄ	17	6,49 %
PITÄÄ	1 168	2,09 %	EI	14	5,34 %
MINÄ	1 160	2,08 %	HÄN	11	4,20 %
MIELESTÄ	651	1,17 %	MINÄ	11	4,20 %
TEHDÄ	637	1,14 %	SE	9	3,44 %
MYÖS	570	1,02 %	JA	7	2,67 %
JOKA	514	0,92 %	HUONE	5	1,91 %
KANNALTA	485	0,87 %	ÄITI	3	1,15 %
KYLLÄ	409	0,73 %	OSATA	3	1,15 %
TULLA	398	0,71 %	PALJON	2	0,76 %
			SIELLÄ	2	0,76 %
			TYKÄTÄ	2	0,76 %
			KOTI	2	0,76 %
			KEITTIÖ	2	0,76 %
			KESÄ	2	0,76 %
			ENÄÄ	2	0,76 %
			OMMELLA	2	0,76 %
			MIKÄ	2	0,76 %
			MIELESTÄ	2	0,76 %

Vähimmäiskynnyksen yläpuolella natiivi- ja oppijanaineistolle on 4 yhteistä kollokaattia. Nämä ovat OLLA, PITÄÄ, MINÄ, MIELESTÄ. Siis natiiveille tyypillisistä kollokaateista vain kolmasosa (36,37 %) esiintyy oppijankielessä. Oppijankielestä nämä muodostavat viidesosan (20 %) kollokaattikirjosta ja näin 80 % oppijanaineiston kollokaateista ei ylitä vähimmäiskynnystä natiiviaineistossa.

Yhteisten kollokaattien käyttö on natiivi- ja oppijanaineistossa aika tasaista. Niiden suhteelliset frekvenssit eroavat vain muutaman prosenttiyksikön verran. Aineistojen välillä eroaa PITÄÄ-kollokaatin käyttö, jolla onkin muutamia eri merkityksiä (ks. KTS). Kielenoppijat käyttävät PITÄÄ-kollokaattia vain ilmaisemaan mieltymystä (esimerkkilause 54), mutta Eduskunnan jäsenet pikemminkin muissa tehtävissä,

esimerkiksi ilmaisemaan suorittamista (esimerkkilause 55) tai ilmaisemaan mielipiteitä (esimerkkilause 56):

- (54) Koska **pidän erittäin** paljon valosta ja lämmöstä, on minun mielestäni kaikkein kaunein vuodenaika ehkä kevät. (ICLFI)
- (55) Täällä edustaja Brax **piti erittäin** ansiokkaan puheenvuoron. (Eduskunta)
- (56) **Pidän erittäin** erikoisena ajatuksena että Yleisradiolta tämmöinen viihde laajasti ymmärrettynä kiellettäisiin. (Eduskunta)

Natiivi- ja oppijanaineistot ovat siinä mielessä samanlaisia, että niissä molemmissa esiintyy vähimmäiskynnyksen yläpuolella eniten verbi- (Eduskunta 35,85 % vrt ICLFI 45,04 %) ja toiseksi eniten pronominkollokaatteja (Eduskunta 7,83 % vrt ICLFI 12,60 %), myös konjunktio- (Eduskunta 1,02 % vrt ICLFI 2,67 %) ja adpositiokollokaattien (Eduskunta 2,03 % vrt. ICLFI 0,76 %) suhteelliset frekvenssit ovat tasaisia.

Aineistot eroavat siitä, että vähimmäiskynnyksen yläpuolella lisäksi natiiviaineistossa esiintyy partikkelikollokaatteja (KYLLÄ), mutta oppijanaineistossa sen sijaan kvanttori- (PALJON), substantiivi- (HUONE, ÄITI, KOTI, KEITTIÖ, KESÄ) ja adverbikollokaatteja (SIELLÄ, ENÄÄ). Siis ICLFI-aineistossa on sekä kollokaattikirjo että erilaisten sanaluokkakollokaattien osuus on suurempi kuin Eduskunta-aineistossa.

#### 4.2.2.4. *Erittäin* oikeanpuoleinen konteksti

Natiiviaineistossa *erittäin* saa oikeanpuoleiseksi kollokaatiksi 13 vähimmäiskynnyksen ylittävää kollokaattia. Nämä kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- <*erittäin*> TÄRKEÄ/HYVÄ/ETTÄ/ASIA/SUURI/PALJON/HYVIN/  
MERKITTÄVÄ/TÄMÄ/VAIKEA/OLLA/VAKAVA/ISO.

Oppijanaineistossa *erittäin* oikeanpuoleisia vähimmäiskynnyksen ylittäviä kollokaatteja on 38. Nämä muodostavat seuraavat kollokaattiorakenteet:

- <*erittäin*> PALJON/JA/HYVÄ/MUKAVA/PIENI/HÄN/KIVA/MINÄ/AIKA/  
HYVIN/VIIHTYISÄ/SE/KOSKA/YSTÄVÄ/OLLA/EI/ÄITI/ISO/PERHE/

MUTTA/KAUNIS/TÄRKEÄ/VIISAS/RAKAS/PITKÄ/KYLMÄ/PITÄÄ/  
SISKO/YSTÄVÄLLINEN/TYÖ/TYTTÖ/VAIKEA/USEIN/SUURI/KOIRA/  
KIINNOSTUA/AKTIIVINEN/HERKULLINEN.

Taulukossa 9 esitetään kaikki vähimmäiskynnyksen ylittävät *erittäin*-sanan oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %). Natiivi- ja oppijanaineiston yhteiset kollokaatit on lihavoitu.

Taulukko 9. *Erittäin* oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.

Eduskunta			ICLFI		
Kollokaatti (t. alue 55 842)	f	f %	Kollokaatti (t. alue 262)	f	f %
<b>TÄRKEÄ</b>	5 988	10,72 %	<b>PALJON</b>	25	9,54 %
<b>HYVÄ</b>	4 676	8,37 %	JA	14	5,34 %
ETTÄ	2 896	5,19 %	<b>HYVÄ</b>	10	3,82 %
ASIA	1 903	3,41 %	MUKAVA	9	3,44 %
<b>SUURI</b>	1 449	2,59 %	PIENI	7	2,67 %
<b>PALJON</b>	1 208	2,16 %	HÄN	6	2,29 %
<b>HYVIN</b>	1 013	1,81 %	KIVA	6	2,29 %
MERKITTÄVÄ	960	1,72 %	MINÄ	5	1,91 %
TÄMÄ	937	1,68 %	AIKA	5	1,91 %
<b>VAIKEA</b>	741	1,33 %	<b>HYVIN</b>	5	1,91 %
<b>OLLA</b>	508	0,91 %	VIIHTYISÄ	5	1,91 %
VAKAVA	468	0,84 %	SE	4	1,53 %
<b>ISO</b>	401	0,72 %	KOSKA	4	1,53 %
			YSTÄVÄ	4	1,53 %
			<b>OLLA</b>	3	1,15 %
			EI	3	1,15 %
			ÄITI	3	1,15 %
			<b>ISO</b>	3	1,15 %
			PERHE	3	1,15 %
			MUTTA	3	1,15 %
			KAUNIS	3	1,15 %
			<b>TÄRKEÄ</b>	3	1,15 %
			VIISAS	3	1,15 %
			RAKAS	3	1,15 %
			PITKÄ	3	1,15 %

KYLMÄ	3	1,15 %
PITÄÄ	2	0,76 %
SISKO	2	0,76 %
YSTÄVÄLLINEN	2	0,76 %
TYÖ	2	0,76 %
TYTTÖ	2	0,76 %
<b>VAIKEA</b>	2	0,76 %
USEIN	2	0,76 %
<b>SUURI</b>	2	0,76 %
KOIRA	2	0,76 %
KIINNOSTUA	2	0,76 %
AKTIIVINEN	2	0,76 %
HERKULLINEN	2	0,76 %

Vähimmäiskynnyksen yläpuolella natiivi- ja oppijanaineistoilla on 8 yhteistä kollokaattia. Nämä ovat TÄRKEÄ, HYVÄ, SUURI, PALJON, HYVIN, VAIKEA, OLLA ja ISO. Natiiviaineistossa yhteiset kollokaatit muodostavat 61,54 % koko kollokaatiokirjosta. Oppijanaineistossa käytetään paljon myös muita kollokaatteja ja näin yhteiset kollokaatit muodostavat kaikista kollokaateista vain viidesosan (21,05 %). Suurin osa (78,95 %) oppijanaineiston kollokaateista ei ylitä vähimmäiskynnystä natiiviaineistossa.

Natiivi- ja oppijanaineistossa SUURI-, HYVIN-, VAIKEA-, OLLA- ja ISO-kollokaattien käyttö on suhteellisen tasaista. Muiden kollokaattien valinnassa näkyy enemmän eroja. Kielenoppijat alikäyttävät natiiviaineiston yleisimpiä kollokaatteja TÄRKEÄ (Eduskunta-aineistossa 10,72 % vrt. ICLFI-aineistossa 1,15 %) ja HYVÄ (Eduskunta-aineistossa 8,37 % vrt. ICLFI-aineistossa 3,82 %), mutta ylikäyttävät kvanttorikollokaattia PALJON (Eduskunta-aineistossa 2,16 % vrt. ICLFI-aineistossa 9,54 %), joka on eniten esiintyvä kollokaatti oppijanaineistossa (esimerkkilauseet 57–58).

(57) Veljeni on urheiluhullu, hän pitää *erittäin paljon* jalkapallosta. (ICLFI)

(58) Mutta oikeastaan me emme ole *erittäin paljon* palkintoja voittanut, koska aina menee jokin pieleen. (ICLFI)

Natiivi- ja oppijanaineistot ovat samanlaisia siinä mielessä, että niissä molemmissa käytetään suhteellisen tasaisesti verbi- (Eduskunta-aineistossa 0,91 % vrt. ICLFI-

aineistossa 3,82 % tarkastelualueesta), adverbi- (Eduskunta-aineistossa 1,81 % vrt. ICLFI-aineistossa 2,67 % tarkastelualueesta), adjektiivi- (Eduskunta-aineistossa 26,29 % vrt. ICLFI-aineistossa 25,95 % tarkastelualueesta) ja konjunktio-kollokaatteja (Eduskunta-aineistossa 5,19 % vrt. ICLFI-aineistossa 8,02 % tarkastelualueesta).

Suuremmat erot tulevat näkyviin pronomini- (Eduskunta-aineistossa 1,68 % vrt. ICLFI-aineistossa 5,73 %), substantiivi- (Eduskunta-aineistossa 3,41 % vrt. ICLFI-aineistossa 8,78 %) ja kvanttorikollokaattien (Eduskunta-aineistossa 2,16 % vrt. ICLFI-aineistossa 9,54 %) valinnassa. Kielenoppijat ylikäyttävät kaikkia niitä. Eduskunta- ja ICLFI-aineistossa sanaluokkakollokaatteja on sama määrä, mutta ICLFI-aineiston kollokaattikirjo on suurempi.

#### 4.2.2.5. *Ihan* vasemmanpuoleinen koteksti

Natiiviaineistossa *ihan* saa vasemmanpuoleisella tarkastelualueella 17 vähimmäiskynnyksen ylittävää kollokaattia. Niistä muodostuneet kollokaatorakenteet ovat yleisyysjärjestyksessä:

- OLLA/SE/TÄMÄ/EI/MINÄ/NYT/JA/KYLLÄ/ETTÄ/VOIDA/SITTEN/MIELESTÄ/TULLA/MUTTA/JOKA/ASIA/PITÄÄ <*ihan*>.

Oppijankielessä *ihan* kollokaatteja vähimmäiskynnyksen yläpuolella on 14. Niistä muodostuneet rakenteet näyttävät yleisyysjärjestyksessä tältä:

- OLLA/SE/JA/MINÄ/HUONE/HÄN/PERHE/KEITTIÖ/MYÖS/TUNTUA/KESÄ/PITÄÄ/KAAPPI/ÄÄNTÄÄ <*ihan*>.

Taulukossa 10 esitetään kaikki vähimmäiskynnyksen ylittävät *ihan*-sanin vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %). Natiivi- ja oppijanaineiston yhteiset kollokaatit on lihavoitu.

Taulukko 10. *Ihan* vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.

Eduskunta			ICLFI		
Kollokaatti (t. alue 34794)	f	f %	Kollokaatti (t. alue 170)	f	f %
OLLA	7 844	22,54 %	OLLA	52	30,59 %
SE	2 134	6,13 %	SE	11	6,47 %
TÄMÄ	1 408	4,05 %	JA	7	4,12 %
EI	1 128	3,24 %	MINÄ	7	4,12 %
MINÄ	742	2,13 %	HUONE	6	3,53 %
NYT	724	2,08 %	HÄN	4	2,35 %
JA	653	1,88 %	PERHE	3	1,76 %
KYLLÄ	524	1,51 %	KEITTIÖ	2	1,18 %
ETTÄ	460	1,32 %	MYÖS	2	1,18 %
VOIDA	431	1,24 %	TUNTUA	2	1,18 %
SITTEN	350	1,01 %	KESÄ	2	1,18 %
MIELESTÄ	304	0,87 %	PITÄÄ	2	1,18 %
TULLA	286	0,82 %	KAAPPI	2	1,18 %
MUTTA	282	0,81 %	ÄÄNTÄÄ	2	1,18 %
JOKA	269	0,77 %			
ASIA	264	0,76 %			
PITÄÄ	249	0,72 %			

Natiivi- ja oppijanaineistolle yhteisiä vähimmäiskynnyksen ylittäviä vasemmanpuoleisia kollokaatteja on 5. Nämä ovat OLLA, SE, MINÄ, JA, PITÄÄ. Yhteiset kollokaatit muodostavat natiiviaineiston kollokaattikirjosta 29,41 % ja oppijanaineistossa 35,71 %. Oppijanaineiston kollokaateista 64,29 % ei ylitä vähimmäiskynnystä natiiviaineistossa.

Verrattuna natiiveihin kielenoppijat ylikäyttävät hieman OLLA-kollokaattia (Eduskunta-aineistossa 22,54 % vrt. ICLFI-aineistossa 30,59 %), muiden kollokaattien käyttö on suhteellisen tasaista. PITÄÄ-kollokaatin käytössä on eroja, koska oppijoiden teksteissä se ilmaisee vain mieltymystä (esimerkkilause 59), mutta natiiviaineistossa PITÄÄ-kollokaattia käytetään pikemmin muissa tehtävissä, esimerkiksi ilmaisemaan aikomusta tehdä jotakin (esimerkkilauseet 60–61)

(59) Eniten haluaisin ratsastaa, koska **pidän** *ihan* paljon hevoisesta. (ICLFI)

(60) Kyllä minusta **pitäisi** *ihan* perusteellisesti kyseenalaistaa tämän päätöksen järkevyyttä ylipäänsäkin. (Eduskunta)

(61) Siitä myöskin **pitää** *ihan* reilusti opposition penkiltä kiitokset antaa.  
(Eduskunta)

Natiivi- ja oppijanaaineistot ovat samanlaisia siinä mielessä, että niissä molemmissa esiintyy eniten verbikollokaatteja (Eduskunta-aineistossa 28,56 % vrt. ICLFI-aineistossa 34,12 % tarkastelualueella) ja toiseksi eniten pronominkollokaatteja (Eduskunta-aineistossa 13,09 % vrt. ICLFI-aineistossa 12,94 % tarkastelualueella), myös konjunktio-kollokaattien käyttö on aika tasaista (Eduskunta-aineistossa 4,01 % vrt. ICLFI-aineistossa 5,29 % tarkastelualueella).

Eroja näkyy substantiivikollokaattien käytössä (Eduskunta-aineistossa 0,76 % vrt. ICLFI-aineistossa 8,82 %), joita kielenoppijoiden teksteissä esiintyy vähimmäiskynnyksen yläpuolella enemmän kuin natiiveilla. Lisäksi Eduskunta-aineistossa esiintyy partikkeli- (KYLLÄ), adpositio- (MIELESTÄ) ja adverbikollokaatteja (NYT, SITTEN). Siis Eduskunta-aineistossa on sekä kollokaattikirjo että erilaisten sanaluokkakollokaattien osuus vähimmäiskynnyksen yläpuolella suurempi kuin ICLFI-aineistossa.

#### 4.2.2.6. *Ihan* oikeanpuoleinen koteksti

Natiiviaineistossa *ihan* saa oikeanpuoleisella tarkastelualueella 18 vähimmäiskynnyksen ylittävää kollokaattia. Näistä muodostuneet kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- <*ihan*> HYVÄ/ETTÄ/OIKEASTI/SE/ASIA/OIKEIN/OIKEA/JA/TÄMÄ/  
SELVÄ/OLLA/MUTTA/OMA/MUU/HYVIN/TOSI/ERI/NIIN.

Oppijanaaineistossa vähimmäiskynnyksen ylittäviä kollokaatteja on 29. Yleisyysjärjestyksessä nämä ovat:

- <*ihan*> PALJON/JA/MUTTA/KOSKA/HYVÄ/PERHE/ISO/MUKAVA/OLLA/  
KIVA/SIELLÄ/VIIHTYISÄ/LÄHELLÄ/TAVALLINEN/PIENI/HAUSKA/SE/  
MINÄ/SUURI/KOTI/ME/KORKEA/TYHMÄ/MONI/PAIKKA/MATKA/  
KUMMALLINEN/KAUNIS/IKKUNA.

Taulukossa 11 esitetään kaikki vähimmäiskynnyksen ylittävät *ihan*-sanan oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %). Natiivi- ja oppijanaineiston yhteiset kollokaatit on lihavoitu.

Taulukko 11. *Ihan* oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.

Eduskunta			ICLFI		
Kollokaatti (t. alue 34794)	f	f %	Kollokaatti (t. alue 170)	f	f %
<b>HYVÄ</b>	1 541	4,43 %	PALJON	8	4,71 %
ETTÄ	1 257	3,61 %	<b>JA</b>	7	4,12 %
OIKEASTI	861	2,47 %	<b>MUTTA</b>	6	3,53 %
<b>SE</b>	860	2,47 %	KOSKA	5	2,94 %
ASIA	672	1,93 %	<b>HYVÄ</b>	5	2,94 %
OIKEIN	657	1,89 %	PERHE	4	2,35 %
OIKEA	628	1,80 %	ISO	4	2,35 %
<b>JA</b>	623	1,79 %	MUKAVA	4	2,35 %
TÄMÄ	616	1,77 %	<b>OLLA</b>	3	1,76 %
SELVÄ	582	1,67 %	KIVA	3	1,76 %
<b>OLLA</b>	484	1,39 %	SIELLÄ	3	1,76 %
<b>MUTTA</b>	372	1,07 %	VIIHTYISÄ	3	1,76 %
OMA	350	1,01 %	LÄHELLÄ	3	1,76 %
MUU	329	0,95 %	TAVALLINEN	3	1,76 %
HYVIN	326	0,94 %	PIENI	3	1,76 %
TOSI	299	0,86 %	HAUSKA	3	1,76 %
ERI	256	0,74 %	<b>SE</b>	2	1,18 %
NIIN	248	0,71 %	MINÄ	2	1,18 %
			SUURI	2	1,18 %
			KOTI	2	1,18 %
			ME	2	1,18 %
			KORKEA	2	1,18 %
			TYHMÄ	2	1,18 %
			MONI	2	1,18 %
			PAIKKA	2	1,18 %
			MATKA	2	1,18 %
			KUMMALLINEN	2	1,18 %
			KAUNIS	2	1,18 %
			IKKUNA	2	1,18 %

Vähimmäiskynnyksen yläpuolella natiivi- ja oppijanaineistolla on 5 yhteistä kollokaattia. Nämä ovat HYVÄ, SE, JA, OLLA, MUTTA. Natiiviaineistossa yhteiset kollokaatit muodostavat 27,78 % kollokaattikirjosta, oppijanaineistossa 17,24 %. Siis oppijanaineiston kollokaateista 82,76 % ei ylitä vähimmäiskynnystä natiiviaineistossa. Yhteisiä kollokaatteja käytetään suhteellisen tasaisesti.

Natiivi- ja oppijanaineistot ovat siinä mielessä samanlaisia, että niissä molemmissa käytetään suhteellisen tasaisesti verbi- (Eduskunta-aineistossa 1,39 % vrt. ICLFI-aineistossa 1,76 % tarkastelualueesta), pronomini- (Eduskunta-aineistossa 5,19 % vrt. ICLFI-aineistossa 3,35 % tarkastelualueesta) ja adverbikollokaatteja (Eduskunta-aineistossa 6,01 % vrt. ICLFI-aineistossa 3,53 % tarkastelualueesta).

Eroja on siinä, että kielenoppijat ylikäyttävät konjunktio- (Eduskunta-aineistossa 6,47 % vrt. ICLFI-aineistossa 10,59 %), substantiivi- (Eduskunta-aineistossa 1,93 % vrt. ICLFI-aineistossa 7,06 %) ja adjektiivikollokaatteja (Eduskunta-aineistossa 10,51 % vrt. ICLFI-aineistossa 22,35 %). Lisäksi oppijanaineistossa esiintyy kvanttorikollokaatteja (PALJON, MONI), joista PALJON on myös oppijanaineiston yleisin kollokaatti (esimerkkilauseet 62–63).

(62) Pidän *ihan paljon* jazzista ja vanheista pop-bändeista. (ICLFI)

(63) Keittiön ja olohuoneen välissä ja ruokailuhuoneen ja olohuoneen välissä ei ole seinää, joten niissä kaikissa on *ihan paljon* valoa. (ICLFI)

Siis ICLFI-aineistossa on sekä kollokaattikirjo että erilaisten sanaluokkakollokaattien osuus suurempi kuin Eduskunta-aineistossa.

#### 4.2.2.7. *Kovin* vasemmanpuoleinen konteksti

Natiiviaineistossa *kovin* saa vasemmanpuoleiselle tarkastelualueelle 14 vähimmäiskynnyksen ylittävää kollokaattia. Näistä muodostuneet kollokaatorakenteet näyttävät yleisyysjärjestyksessä tältä:

- OLLA/EI/SE/TÄMÄ/NIIN/ETTÄ/KYLLÄ/NYT/MINÄ/VOIDA/MUTTA/TUNTUA/JA/TULLA <*aivan*>.

Oppijanaineiston pienuudesta johtuen vähimmäiskynnyksen (ja vähintään kahden esiintymän rajan) ylitti vain 3 kollokaattia. Kielenoppijoiden käyttämät kollokaattiorakenteet näyttävät yleisyysjärjestyksessä tältä:

- OLLA/EI/PITÄÄ <*kovin*>.

Taulukossa 12 esitetään kaikki vähimmäiskynnyksen ylittävät *kovin*-sanon vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %). Natiivi- ja oppijanaineiston yhteiset kollokaatit on lihavoitu.

Taulukko 12. *Kovin* vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.

Eduskunta			ICLFI		
Kollokaatti (t. alue 9096)	f	f %	Kollokaatti (t. alue 38)	f	f %
<b>OLLA</b>	2 120	23,31 %	<b>OLLA</b>	14	36,84 %
<b>EI</b>	1 409	15,49 %	<b>EI</b>	7	18,42 %
SE	368	4,05 %	PITÄÄ	2	5,26 %
TÄMÄ	220	2,42 %			
NIIN	165	1,81 %			
ETTÄ	145	1,59 %			
KYLLÄ	123	1,35 %			
NYT	101	1,11 %			
MINÄ	86	0,95 %			
VOIDA	84	0,92 %			
MUTTA	75	0,82 %			
TUNTUA	73	0,80 %			
JA	70	0,77 %			
TULLA	70	0,77 %			

Vähimmäiskynnyksen yläpuolella natiivi- ja oppijanaineistolle yhteisiä vasemmanpuoleisia kollokaatteja on 2. Nämä ovat OLLA ja EI, jotka ovat yleisimmät kollokaatit kummassakin aineistossa (esimerkkilauseet 64–67). Ne muodostavat natiiviaineiston kollokaattikirjosta 14,29 % ja oppijanaineistossa 66,67 %. Siis 33,33 % kielenoppijoiden kollokaateista ei ylitä vähimmäiskynnystä natiiviaineistossa. Verrattuna

natiiviaineistoon kielenoppijat ylikäyttävät OLLA-kollokaattia (Eduskunta-aineistossa 23,31 % vrt. ICLFI-aineistossa 36,84 % tarkastelualueesta), mutta EI-kollokaattia käytetään suhteellisen tasaisesti.

- (64) Kun on lämmin ilma ja aurinko paistaa, **on** *kovin* kiva ottaa polkupyörä ja kori sekä ajella Tarton keskustan torille. (ICLFI)
- (65) Näen että **olen** *kovin* yksinäinen tässä puhujalistalla tällä hetkellä. (Eduskunta)
- (66) Myöskään nimipäivät **eivät ole** *kovin* tärkeitä, niiden juhlimista ei juuri ole. (ICLFI)
- (67) Tämä on tietysti tällainen tulevaisuuden toive ja välttämättä se **ei** *kovin* nopeasti toteudu. (Eduskunta)

Natiivi- ja oppijanaaineisto ovat siinä mielessä samanlaisia, että niiden kummankin vasemmanpuoleisella tarkastelualueella esiintyy eniten verbejä (Eduskunta-aineistossa 41,29 % vrt. ICLFI-aineistossa 61 %), tosin ICLFI-aineistossa on niitä huomattavasti enemmän kuin Eduskunta-aineistossa.

Aineistot eroavat keskenään siitä, että ICLFI-aineistossa ei olekaan muita kuin verbikollokaatteja, mutta Eduskunta-aineistossa on lisäksi myös pronomini- (SE, TÄMÄ, MINÄ), konjunktio- (ETTÄ, MUTTA, JA), adverbi- (NIIN, NYT) ja partikkelikollokaatteja (KYLLÄ). Siis Eduskunta-aineistossa on sekä kollokaattikirjo että erilaisten sanaluokkakollokaattien määrä suurempi kuin ICLFI-aineistossa.

#### 4.2.2.8. *Kovin* oikeanpuoleinen konteksti

Natiiviaineistossa *kovin* saa oikeanpuoleisella tarkastelualueella 18 vähimmäiskynnyksen ylittävää kollokaattia. Niistä muodostuneet kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- <*kovin*>PALJON/SUURI/OLLA/JA/SE/TÄMÄ/HYVÄ/MUTTA/HELPPU/EI/VAIKEA/HYVIN/ERILAINEN/PALJO/VÄHÄN/PITKÄ/ASIA/ETTÄ.

Oppijanaineistossa enemmän kuin kerran esiintyviä kollokaatteja on 4. Kielenoppijoiden käyttämät kollokaattiorakenteet näyttävät yleisyysjärjestyksessä tältä:

- <kovin> PALJON/SE/KAUKANA/KESKUSTA.

Taulukossa 13 esitetään kaikki vähimmäiskynnyksen ylittävät *kovin*-sanana oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %). Natiivi- ja oppijanaineiston yhteiset kollokaatit on lihavoitu.

Taulukko 13. *Kovin* oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.

Eduskunta			ICLFI		
Kollokaatti (t. alue 9096)	f	f %	Kollokaatti (t. alue 38)	f	f %
<b>PALJON</b>	736	8,09 %	<b>PALJON</b>	4	10,53 %
SUURI	384	4,22 %	<b>SE</b>	2	5,26 %
OLLA	244	2,68 %	KAUKANA	2	5,26 %
JA	233	2,56 %	KESKUSTA	2	5,26 %
<b>SE</b>	223	2,45 %			
TÄMÄ	210	2,31 %			
HYVÄ	206	2,26 %			
MUTTA	134	1,47 %			
HELPPO	126	1,39 %			
EI	107	1,18 %			
VAIKEA	92	1,01 %			
HYVIN	90	0,99 %			
ERILAINEN	82	0,90 %			
PALJO	80	0,88 %			
VÄHÄN	79	0,87 %			
PITKÄ	76	0,84 %			
ASIA	74	0,81 %			
ETTÄ	70	0,77 %			

Natiivi- ja oppijanaineistolle yhteisiä vähimmäiskynnyksen ylittäviä oikeanpuoleisia kollokaatteja on kaksi. Nämä ovat PALJON ja SE. Yhteiset kollokaatit muodostavat natiiviaineiston kollokaattikirjosta 11,11 % ja oppijanaineiston kollokaattikirjosta 50 %.

Näin oppijanaineiston kollokaateista puolet eivät ylitä vähimmäiskynnystä natiiviaineistossa.

Natiiviaineiston kolligaatioanalyysistä kävi ilmi, että *kovin* määrittää tasaisesti sekä adjektiiveja että kvanttoreita, mutta adverbejä suhteellisen vähän. Oppijanaineistosta näkyy, että kvanttorikollokaatit ovat sielläkin tavallisia (esimerkkilause 68), mutta vähimmäiskynnyn yläpuolella ei esiinny adjektiiveja. Tämä tulos johtuu aineiston pienuudesta; aineistossa kyllä esiintyy myös adjektiiveja oikeanpuoleisessa kontekstissa (ks. esimerkkilauseet 64, 66).

(68) Hän ei pitää *kovin* **paljon** ruuanlaitosta, mutta hänen täytyy tehdä sen.  
(ICLFI)

Natiivi- ja oppijanaineisto ovat siinä miehestä samanlaisia, että niissä käytetään kvanttori- (Eduskunta-aineistossa 9,84 % vrt. ICLFI-aineistossa 10,53 % tarkastelualueesta) ja pronominikollokaatteja (Eduskunta-aineistossa 4,76 % vrt. ICLFI-aineistossa 5,26 % tarkastelualueesta) aika tasaisesti. Eroja näkyy siinä, että kielenoppijat ylikäyttävät hieman adverbi- (Eduskunta-aineistossa 0,99 % vrt. ICLFI-aineistossa 5,26 % tarkastelualueesta) ja substantiivikollokaatioita (Eduskunta-aineistossa 0,81 % vrt. ICLFI-aineistossa 5,26 % tarkastelualueesta). Lisäksi Eduskunta-aineistossa esiintyy verbi- (OLLA, EI), konjunktio- (JA, MUTTA, ETTÄ)- ja adjektiivikollokaatteja (SUURI, HYVÄ, HELPPO, VAIKEA, ERILAINEN, PITKÄ). Siis Eduskunta-aineistossa on sekä kollokaattikirjo että erilaisten sanaluokkakollokaattien määrä suurempi kuin ICLFI-aineistossa.

#### 4.2.2.9. *Oikein* vasemmanpuoleinen konteksti

Natiiviaineistossa *kovin* saa vasemmanpuoleisella tarkastelualueella 20 vähimmäiskynnyn ylittävää kollokaattia, jotka muodostavat yleisyysjärjestyksessä seuraavat kollokaattiorakenteet:

- OLLA/EI/SE/TÄMÄ/MINÄ/JA/NYT/JOS/YMMÄRTÄÄ/KYLLÄ/ETTÄ/  
ASIA/PITÄÄ/KUN/MIELI/TEHDÄ/TULLA/VOIDA/MENNÄ/SAADA  
<*oikein*>.

Oppijanaineistossa vähimmäiskynnyksen ylittäviä vasemmanpuoleisia kollokaatteja on 17. Niistä muodostuneet kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- OLLA/SE/PITÄÄ/JA/MINÄ/ME/HUONE/HÄN/MUTTA/SIELLÄ/JOKA/  
PERHE/MYÖS/PUHUA/TEHDÄ/LOMA/TÄMÄ <*oikein*>.

Taulukossa 14 esitetään kaikki vähimmäiskynnyksen ylittävät *oikein*-sanan vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %). Natiivi- ja oppijanaineiston yhteiset kollokaatit on lihavoitu.

Taulukko 14. *Oikein* vasemmanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.

Eduskunta			ICLFI		
Kollokaatti (t. alue 3492)	f	f %	Kollokaatti (t. alue 554)	f	f %
<b>OLLA</b>	618	17,70 %	<b>OLLA</b>	165	29,78 %
EI	251	7,19 %	<b>SE</b>	27	4,87 %
<b>SE</b>	149	4,27 %	<b>PITÄÄ</b>	22	3,97 %
<b>TÄMÄ</b>	138	3,95 %	<b>JA</b>	19	3,43 %
<b>MINÄ</b>	93	2,66 %	<b>MINÄ</b>	17	3,07 %
<b>JA</b>	72	2,06 %	ME	17	3,07 %
NYT	59	1,69 %	HUONE	11	1,99 %
JOS	54	1,55 %	HÄN	9	1,62 %
YMMÄRTÄÄ	46	1,32 %	MUTTA	8	1,44 %
KYLLÄ	41	1,17 %	SIELLÄ	7	1,26 %
ETTÄ	35	1,00 %	JOKA	5	0,90 %
ASIA	34	0,97 %	PERHE	4	0,72 %
<b>PITÄÄ</b>	34	0,97 %	MYÖS	4	0,72 %
KUN	32	0,92 %	PUHUA	4	0,72 %
MIELESTÄ	30	0,86 %	<b>TEHDÄ</b>	4	0,72 %
<b>TEHDÄ</b>	30	0,86 %	LOMA	4	0,72 %
TULLA	29	0,83 %	<b>TÄMÄ</b>	4	0,72 %
VOIDA	27	0,77 %			
MENNÄ	25	0,72 %			
SAADA	25	0,72 %			

Vähimmäiskynnyksen yläpuolella natiivi- ja oppijanaineistoille yhteisiä vasemmanpuoleisia kollokaatteja on 7. Nämä ovat: OLLA, SE, TÄMÄ, MINÄ, JA, PITÄÄ, TEHDÄ. Natiiviaineiston kollokaatiokirjosta nämä muodostavat 35 %, oppijanaineistosta 41,18 %. Näin 58,82 % oppijanaineiston kollokaateista ei ylitä vähimmäiskynnystä natiiviaineistossa. Kielenoppijat ylikäyttävät OLLA-kollokaattia (Eduskunta-aineistossa 17,70 % vrt. ICLFI-aineistossa 29,78 %), mutta muiden kollokaattien käyttö on frekvenssien perusteella melko samanlaista.

Eroja on PITÄÄ-kollokaatin käytössä, koska oppijoiden teksteissä se ilmaisee yleensä mieltymystä (esimerkkilause 69) ja on harvemmin muissa tehtävissä kuten ilmaisemassa pakkoa (esimerkkilause 70), mutta natiiviaineistossa PITÄÄ käytetään pikemmin muissa tehtävissä kuin mieltymyksen ilmaisemisessa, esimerkiksi ilmaisemassa mielipiteitä (esimerkkilause 71):

- (69) **Pidän oikein** paljon keväästä, koska silloin linnut laulavat ja puihin tulee vähitellen vihreitä lehtiä. (ICLFI)
- (70) Viime vuonna minun **piti** olla *oikein* kauan Tartossa koska tenttikausi oli kesäkuussa. (ICLFI)
- (71) **Pidin** tätä *oikein* hyvänä. (Eduskunta)

Natiivi- ja oppijanaineisto ovat siinä mielessä samanlaisia, että niissä molemmissa käytetään aika tasaisesti verbi- (Eduskunta-aineistossa 31,07 % vrt. ICLFI-aineistossa 35,20 % tarkastelualueesta), pronomini- (Eduskunta-aineistossa 10,88 % vrt. ICLFI-aineistossa 14,26 % tarkastelualueesta), konjunktio- (Eduskunta-aineistossa 5,53 % vrt. ICLFI-aineistossa 5,60 % tarkastelualueesta), adverbi- (Eduskunta-aineistossa 1,69 % vrt. ICLFI-aineistossa 1,26 % tarkastelualueesta) ja substantiivikollokaatteja (Eduskunta-aineistossa 0,97 % vrt. ICLFI-aineistossa 3,43 % tarkastelualueesta). Lisäksi Eduskunta-aineistossa esiintyy adpositio- (MIELESTÄ) ja partikkelikollokaatteja (KYLLÄ). Siis Eduskunta-aineistossa on sekä kollokaattikirjo että erilaisten sanaluokkakollokaattien määrä suurempi kuin ICLFI-aineistossa.

#### 4.2.2.10. *Oikein* oikeanpuoleinen koteksti

Natiiviaineistossa *oikein* saa oikeanpuoleiselle tarkastelualueelle 12 vähimmäiskynnyksen ylittävää kollokaattia. Niistä muodostuneet kollokaatorakenteet ovat yleisyysjärjestyksessä:

- *<oikein>* HYVÄ, HYVIN, JA, ETTÄ, SE, TÄMÄ, KUNNOLLA, OLLA, PALJON, ASIA, KUN, KUNNON.

Oppijanaineistossa *oikein* saa oikeanpuoleiseksi kollokaatiksi 25 vähimmäiskynnyksen ylittävää kollokaattia. Niistä muodostuneet kollokaatorakenteet ovat yleisyysjärjestyksessä:

- *<oikein>* JA/PALJON/HYVÄ/KAUNIS/MUKAVA/HYVIN/SE/MINÄ/OLLA/PIENI/VIIHTYISÄ/VAIKEA/KOSKA/HAUSKA/KIVA/TÄRKEÄ/PITÄÄ/HÄN/MUTTA/ISO/LUONTO/PERHE/ONNELLINEN/PAIKKA/VIRO.

Taulukossa 15 esitetään kaikki vähimmäiskynnyksen ylittävät *oikein*-sanon oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %). Natiivi- ja oppijanaineiston yhteiset kollokaatit on lihavoitu.

Taulukko 15. *Oikein oikeanpuoleiset kollokaatit Eduskunta- ja ICLFI-aineistossa.*

Eduskunta			ICLFI		
Kollokaatti (t. alue 3492)	f	f %	Kollokaatti (t. alue 554)	f	f %
<b>HYVÄ</b>	411	11,77 %	<b>JA</b>	58	10,47 %
<b>HYVIN</b>	232	6,64 %	<b>PALJON</b>	53	9,57 %
<b>JA</b>	122	3,49 %	<b>HYVÄ</b>	21	3,79 %
ETTÄ	113	3,24 %	KAUNIS	20	3,61 %
<b>SE</b>	101	2,89 %	MUKAVA	17	3,07 %
TÄMÄ	80	2,29 %	<b>HYVIN</b>	13	2,35 %
KUNNOLLA	74	2,12 %	<b>SE</b>	11	1,99 %
<b>OLLA</b>	46	1,32 %	MINÄ	10	1,81 %
<b>PALJON</b>	44	1,26 %	<b>OLLA</b>	9	1,62 %
ASIA	39	1,12 %	PIENI	9	1,62 %
KUN	30	0,86 %	VIIHTYISÄ	8	1,44 %
KUNNON	27	0,77 %	VAIKEA	7	1,26 %
			KOSKA	6	1,08 %
			HAUSKA	6	1,08 %
			KIVA	6	1,08 %
			TÄRKEÄ	6	1,08 %
			PITÄÄ	5	0,90 %
			HÄN	5	0,90 %
			MUTTA	5	0,90 %
			ISO	5	0,90 %
			LUONTO	5	0,90 %
			PERHE	4	0,72 %
			ONNELLINEN	4	0,72 %
			PAIKKA	4	0,72 %
			VIRO	4	0,72 %

Natiivi- ja oppijanaineistolla on 6 yhteisiä kollokaattia. Nämä ovat: HYVÄ, HYVIN (esimerkkilauseet 72–75), JA, SE, OLLA, PALJON. Yhteiset kollokaatit muodostavat natiiviaineiston kollokaattikirjosta puolen (50 %), mutta oppijanaineistossa noin neljäsosan (24 %). Siis kielenoppijoiden kollokaattikirjosta 76 % ei yllä vähimmäiskynnystä natiiviaineistossa.

Kielenoppijat alikäyttävät HYVÄ- (Eduskunta-aineistossa 11,77 % vrt. ICLFI-aineistossa 3,79 % tarkastelualueesta) ja HYVIN-kollokaatteja (Eduskunta-aineistossa

6,64 % vrt. ICLFI-aineistossa 2,35 % tarkastelualueesta) ja ylikäyttävät JA- (Eduskunta-aineistossa 3,49 % vrt. ICLFI-aineistossa 10,47 % tarkastelualueesta) ja PALJON-kollokaatteja (Eduskunta-aineistossa 1,26 % vrt. ICLFI-aineistossa 9,57 % tarkastelualueesta). SE- ja OLLA-kollokaattien käyttö on suhteellisen tasaista.

- (72) Se oli mielenkiintoista työtä ja *oikein hyvä* kokemus. (ICLFI)
- (73) Kollegoille ja puhemiehelle *oikein hyvää* rauhallista joulua. (Eduskunta)
- (74) Puhun *oikein hyvin* viroa ja englantia, vähän venäjää. (ICLFI)
- (75) Vuoden 91 vaalit muistan *oikein hyvin* tässä suhteessa. (Eduskunta)

Natiivi- ja oppijaaineisto ovat siinä mielessä samanlaisia, että niissä kummassakin käytetään aika tasaisesti verbi- (Eduskunta-aineistossa 1,32 % vrt. ICLFI-aineistossa 2,53 % tarkastelualueesta), pronomini- (Eduskunta-aineistossa 5,18 % vrt. ICLFI-aineistossa 4,69 % tarkastelualueesta) ja substantiivikollokaatteja (Eduskunta-aineistossa 1,12 % vrt. ICLFI-aineistossa 3,07 % tarkastelualueesta). Eroja puolestaan on siinä, että kielenoppijat ylikäyttävät konjunktio-kollokaatteja (Eduskunta-aineistossa 7,59 % vrt. ICLFI-aineistossa 12,45 % tarkastelualueesta). Suuret erot ovat myös kvanttori- (Eduskunta-aineistossa 1,26 % vrt. ICLFI-aineistossa 9,57 % tarkastelualueesta), adverbii- (Eduskunta-aineistossa 8,76 % vrt. ICLFI-aineistossa 2,35 % tarkastelualueesta) ja adjektiivikollokaattien (Eduskunta-aineistossa 12,54 % vrt. ICLFI-aineistossa 19,68 % tarkastelualueesta) käytössä. Kaikkiaan oppijanaineiston kollokaattikirjo on suurempi kuin natiiviaineistossa.

#### 4.2.2.11 *Sangen* vasemmanpuoleinen koteksti

Natiiviaineistossa *sangen*-sanalla on 12 vähimmäiskynnyksen ylittävää kollokaattia. Nämä muodostavat noodin kanssa seuraavat kollokaattiorakenteet:

- OLLA/TÄMÄ/SE/JO/JA/ESITYS/MINÄ/JOKA/HALLITUS/VIELÄ/TIETÄÄ/  
KYLLÄ <*sangen*>.

Oppijanaineistossa *sangen*-sanaa ei esiinny.

Taulukossa 16 esitetään kaikki vähimmäiskynnyksen ylittävät *sangen*-sanana vasemmanpuoleiset kollokaatit Eduskunta-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %).

Taulukko 16. *Sangen* vasemmanpuoleiset kollokaatit Eduskunta-aineistossa.

Kollokaatti (t. alue 282)	f	f %
OLLA	66	23,40 %
TÄMÄ	8	2,84 %
SE	7	2,48 %
JO	6	2,13 %
JA	4	1,42 %
ESITYS	4	1,42 %
MINÄ	4	1,42 %
JOKA	3	1,06 %
HALLITUS	2	0,71 %
VIELÄ	2	0,71 %
TIETÄÄ	2	0,71 %
KYLLÄ	2	0,71 %

#### 4.2.2.12. *Sangen* oikeanpuoleinen konteksti

Natiiviaineistossa *sangen*-sanalla on 21 kollokaattia. Niistä muodostuneet kollokaattiorakenteet ovat yleisyysjärjestyksessä:

- <*sangen*> JA/TÄMÄ/HYVIN/VÄHÄN/ETTÄ/YKSIMIELINEN/SE/  
MIELENKIINTOINEN/SUURESTI/SUURI/HYVÄ/RATKAISU/LYHYT/  
OLLA/HALLITUS/TUO/SIKSI/ASIA/NIIN/MUTTA.

Taulukossa 17 esitetään kaikki *sangen*-sanana vähimmäiskynnyksen ylittävät oikeanpuoleiset kollokaatit Eduskunta-aineistossa, niiden todelliset frekvenssit (f) ja suhteelliset osuudet tarkastelualueella (f %).

Taulukko 17. *Sangen oikeanpuoleiset kollokaatit Eduskunta-aineistossa.*

Kollokaatti (t. alue 282)	f	f %
JA	17	6,03 %
TÄMÄ	8	2,84 %
HYVIN	5	1,77 %
VÄHÄN	5	1,77 %
ETTÄ	4	1,42 %
YKSIMIELINEN	4	1,42 %
SE	3	1,06 %
MIELENKIINTOINEN	3	1,06 %
SUURESTI	3	1,06 %
SUURI	3	1,06 %
HYVÄ	3	1,06 %
RATKAISU	3	1,06 %
LYHYT	3	1,06 %
OLLA	2	0,71 %
ESITYS	2	0,71 %
HALLITUS	2	0,71 %
TUO	2	0,71 %
SIKSI	2	0,71 %
ASIA	2	0,71 %
NIIN	2	0,71 %
MUTTA	2	0,71 %

### 4.3. Pohdintaa

Usein astemääritteitä käsitellään synonyymisinä sanoina. Onhan totta, että niiden tehtävät ovat samanlaisia ja myös käyttöympäristöt limittyvät osittain, mutta yksityiskohtaisempi analyysi paljastaa, että yleensä ne eivät ole keskenään vaihdettavissa. Kaikille astemääritteille yhteisiä kollokaatteja on niukalti. Jokaisella astemääritteellä on omat erityispiirteensä ja kahta sellaista astemääritettä, joiden käyttö olisi identtistä, ei löydy. Eroja näkyi melkein kaikilla osa-alueilla, joita tarkasteltiin. Nyt kiinnitetään huomiota

vain eniten esiintyviin piirteisiin, koska lähdetään oletuksesta, että enemmän esiintyvä piirre on analyysin kannalta oleellisempi.

Edellinen luku (4.2) tarjosi runsaasti esimerkkejä siitä, että astemääritteet eivät esiinny kielessä satunnaisesti, vaan niillä on taipumuksia esiintyä tietynlaisessa kontekstissa. Astemääritteiden vasemmanpuoleisella tarkastelualueella esiintyy ensi sijassa verbi- ja pronominikolligaatteja. Nämä kattavat vähintään kolmasosan (*sangen*) tai jopa melkein puolet astemääritteen (*kovin*) V2–V1-position tarkastelualueesta.

Kollokaatioanalyysistä puolestaan ilmeni, että ylivoimaisesti eniten käytetty verbi kaikkien astemääritteiden vasemmanpuoleisella tarkastelualueella on suomen kielen yleisin sana OLLA. Lisäksi esiintyi muitakin verbikollokaatteja, mutta suurin osa niistäkin kuuluu suomen kielen yleisimpiin sanoihin. Niiden joukossa on myös EI, jonka esiintyminen erottaa astemääritteet selvästi toisistaan. (ks. Saukkonen ym. 1979)

EI on tavallisin *kovin*- (37,3 % kaikista verbikollokaateista) ja *oikein*- (23,13 % kaikista verbikollokaateista) tarkastelualueella. Vähemmän sitä esiintyy *ihan*- (11,35 % kaikista verbikollokaateista) ja *aivan*- (4,25 % kaikista verbikollokaateista) tarkastelualueella. *Erittäin*- ja *sangen*- tarkastelualueella EI-kollokaattia ei käytetä niin paljon, että se ylittäisi vähimmäiskynnyksen. Toisin kuin tässä, Jantunen (2004: 97) nostaa esiin, että hänen aineistossaan *oikein*-määritteellä on erittäin voimakas negatiivinen preferenssi kieltohakuisuuteen. Hänen aineistossaan *oikein*-sanan tarkastelualueella (V3–O2) EI-kieltoverbi muodostaa kaikista verbikollokaateista vain 5 % (mts. 90).

Astemääritteiden vasemmanpuoleinen tarkastelualue on suhteellisen homogeeninen, koska verbien osalta siellä esiintyvät tietyt yleiset verbit ja pronomineja taas on yleisesti niin vähän, että kirjoja ei tule paljon. Pronominikollokaatteja on vähimmäiskynnyksen yläpuolella vain neljä ja kaksi niistä (TÄMÄ, MINÄ) ovat kaikille astemääritteille yhteisiä. Pronominikolligaattien runsas ja substantiivikolligaattien niukka käyttö antavat olettaa, että astemääritteiden vasemmanpuoleisessa kontekstissa tarkoitteisiin viitataan pikemmin pronomineilla. Pronomineilla tarkoitteisiin viittaminen tuntuu johdonmukaiselle, koska astemääritteitä käytetään henkilökohtaisen suhtautumisen ilmaisemiseen ja todennäköisesti aihe, johon otetaan kantaa, on ollutkin puheena jo

aiemmin. Sen tarkisteleminen vaatisi kuitenkin erillistä tutkimusta, koska tässä tutkimuksessa suuri osa substantiivikollikaatteja eivät yltäneet vähimmäiskynnyksen yläpuolelle.

Määritelmän mukaan astemääritteet toimivat adjektiivien, adverbien, kvanttoreiden ja adpositioiden/adpositiolausekkeiden etumääritteinä. Oikeanpuoleisen tarkastelualueen kolligaatioanalyysistä selvisi, että astemääritteet eroavat keskenään nimenomaan siitä, millaisia sanaluokkia ne määrittävät. Esiin tuli, että aivan määrittää aika tasaisesti sekä adjektiiveja (13,67 %) että adverbejä (10,25 %), mutta ei ollenkaan kvanttoreita. Erittäin määrittää pääasiassa adjektiiveja (26,29 %), samalla kun kvanttorien (2,16 %) ja adverbien (1,81 %) osuudet ovat suhteellisen pieniä. Ihan määrittää tyypillisesti adjektiiveja (10,51 %) ja vähemmän adverbejä (6,01 %), mutta ei ollenkaan kvanttoreita. Kovin määrittää tasaisesti sekä adjektiiveja (10,62 %) että kvanttoreita (9,84 %), mutta adverbejä suhteellisen vähän (0,99 %). Oikein toimii enimmäkseen adjektiivien (12,54 %) ja adverbien (8,76 %) etumääritteenä, samalla kun kvanttoreiden osa on marginaalinen (1,26 %). Sangen määrittää muita tasaisemmin adjektiiveja (5,67 %), adverbejä (4,26 %), mutta myös kvanttoreita (1,77 %), vaikka niiden osuus on hieman vähäisempi.

Edellinen havainto on siinä mielessä kiinnostava, että Jantusen (2004: 113) analyysistä ilmeni, että sanaluokkakolligaattien keskeinen järjestys ei erota astemääritteitä toisistaan vaan on kaikilla sama (yleisyysjärjestyksessä adjektiivit, adverbit ja kvanttorit), vaikkakin prosentuaalisissa osuuksissa näkyy myös hänen aineistossaan huomattavia eroja. Tulokset ovat samanlaisia siinä mielessä, että Jantunen (mts. 113) toteaa adpositiokolligaattien osuuden olevan häviävän pieni (alle 2 %). Tässä tutkimuksessa adpositiokolligaatit eivät edes yltäneet vähimmäiskynnyksen yläpuolelle. Tutkimusten välisiin eroihin ilmeisesti johtaa toisaalta se, että Jantunen (2004) tarkasteli vain *hyvin-*, *kovin-* ja *oikein-*astemääritteitä, toisaalta se, että tutkimuksissa käytettiin erilaisia korpuksia. Jantusen natiiviaineiston tarkastelu perustui kaunokirjallisuuteen ja tämän tutkimuksen natiiviaineistona käytettiin Eduskunnan täysistuntoja. (ks. mts. 85, 227). Lisäksi analyysimenetelmä voi vaikuttaa, koska tässä tutkimuksessa korpusten kokoeron

vuoksi vähimmäiskynnys laitettiin melko korkealle, mikä jätti suhteellisen paljon materiaalia myös analyysin ulkopuolelle.

Kollokaatioanalyysistä ilmeni, että adjektiiveihin, adverbeihin ja kvanttoreihin kuuluvien kollokaattien kirjo on laaja ja sellaisia adjektiivi- ja kvanttorikollokaatteja, jotka esiintyisivät kaikkien astemääritteiden oikeanpuoleisella tarkastelualueella, ei löytynyt. Ainoa kaikille astemääritteille yhteinen määritettävä sana oli adverbikollokaatti HYVIN. Osa havaitusta variaatiosta voi johtua aineistosta, mutta osa myös laajemmin sanojen preferensseistä ja rajoituksista. Esimerkiksi Kielitoimiston sanakirja (KTS) nostaa esiin, että astemääritettä *aivan* ei käytetä sanojen *suuri* ja *raskas* määrittämiseen. Sellaisia yleistyksiä sanojen käyttöä koskevista rajoituksista ei voi kuitenkaan tehdä vain yhden korpuksen perusteella.

Natiivi- ja oppijanaineiston vertailusta kävi ilmi, että suomalaiset ja virolaiset B1–B2-tason suomen kielen oppijat käyttävät astemääritteitä melko samankaltaisesti. Natiivien aineistossa astemääritteitä on 31,1 esiintymää 10 000 sanetta kohti ja kielenoppijoilla 28,7 sanetta 10 000 saneeseen verrattuna. Samanlaisten frekvenssien taustalla on kuitenkin eroja, koska kielenoppijat ylikäyttävät astemääritettä *oikein* ja alikäyttävät kaikkia muita. *Sangen*-määritettä ei käytetä kertaakaan. *Oikein*-määritteen ylikäytöstä suomea vieraana kielenä opiskelevien kielenkäytössä on huomauttanut myös Isohäätä (2014: 90). Vaikka hän tutki B1–B2-tason tekstejä niin kuin tässäkin tutkimuksessa, Isohäätäällä kielenoppijoiden äidinkielen kirjo oli laajempi kuin vain viro (mts. 3, 5).

Vertailtaessa natiivi- ja oppijanaineistojen kollokaattikirjoa vähimmäiskynnyksen yläpuolella ilmenee, että kielenoppijat käyttävät vähiten erilaisia natiiveille ominaisia kollokaattiorakenteita *aivan*- ja *kovin*-määritteiden tarkastelualueilla. Siihen voi vaikuttaa aineiston pienuus, koska oppijanaineistossa sekä *aivan* että *kovin* esiintyvät vähemmän kuin 20 kertaa. Siitä johtuen on mahdollista, että muut natiiveille ominaiset kollokaattiorakenteet eivät yksinkertaisesti ylittäneet tämän tutkimuksen vähimmäiskynnystä. Kaikkien muiden astemääritteiden tarkastelualueilla hyödynnetään ainakin neljäsosaa (*ihan*), mutta joskus jopa puolia tai enemmän kuin puolia (*erittäin* ja *oikein* oikeanpuoleisella tarkastelualueella) natiivien kielenkäytössä tavallisista

kollokaateista. *Ihan*-sanasta hyödynnettiin enemmän natiiviaineiston vasemmanpuoleisia kollokaatteja, *erittäin*- ja *oikein*-sanoista puolestaan oikeanpuoleisia kollokaatteja. Oppijanaineistossa toisia yhteisiä kollokaatteja ylikäytettiin ja toisia alikäytettiin, samalla esiintyi myös niitä, joiden käyttö oli tasaista.

Lisäksi kielenoppijat käyttävät sellaisia kollokaatorakenteita, joita ei esiinny natiiviaineistossa vähimmäiskynnyksen yläpuolella. Yleensä natiiviaineiston kannalta epätyypilliset kollokaatorakenteet esiintyvät astemääritteen oikeanpuoleisella tarkastelualueella (paitsi *erittäin*-sanalla, jolla on noin prosentin ero toisinpäin). Vähiten epätyypillisiä kollokaatorakenteita koko kollokaattikirjosta oli *kovin*- (50 %) ja eniten *aivan*-sanan oikeanpuoleisella (100 %) tarkastelualueella. Niin kuin todettiin, niitä molempia esiintyi aineistossa suhteellisen vähän ja vähimmäiskynnyksen yläpuolelle ylsivät vain yksittäiset kollokaatit. Kaikkien muiden astemääritteiden epätyypillisten kollokaatorakenteiden osuus kaikkien kollokaattien kirjosta oli 76 %:stä (*oikein*) 82,76 %:iin (*ihan*) oikeanpuoleisella tarkastelualueella. Epätyypillisuus ei siis tarkoita välttämättä, että kollokaatorakenne olisi natiivien kielenkäytössä mahdoton, se vain ei ylittänyt vähimmäiskynnystä Eduskunta-aineistossa. Se, kuinka suuri osa oppijanaineiston unikaaleista kollokaateista on natiiviaineistossa mahdottomia, vaatisi erillistä tutkimusta.

Lisäksi analyysissa nousi esiin se, että kielenoppijat käyttävät verrattuna natiiveihin enemmän erilaisia substantiivikollokaatteja, mutta pronomien käyttö on suhteellisen tasaista. Siihen vaikuttaa ilmeisesti se, että oppijanaineiston teksteissä käsiteltävien aiheiden piiri on suppeampi kuin natiiviaineistossa ja näin oppijanaineistosta nousevatkin esiin suuressa määrin esimerkiksi perheeseen ja kotiin viittaavat substantiivit. Aineistojen välillä eroaa myös PITÄÄ-kollokaatin käyttö, jota kielenoppijat käyttävät ensisijaisesti mieltymyksen ilmaisemisessa.

Vaikka oppijanaineistossa löytyy sellaisia kollokaatteja, joita ei esiinny natiiviaineistossa vähimmäiskynnyksen yläpuolella, kielenoppijat hyötyntävät taitavasti natiivien kielenkäytössä tavallisia kollokaatteja. Oppijoiden kielitaidon kehittämiseksi oppijoita voisi kehottaa käyttämään enemmän erilaisia astemääritteitä *oikein*-sanän lisäksi. Lisäksi

voisi kiinnittää huomiota vasemmanpuoleisella tarkastelualueella EI-kollokaattiin, koska muutamat astemääritteet (esim. *kovin*) ovat kieltohakuisia ja muutamat eivät (esim. *sangen*). Oikeanpuoleisella tarkastelualueella voisi kiinnittää huomiota adverbi-, adjektiivi- ja kvanttorikolligaattien yleisjärjestykseen (tai muutaman sanaluokkakolligaatin puuttumiseen) ja tavallisimpiin kollokaatteihin. Jälkimmäistä tehtävää varten liitteenä (liite 1) on taulukko, josta näkee kaikki ne astemääritteiden *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* oikeanpuoleisella tarkastelualueella esiintyvät adjektiivi-, adverbi- ja kvanttorikollokaatit, jotka ylsivät vähimmäiskynnyksen yläpuolelle Eduskunta-aineistossa.

## YHTEENVETO

Tämän tutkimuksen tavoitteena oli luoda katsaus intensiteettiä ilmaisevien astemääritteiden *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* keskinäisiin samanlaisuuksiin ja eroihin natiivien kielenkäytössä ja sen jälkeen vertailla, miten virolaisten B1–B2-taitotason kielenoppijoiden astemääritekäyttö vastaa natiivien astemääritekäyttöä.

Tutkimustavoitteiden toteuttamiseksi tarkasteltiin astemääritekäyttöä korpusten perusteella. Natiiviaineistona käytettiin Eduskunnan täysistuntoja sisältävää korpusta (Eduskunta) ja oppijanaaineistona Kansainvälistä oppijansuomen korpusta (ICLFI). Eduskunta-aineistossa astemääritteitä esiintyi yhteensä 69 832 kertaa ja ICLFI-aineistossa 529 kertaa. Aineistojen keskeistä vertailtavuutta varten astemääritekäyttöä tarkasteltiin sekä suhteellisten frekvenssien (10 000 sanaa kohti) että suhteellisten osuuksien perusteella (prosentuaalisesti kaikista esiintymistä). Astemääritteitä tutkittiin tarkastelemalla niiden kontekstuaalisia erityispiirteitä. Tarkemmin keskityttiin kolligaatioon ja kollokaatioon.

Natiiviaineiston analyysistä ilmeni, että astemääritteiden vasemmanpuoleinen tarkastelualue on verrattuna oikeanpuoleiseen tarkastelualueeseen suhteellisen homogeeninen, koska siellä esiintyy eniten verbi- ja pronomnikolligaaatteja. Suurin osa verbikollokaateista puolestaan kuuluu suomen kielen yleisimpiin sanoihin ja myös pronominien luokka on rajallinen. Eroja näkee kuitenkin EI-kollokaatin käytössä. Oikeanpuoleisella tarkastelualueella esiintyi odotuksenmukaisesti eniten adjektiivivi-, adverbi- ja kvanttorikolligaaatteja. Analyysi paljasti, että niiden sanaluokkakolligaattien keskinäinen järjestys erottaa astemääritteet toisistaan. Vaikka usein astemääritteitä käsitellään synonyymisinä sanoina, yleensä ne eivät ole keskenään vaihdettavissa.

Kaikille astemääritteille yhteisiä kollokaatteja on niukalti. Jokaisella astemääritteellä on omat erityispiirteensä ja kahta sellaista astemääritettä, joiden käyttö olisi identtistä, ei löydy.

Tutkimuksen johdannossa pystytetty hypoteesi synonyymisten astemääritteiden *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* haasteellisuudesta kielenoppijan kannalta pitää tulosten perusteella paikkansa. Kielenoppijat käyttävät astemääritteitä natiivien kanssa melko samankaltaisesti, mutta he ylikäyttävät *oikein*-astemääritettä ja alikäyttävät kaikkia muita (etenkin *sangen*, jota oppijanaineistossa ei esiinny). Vaikka kielenoppijat käyttävät taitavasti natiiveille ominaisia kollokaatteja, heidän kielenkäytössään on myös paljon sellaisia kollokaatteja, jotka eivät ole natiivien kielenkäytössä (niin) tavallisia.

Oppijoiden kielitaidon kehittämiseksi voisi toisaalta kehottaa heitä käyttämään enemmän erilaisia astemääritteitä *oikein*-sanana lisäksi, toisaalta kiinnittää huomiota astemääritteen vasemmanpuoleisessa kontekstissa EI-kollokaatin mahdollisuuteen. Lisäksi voisi kiinnittää enemmän huomiota oikeanpuoleisella tarkastelualueella adjektiivi-, adverbi- ja kvanttorikolligaattien keskeiseen järjestykseen ja tavallisimpiin kollokaatteihin.

Tässä tutkimuksessa astemääritteiden tarkastelualueella esiintyy enemmän pronomineja kuin substantiiveja. Tulevaisuudessa voisi tutkia, mistä se johtuu: onko kyse siitä, että astemääritteiden käyttöympäristössä tarkoitteisiin viitataan enemmän pronomineilla vai onko taustalla tämän tutkimuksen vähimmäiskynnys, jota substantiivit eivät ylittäneet. Lisäksi voisi tutkia, kuinka suuri osa oppijanaineiston unikaaleista kollokaateista tuntuu natiiveille epätavallisille.

# LÄHTEET

## Tutkimusaineisto

**Korp = Borin, Lars, Markus Forsberg and Johan Roxendal 2012.** Korp – the corpus infrastructure of Språkbanken. BibTeX. Saatavilla: <https://korp.csc.fi>.

**Eduskunta = Eduskunta 2017.** Eduskunnan täysistunnot, Kielipankin Korp-versio 1.5 [tekstikorpus]. Kielipankki. Saatavilla: <http://urn.fi/urn:nbn:fi:lb-2019101621>.

**ICLFI = Jantunen, Jarmo Harri, Sisko Brunni ja Oulun yliopisto, suomen kielen oppiaine 2013.** Kansainvälinen oppijansuomen korpus [tekstikorpus]. Kielipankki. Saatavilla: <http://urn.fi/urn:nbn:fi:lb-20140730163>.

## Kirjallisuus

**Ahlman, Erik 1933.** Adverbeista. – Virittäjä 37, 137–159; <https://journal.fi/virittaja/article/view/29416>. Katsottu 3.1.2019.

**Biber, Douglas 1988.** Variation Across Speech and Writing. Cambridge: Cambridge University Press.

**Carter, Ronald 2004.** Introduction. – Trust the text. Language, corpus and discourse. Ed. John Sinclair, with Ronald Carter. London, New York: Routledge, 1–6; ProQuest Ebook Central, <https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=200411>. Katsottu 14.12.2019.

**Collins Dictionary = Collins Online English Dictionary.**

[https://www.collinsdictionary.com/cobuild/?\\_\\_cf\\_chl\\_jschl\\_t](https://www.collinsdictionary.com/cobuild/?__cf_chl_jschl_t). Katsottu 8.3.2020.

**ConLexis = Oulun yliopiston suomi toisena ja vieraana kielenä -oppiaine, Oulun Aikuiskoulutuskeskuksen suomen kielen opetuksen yksikkö.** ConLexis verkkosanakirja. 1. versio. <http://wiki.osao.fi/conlexis/Etusivu>. Katsottu 5.5.2020.

**Cowie, Anthony Paul 2001 (1998).** Introduction. – *Phraseology: Theory, Analysis, and Applications*. Ed. Anthony Paul Cowie. Oxford: Oxford University Press, 1–20.

**Ellis, Nick C. 2008.** *Phraseology: The periphery and the heart of language*. – *Phraseology in Foreign Language Learning and Teaching*. Ed. Fanny Meunier, Sylviane Granger. Amsterdam, Philadelphia: John Benjamins Publishing Company, 1–13; ProQuest Ebook Central, <https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=622977>. Katsottu 18.12.2019.

**Erman Britt, Beatrice Warren 2000.** The idiom principle and the open choice principle. – *Text* 20 (1), 29–62; <https://www.semanticscholar.org/paper/The-idiom-principle-and-the-open-choice-principle-Erman-Warren/bb9b41f92fb34faa965ad126a10340629276de04>. Katsottu 6.3.2020.

**Firth, John Rupert 1962.** A synopsis of linguistic theory, 1930–1955. – *Studies in linguistic analysis*. Ed. John Rupert Firth. Oxford: Basil Blackwell, 1–32; <http://cs.brown.edu/courses/csci2952d/readings/lecture1-firth.pdf>. Katsottu 2.3.2020.

**Granger, Sylviane 1996.** From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. – *Languages in contrast: Papers from a symposium on text-based cross-linguistic studies, Lund 4-5 March 1994*. Ed. Karin Aijmer, Bengt Altenberg, Mats Johansson. *Lund Studies in English*, Vol. 88. Lund: Lund University Press, 37–51; <http://hdl.handle.net/2078.1/75847>. Katsottu 5.5.2020.

**Granger, Sylviane 1998.** The computer learner corpus: A versatile new source of data for SLA research. – *Learner English on Computer*. Ed. Sylviane Granger. London, New York: Longman, 3–18.

**Granger, Sylviane 2001 (1998).** Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. – *Phraseology: Theory, Analysis and Applications*. Ed. Anthony Paul Cowie. Oxford: Oxford University Press, 145–160.

**Granger, Sylviane 2005.** Pushing back the limits of phraseology. How far can we go? – *Proceedings of the Phraseology 2005 Conference, Louvain-la-Neuve, 13–15 October 2005*. Ed. Christelle Cosme, Céline Gouverneur, Fanny Meunier, Magali Paquot. Louvain-la-Neuve: Presses universitaires de Louvain, 165–168; <http://hdl.handle.net/2078.1/75667>, 1–4. Katsottu 15.2.2020.

**Granger, Sylviane, Paul Rayson 1998.** Automatic profiling of learner texts. – *Learner English on Computer*. Ed. Sylviane Granger. London, New York: Longman, 119–131

**Hakulinen, Auli, Fred Karlsson 1988.** *Nykysuomen lauseoppia*. 2. painos. Helsinki: Suomalaisen Kirjallisuuden Seura.

**Halme, Suvi 2017.** Sangen monet merkitykset: sangen-sanan syntaktis-semanttista tarkastelua vanhasta kirjasuomesta varhaisnykysuomeen. Pro gradu -tutkielma; [https://epublications.uef.fi/pub/urn\\_nbn\\_fi\\_uef-20170271/urn\\_nbn\\_fi\\_uef-20170271.pdf](https://epublications.uef.fi/pub/urn_nbn_fi_uef-20170271/urn_nbn_fi_uef-20170271.pdf). Katsottu 19.5.2020.

**Heikkinen, Vesa, Outi Lehtinen, Mikko Lounela 2001.** Kuvia kirjoitetusta suomesta. – *Kielikello* 3/2001, 12–15; <https://www.kielikello.fi/-/kuvia-kirjoitetusta-suomesta>. Katsottu 17.5.2020.

**Howarth, Peter 2001 (1998).** *The Phraseology of Learners' Academic Writing*. – *Phraseology: Theory, Analysis, and Applications*. Ed. Anthony Paul Cowie. Oxford: Oxford University Press, 161–186.

**Huomo, Tuomas 2019.** Melkein kohti, ihan lähellä, melko edessä: Mitä astemääritteet kertovat spatiaalisten grammien semantiikasta. – *Sananjalka* 61, 7–29; <https://doi.org/10.30673/sja.80150>. Katsottu 27.1.2020.

**Isohätälä, Jaana 2014.** Kielenoppimisympäristön vaikutus oppijansuomen sanastoon. Pro gradu -tutkielma; <http://jultika.oulu.fi/Record/nbnfioulu-201405291585>. Katsottu 27.11.2019.

**Jantunen, Jarmo Harri 2001.** "Tärkeä seikka" ja "keskeinen kysymys": Mitä korpuslingvistinen analyysi paljastaa lähisynonyymeista? – *Virittäjä* 105 (2), 170–192; <https://journal.fi/virittaja/article/view/40082>. Katsottu 4.3.2020.

**Jantunen, Jarmo Harri 2004.** Synonymia ja käännessuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännskielen leksikaalisiin erityispiirteisiin. Väitöskirja. Joensuun yliopiston humanistisia julkaisuja 35; [https://epublications.uef.fi/pub/urn\\_isbn\\_952-458-479-4/](https://epublications.uef.fi/pub/urn_isbn_952-458-479-4/). Katsottu 24.1.2020.

**Jantunen, Jarmo Harri 2008.** Haasteita oppijankielen korpusanalyysille: oppijankielen universaalit. – *Õppijakeele analüüs: võimalused, probleemid, vajadused*. Toim. Pille Eslon. Tallinna Ülikooli eesti filoloogia osakonna toimetised 10. Tallinn: Tallinna Ülikooli kirjastus; 67–92.

**Jantunen, Jarmo Harri 2009.** Minulla on aivan paljon rahaa – Fraseologiset yksiköt suomen kielen opetuksessa [‘I have really lots of money’ – Phraseological units in the teaching of Finnish]. – *Virittäjä* 113 (3), 356–381; <https://journal.fi/virittaja/article/view/4202>. Katsottu 28.11.19.

**Jantunen, Jarmo Harri 2015.** Oppimiskontekstin vaikutus oppijanpragmatiikkaan: astemääritteet leksikaalisina nallekarhuina. – *Lähivõrdlusi. Lähivertailuja* 25, 105–136; <http://arhiiv.rakenduslingvistika.ee/ajakirjad/index.php/lahivordlusi/article/view/LV25.05/352>. Katsottu 14.10.2019.

**Jantunen, Jarmo Harri, Sisko Brunni 2012.** Morfologinen priming ja fraseologia vieraan kielen oppimisessa: korpustutkimus oppijansuomesta. – *Lähivõrdlusi. Lähiveriluja* 22, 71–100; <http://arhiiv.rakenduslingvistika.ee/ajakirjad/index.php/lahivordlusi/article/view/LV22.03>. Katsottu 23.3.2020.

**Johansson, Stig 2011.** Corpus, lexis, discourse: a tribute to John Sinclair. – The Phraseological View of Language: A Tribute to John Sinclair. Ed. Thomas Herbst, Susen Faulhaber, Peter Uhrig. Berlin, Boston: De Gruyter Mouton, 17–26; ProQuest Ebook Central, <https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=827309>. Katsottu 17.3.2020.

**Kennedy, Graeme 2003.** Amplifier Collocations in the British National Corpus: Implications for English Language Teaching. – TESOL Quarterly, 37 (3), 467–487; <https://www.jstor.org/stable/3588400?seq=1>. Katsottu 10.3.2020.

**Kennedy, Graeme D. 2010 (1998).** An Introduction to Corpus Linguistics. London, New York: Longman.

**Kitsnik, Mare 2006.** Keelekorpused ja võõrkeeleõpe. – Eesti Rakenduslingvistika aastaraamat 2. Toim. Helle Metslang, Margit Langemets. Tallinn: Eesti Keele Sihtasutus, 93–107; <http://arhiiv.rakenduslingvistika.ee/ajakirjad/index.php/aastaraamat/article/view/ERYa2.07/163> Katsottu 6.1.2020.

**Kjellmer, Göran 1984.** Some Thoughts on Collocational Distinctiveness. – Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research. Ed. Jan Aarts, Willem Meijs. Amsterdam: Rudopi, 163–171.

**Kjellmer, Göran 2013 (1991).** A Mint of Phrases. – English Corpus Linguistics. Ed. Karin Aijmer, Bengt Altenberg. New York: Routledge, 111–127; <https://www-taylorfrancis-com.ezproxy.utlib.ut.ee/books/e/9781315845890>. Katsottu 10.5.2020.

**Koski, Mauno 1988.** Adjektiiviadverbien semantiikkaa ja pragmatiikkaa. Virittäjä 92 (3), 407–412; <https://journal.fi/virittaja/article/view/38240>. Katsottu 27.1.2020.

**KTS = Kielitoimiston sanakirja 2020.** Helsinki: Kotimaisten kielten keskus; verkkojulkaisu, päivitettävä julkaisu; <https://www.kielitoimistonsanakirja.fi>. Katsottu 16.4.2020.

**Kuusi, Anna-Leena 1971.** Johdatusta suomen kielen fraseologiaan. Helsinki: Suomalaisen Kirjallisuuden Seura.

**Lappalainen, Hanna 2001.** Sosiolingvistinen katsaus suomalaisnuorten nykypuhekieleen ja sen tutkimukseen. – *Virittäjä*, 105 (1), 74–101; <https://journal.fi/virittaja/article/view/40068>. Katsottu 20.2.2020

**Lim, Ni-Eng, Huaqing Hong 2012.** Intensifiers as stance markers: A corpus study on genre variation in Mandarin Chinese. – *Chinese Language and Discourse* 3 (2), s 129–166; <http://dx.doi.org/10.1075/cld.3.2.01lim>. Katsottu 25.11.2019.

**Lorenz, Gunter R. 1999.** Adjective Intensification: Learners Versus Native Speakers. A Corpus Study of Argumentative Writing. – *Language and Computers: Studies in Practical Linguistics*, No. 27. Ed. Jan Aarts and Willem Meijs. Amsterdam, Atlanta: Rodopi.

**Louw, Bill 1993.** Irony in the Text or Insincerity in the Writer? – The Diagnostic Potential of Semantic Prosodies. – *Text and Technology: In honour of John Sinclair*. Ed. Mona Baker, Gill Francis, Elena Tognini-Bonelli. Philadelphia, Amsterdam: John Benjamins Publishing Company, 157–176; ProQuest Ebook Central, <https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=680963>. Katsottu 3.4.2020.

**Meunier, Fanny 1998.** Computer tools for the analysis of learner corpora. – *Learner English on Computer*. Ed. Sylviane Granger. London, New York: Longman, 19–37.

**Nesselhauf, Nadja 2003.** The Use of Collocations by Advanced Learners of English and Some Implications for Teaching – *Applied Linguistics*, Vol. 24 (2), 223–242; <https://doi-org.ezproxy.utlib.ut.ee/10.1093/applin/24.2.223>. Katsottu 11.12.2019.

**Orpana, Terttu 1998.** Kuvaus tai kommentti. Tutkimus suomen kielen adjektiiviadverbien semanttisesta tulkinnasta. *Opera Fennistica ja Linguistica*, 2. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.

**Pajusalu, Renate 2009.** Sõna ja tähendus. Tallinn: Eesti Keele Sihtasutus.

**Palmanto, Mirja 2004.** Suomenoppijoiden kokemuksia kielenkäytöstä ja kielitaidon kehittymisestä työharjoittelussa – Virittäjä 108 (3), 405–114; <https://journal.fi/virittaja/article/view/40352>. Katsottu 26.4.2020.

**Partington, Alan 1998.** Patterns and Meanings. Using Corpora for English Language Research and Teaching. Amsterdam, Philadelphia: John Benjamins Publishing Company; ProQuest Ebook Central, <https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=622950>. Katsottu 9.1.2020.

**Pérez-Paredes, Pascual, María Belén Díez-Bedmar 2012.** The use of intensifying adverbs in learner writing. – Developmental and Crosslinguistic Perspectives in Learner Corpus Research. Ed. Yukio Tono, Yuji Kawaguchi, Makoto Minegishi. Tokyo University of Foreign Studies. Studies in Linguistics, 4. Amsterdam, Philadelphia: John Benjamins Publishing Company, 105–124; <https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=1161965>. Katsottu 28.11.2019.

**Pällin, Kristi, Annekatrin Kaivapalu 2012.** Suomen mennä ja lähteä vertailussa: lähtökohtana vironkielinen suomenoppija. – Lähivördlusi. Lähivertailuja 22, 287–323; <http://www.rakenduslingvistika.ee/ajakirjad/index.php/lahivordlusi/article/view/LV22.10/199>. Katsottu 30.12.2019.

**Ranalli, James M. 2003.** ELT coursebooks in the age of corpus linguistics: constraints and possibilities; <https://www.birmingham.ac.uk/Documents/college-artslaw/cels/essays/corpuslinguistics/Ranalli6.pdf>. Katsottu 14.2.2020.

**Räsänen, Matti 2010.** Päivä päivältä enemmän ja enemmän: Suomen toistokonstruktioita. Väitöskirja. Espoo: Matti Räsänen; <https://helda.helsinki.fi/handle/10138/19663>. Katsottu 24.4.2020.

**Saukkonen, Pauli, Marjatta Haipus, Antero Niemikorpi, Helena Sulkala 1979.** Suomen kielen taajuussanasto. A Frequency Dictionary of Finnish. Porvoo, Helsinki, Juva: Werner Söderström Osakeyhtiö.

**Segalowitz, Norman, Barbara F. Freed 2004.** Context, contact, and cognition in oral fluency acquisition. Learning Spanish in at home and study abroad contexts. – *Studies in Second Language Acquisition* 26 (2), 173–199; <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/context-contact-and-cognition-in-oral-fluency-acquisition-learning-spanish-in-at-home-and-study-abroad-contexts/>. Katsottu 28.11.2019.

**Sinclair, John McHardy 1987a.** Introduction – Looking Up. An account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English Language Dictionary. Ed. John McHardy Sinclair. London, Glasgow: Collins ELT, VII-IX.

**Sinclair, John McHardy 1987b.** The Nature of the Evidence – Looking Up. An account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English Language Dictionary. Ed. John McHardy Sinclair. London, Glasgow: Collins ELT, 150–159.

**Sinclair, John 1991.** Corpus, Concordance, Collocation. Oxford: Oxford University Press; <https://ru.b-ok2.org/book/2173916/43692c>. Katsottu 10.5.2020.

**Sinclair, John 2004a.** Trust the text. Language, corpus and discourse. Ed. with Ronald Carter. London, New York: Routledge; ProQuest Ebook Central, <https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=200411>. Katsottu 14.12.2019.

**Sinclair, John 2004b.** Introduction. – How to use corpora in language teaching. Ed. John McHardy Sinclair. *Studies in Corpus Linguistics*, Vol. 12. Amsterdam, Philadelphia: John Benjamins Publishing Company, 1–10; ProQuest Ebook Central,

<https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=622666>. Luettu 16.12.2019.

**Stubbs, Michael 1993.** British Traditions in Text Analysis – From Firth to Sinclair. – Text and Technology: In honour of John Sinclair. Ed. Mona Baker, Gill Francis, Elena Tognini-Bonelli. Philadelphia, Amsterdam: John Benjamins Publishing Company, 1–33; ProQuest Ebook Central, <https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=680963>. Katsottu 8.5.2020.

**Stubbs, Michael 1996.** Text and Corpus Analysis. Corpus-assisted Studies of Language and Culture. Oxford: Blackwell.

**Tanner, Johanna 2008.** S2-oppikirjojen dialogien tarkastelua. – Suomenoppiminen, kielitaito ja tutkimus. Toim. Johanna Tanner, Marja Kokkonen. Kakkoskieli 6. Helsingin yliopisto: Suomen kielen ja kotimaisen kirjallisuuden laitos, 163–189.

**Tognini-Bonelli, Elena 2001.** Corpus Linguistics at Work. Studies in Corpus Linguistics, Vol. 6. Amsterdam, Philadelphia: John Benjamins Publishing Company.

**Tsui, Amy B. M. 2004.** What teachers have always wanted to know – and how corpora can help. – How to use corpora in language teaching. Ed. John McHardy Sinclair. Studies in Corpus Linguistics, Vol. 12. Amsterdam, Philadelphia: John Benjamins Publishing Company, 39–61; ProQuest Ebook Central, <https://ebookcentral-proquest-com.ezproxy.utlib.ut.ee/lib/tartu-ebooks/detail.action?docID=622666>. Katsottu 11.12.2019.

**Vanhatalo, Ulla 2003.** Kyselytekstit vs. korpuslingvistiikka lähisynonyymien semanttisten sisältöjen arvioinnissa - mitä vielä keskeisestä ja tärkeästä? – Virittäjä, 107(3), 351–369; <https://journal.fi/virittaja/article/view/40274>. Katsottu 4.3.2020.

**Vilkuna, Maria 1996.** Suomen lauseopin perusteet. 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. Verkkoversio; <http://scripta.kotus.fi/visk>. Katsottu 20.11.2019.

**Vuotilainen, Eero 2019.** Onko eduskunnan täysistunto pelkkää teatteria? – Kielikello, 4/2019; <https://www.kielikello.fi/-/onko-eduskunnan-taysistunto-pelkkaa-teatteria>. Katsottu 19.12.2019.

**Õim, Asta, Katre Õim 2019.** Lähtekohti eesti fraseoloogia käsitlemiseks. Toim. Pille Eslon. Eesti Teaduste Akadeemia Emakeele Seltsi toimetised nr 76. Tallinn: EKSA.

**Yaguchi, Michiko, Yoko Iyeiri, Yasumasa Baba 2010.** Speech style and gender distinctions in the use of very and real/really: An analysis of the Corpus of Spoken Professional American English. – *Journal of Pragmatics* 42 (3), 585–597; <http://dx.doi.org/10.1016/j.pragma.2009.08.002>. Katsottu 13.1.2020.

# KEELEÕPPIJA FRASEOLOOGIATIHNİKUS: MÄÄRAMÄÄRSÕNADE AIVAN, ERITTÄIN, IHAN, KOVIN, OIKEIN JA SANGEN KASUTUSKOHAD

## RESÜMEE

Magistritöö eesmärk on välja selgitada soomekeelsete sünonüümsete määrämäärsõnade (*astemäärite*) *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* kasutamiskohad tegeliku keelekasutuse põhjal nii emakeelsetel keelekõnelejal kui ka keeleõppijatel ning neid omavahel võrrelda. Sageli seletatakse sõnade tähendusi sünonüümide abil, kuid neist suurem osa ei ole kõikides kontekstides üksteisega asendatavad. Sellest tulenevalt võib sünonüüme vahetades moodustuda lause, mis on küll grammatiliselt korrektne, ent mis kõlab emakeelena kõnelejale ebatavaliselt. Teatud sõnad kipuvad sageli koos esinema ja teised mitte. Selliseid sõnade koosinemise seaduspärasid uurib Briti fraseoloogia koolkond, mille ideedest lähtutakse ka siinses töös. Uurimuse aluseks on kvantitatiivne analüüs.

Emakeele kõneleja valib konteksti sobiva sõna intuiitiivselt, aga keeleõppija seisab silmitsi raskustega. Õppijat mõjutavad valiku juures näiteks tema emakeel, keeletase ja seegi, kas keelt õpitakse teise või võõrkeelena. Sellest lähtuvalt püstitati hüpotees, et keeleõppijale valmistab sünonüümsete määrämäärsõnade kasutamine raskusi. Hüpoteesi kontrollimiseks sõnastati kaks uurimisküsimust. Esiteks: mille poolest sarnanevad ja erinevad sõnad *aivan*, *erittäin*, *ihan*, *oikein*, *kovin* ja *sangen* emakeelena kõnelejate keeles? Teiseks: kuidas vastab emakeelena kõnelejate valikutele B1–B2-keelelasemega eestlastest keeleõppijate määrämäärsõnade *aivan*, *erittäin*, *ihan*, *oikein*, *kovin* ja *sangen* kasutus?

Töö koosneb neljast põhipeatükist. Esimeses antakse ülevaade määramäärsõnadest. Teises tutvustakse fraseoloogiat uurimissuunana, keskendudes just Briti fraseoloogia koolkonna vaadetele, ning tutvustatakse siinses töös olulisemaid mõisteid (näiteks *kollokatsioon* ja *kolligatsioon*). Kolmandas peatükis kirjeldatakse tegureid, mis võivad õppijate keelekasutust mõjutada, ning antakse ülevaade senistest õppijakeele määramääraõnade uurimustest. Neljas peatükk keskendub analüüsile.

Kuna tegeliku keelekasutusena käsitatakse Briti fraseoloogia koolkonnas muuhulgas keelekorpuste tekste, võetakse õppijakeele näited soome õppijakeele korpusest (ICLFI). Emakeelena kõnelejate tekstinäited pärinevad parlamendi täisistungeid sisaldavast korpusest (Eduskunta). Neist esimeses kasutatakse määramäärsõnu 529 korral ning teises 69 832 korral. Kollokatsioonide uurimiseks otsiti programmi WordSmithTools abil kõik need määramäärsõnade vasak- ja parempoolsed kollokaadid, mille esinemissagedus oli suurem kui 1,4% määramäärsõna esinemissagedusest. Kollokaate otsiti nii paremalt kui ka vasakult poolt kahe sõna ulatuses. Seejärel määrati kõikide kollokaatide sõnaliik.

Täpsem kvantitatiivne analüüs tehti Excelis, kus võrreldi määramäärsõnade vasak- ja parempoolsete kollokaatide ja kolligaatide esinemissagedusi. Selgus, et kõikide sõnade vasakpoolne kontekst on võrdlemisi homogeenne, sisaldades kõige rohkem verbi- ja pronoomenikolligaate. Kõikide määramäärsõnade ülekaalukaim verbikollokaat oli soome keeles enim esinev sõna OLLA, ent ka suur osa teistest verbikollokaatidest on soome keeles sagedasti kasutatavad sõnad. Ehkki pronoomenikollokaate kasutati suurel hulgal, oli nende hulgas varieeruvust vähe.

Suuremad erinevused ilmneseid sõna parempoolses kontekstis. Sõnad erinevad eelkõige adjektiiv-, adverbi- ja kvantorikolligaatide järjestuse poolest. Sellest võib järeldada, et määramäärsõnad erinevad omavahel eelkõige selle poolest, kas täpsustavad enim omaduse, viisi või hulga astet. Kõikidele sõnadele ühiseid kollokaate leidub napilt. Analüüsist selgus, et ehkki sõnu peetakse sünonüümideks, on neile kõigile ühiseid vasak- ja parempoolseid kollokaate kokku vaid seitse. Seega on *aivan*, *erittäin*, *ihan*, *oikein*, *kovin* ja *sangen* pigem osasünonüümid, mis ei ole kõikides kontekstides omavahel vahetatavad.

Eduskunta- ja ICLFI-korpuse võrdlus näitas, et õppijad ja emakeelena kõnelejad kasutavad määramäärsõnu üsna võrdselt. Siinjuures õppijad ülekasutavad *oikein*-sõna ja alakasutavad teisi, määramäärsõna *sangen* ei kasutata. Õppijad tarvitavad ühelt poolt emakeelena kõnelejatele omaseid kollokaate, kuid teisalt ka niisuguseid, mis ei ole viimaste puhul kuigi tavalised. Ühiste kollokaatide kasutamisel on mõningaid erinevusi, sest siin esineb nii üle- kui ka alakasutamist. Püstitatud hüpotees, et õppijatele on määramäärsõnade *aivan*, *erittäin*, *ihan*, *kovin*, *oikein* ja *sangen* kasutamine keerukas, leidis kinnitust.

Uurimuses esines määramäärsõnade kontekstis rohkem pronoomeneid kui substantiive. Edaspidi võiks uurida, kas siin on põhjuseks uurimuses kehtestatud alampiir, mida suurem osa substantiive ei ületanud või viidataksegi määramäärsõnade kontekstis erinevatele objektidele pigem pronoomenitega. Lisaks võiks uurida, kui suurt osa vaid õppijakeelele omaseid kollokaate tajuvad emakeelena kõnelejad ebatavalisena.

## Liite 1. Oikeanpuoleiset adjektiiv-, adverb- ja kvanttorikolligaatit Eduskunta-aineistossa.

	aivan	erittäin	ihan	kovin	oikein	sangen
eri	X		X			
erilainen				X		
erinomainen	X					
erityisesti	X					
helppo				X		
hyvin	X	X	X	X	X	X
hyvä		X	X	X	X	X
iso		X				
keskeinen	X					
kunnolla					X	
kunnon					X	
käsittämätön	X					
liian	X					
lyhyt						X
merkittävä		X				
mielenkiintoinen						X
oikea	X				X	
oikeasti			X			
oikein	X		X			
oma			X			
paljo				X		
paljon		X		X	X	
pitkä				X		
selvä	X		X			
siksi						X
suuresti						X
suuri		X		X		X
tosi	X		X			
tärkeä		X				
uusi	X					
vaikea		X		X		
vakava		X				

varma	X					
varmasti	X					
vähän				X		X
välttämätön	X					
yksimielinen						X

## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Ewa Pikk,el,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

Kielenoppija fraseoloogisessa viidakossa: astemäärítteiden *aivan, erittäin, ihan, kovin, oikein* ja *sangen* käyttökontekstit,

mille juhendaja on Tiina Liisa Marjatta Nahkola,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Ewa Pikk*

**25.05.2020**