

LIKIMÄÄRÄN ILMAISIMET NATIIVI- JA OPPIJANSUOMESSA

Lekseemien *lähes* ja *melkein* vierussanatarkastelu

Suomen kielen
kandidaatintutkielma
Oulun yliopisto
15.1.2016

Virpi Kukkonen

SISÄLLYS

1. JOHDANTO	1
1.1. Tutkimuskysymykset ja hypoteesit	1
1.2. Aineistoina ICLFI ja supisuomi	2
1.3. Tutkimusmenetelmä	3
2. TEOREETTINEN VIITEKEHYS	6
2.1. Kontekstuaalinen semantiikka	6
2.2. Keskeiset käsitteet	7
2.3. Oppijankielen korpustutkimus	9
3. LIKIMÄÄRÄN ILMAISIMET	11
3.1. Yleistä adverbista ja partikkeleista	11
3.2. Likimäärän ilmaisu suomen kielessä	12
3.3. Havaintoyksiköt <i>lähes</i> ja <i>melkein</i>	13
4. ANALYYSI	16
4.1. Likimäärän ilmaisinten frekvenssitarkastelu	16
4.2. Tapaus 1: <i>melkein</i>	18
4.2.1. Kollokaattitarkastelu	18
4.2.2. Klusterit ja muotit	19
4.3. Tapaus 2: <i>lähes</i>	21
4.3.1. Kollokaattitarkastelu	22
4.3.2. Klusterit ja muotit	23
4.4. Määritteiden <i>lähes</i> ja <i>melkein</i> keskinäistä vertailua	24
5. PÄÄTÄNTÖ	26
LÄHTEET JA LYHENTEET	28
LIITTEET	

1. JOHDANTO

Kandidaatintutkielmassani tarkastelen suomea vieraana kielenään opiskelevien suomenoppijoiden¹ tapaa käyttää likimäärän ilmaisimia, joita ovat esimerkiksi *lähes*, *melkein*, *suunnilleen* ja *noin*. Aiheeseen minua motivoi alkujaan kiinnostus lekseemien *lähes* ja *melkein* synonymiaan, minkä vuoksi keskityn tarkemmin tutkimaan näitä kahta astemääritettä ja niiden määrittämien vierussanojen laatua. *lähes* ja *melkein* ovat likimäärän ilmaisimista yleisimpien joukossa sekä oppijansuomen että natiivisuomen korpuksissa, mutta myös Suomen kielen taajuusanastossa (TS) (Saukkonen, Haipus, Niemikorpi & Sulkala 1979).

Aiempaa tutkimusta tämän kielen piirteen, likipitäisyyden, natiivikäytöstä en ole löytänyt. Vertailevan tutkimusotteeni vuoksi kuvaan työssäni äidinkielisten käyttötappaa. Astemääritteiden *hyvin*, *kovin* ja *oikein* synonymiaa käännössuomesta on tutkinut Jarmo Jantunen. Hänen havaintoyksiköiden valintakriteereistään omaan tapaukseeni sopii myös se, että astemääritteille on vastine ainakin indoeurooppalaisissa ja suomalais-ugrilaisissa kielissä. On siis todennäköistä, että suomenoppijoiden korpusaineistosta löytyy tutkittavaksi tarpeeksi likimäärää koskevia tapauksia. (Jantunen 2004: 68–69.)

Yleiskuvaukseni ilmiöstä perustuu lähinnä Isoon suomen kielioppiin (ISK 2004 § 615, 661, 662, 792, 855) ja Kielitoimiston sanakirjan (KS) internet-version sana-artikkeleihin. Tutkimuksessa käytän korpuslingvistisiä menetelmiä.

1.1. Tutkimuskysymykset ja hypoteesit

Tarkoitukseni on muodostaa yleiskuva suomenoppijoiden ja natiivien tavasta ilmaista likimäärää ja käyttää lekseemejä *lähes* ja *melkein*. Apunani ovat seuraavat kysymykset:

1. Mitä likimäärän ilmaisimia suomenoppijat ja natiivit käyttävät ja kuinka frekventisti?

¹ S2-lyhennettä käytetään sekä suomi toisena että suomi vieraana kielenä -oppiaineesta. Toisena kielenä kieltä opitaan kohdekielisessä ympäristössä, kun taas vieraana kielenä kieltä opitaan kohdekielisen ympäristön ulkopuolella. Kansainvälisesti toisen ja vieraan kielen oppialasta käytetään lyhennettä SLA (*second language acquisition*). (Martin 1999: 157; Nissilä 2011: 37–38.)

2. Millaisia asioita likimäärän ilmaisimet *lähes* ja *melkein* suomenoppijoiden teksteissä määrittävät eli millaisia **vieruskollokaatteja** (kerasanoja) ne saavat oikeanpuoleisessa **kotektissaan** eli tekstuaalisessa lähiympäristössään natiiviteksteihin verrattuina?
3. Millaisten rakenteiden, **muottien**², osaksi *lähes* ja *melkein* asettuvat aineistoissa?

Oletan ensinnäkin, että suomenoppijat hallitsevat vain yleisimmät natiivien käyttämistä likimäärän ilmaisimista. Toisena hypoteesinani esitän, että suomenoppijoilla on havaittavissa natiivien käyttötapaan nähden epätyypillistä käyttöä. Epätyypillisellä käytöllä tarkoitan sellaisia kielenoppijan astemäärite- ja kollokaattivalintoja, jollaisia natiivi ei vertailuaineiston eikä ISK:n ja KS:n esimerkkien perusteella tee. Virheet ortografiassa jätän huomiotta.

Astemäärite on lauseessa valinnainen elementti; sitä ilmeikkin lause on ehjä kokonaisuus, mutta kun se lisätään, merkitys muuttuu (ISK 2004 § 439). Koska itse määrite on taipumaton ja koska sen lisäys ei aiheuta sen määrittämässä sanassa tai lausekkeessa muutosta, uskon suomenoppijoiden onnistuvan likimäärän ilmaisinten käytössä paremmin kuin monien muiden kielenpiirteiden käytössä. Jätän kielenoppijoiden taitotasot ja oppijan lähtökielen vaikutuksen (*transfer*) satunnaisiksi huomioiksi. Tutkittavaa ilmiötä pidän neutraalina, tekstilajista ja oppijoiden tehtävänannosta riippumattomana kielenpiirteenä.

1.2. Aineistoina ICLFI ja supisuomi

Oppijankielen aineistona on Kansainvälinen oppijansuomen korpus, ICLFI (*International Corpus of Learner Finnish*), jota alettiin koostaa 2007 Oulun yliopiston ja neljän muun pohjoisen yliopiston Korpustutkimus oppijankielen kielikohtaisista ja universaaleista ominaisuuksista -hankkeessa. ICLFI koostuu yli 20 ulkomaisessa yliopistossa suomea vieraana kielenä opiskelevien oppijoiden sekä fiktiivisistä että ei-fiktiivisistä

² **Muotti** on rakennehahmo tai kehikko, jonka avulla muodostetaan kielellisiä ilmauksia. Sitä on käytetty analogiaan perustuvana muodostusmallina esimerkiksi sanoja johdattaessa ja kun muodostetaan verbin täydennyksiä (esim. resultatiivimuotti). (ISK 2004 § 148; 474; Alho & Kauppinen 2009: 23.)

teksteistä. Tekstejä on 18 maasta, ja äidinkielenä ainakin englantia, espanjaa, hollantia, islantia, italiaa, kiinaa, puola, ranska, romanian, ruotsi, saksa, slovakia, tsekki, venäjä, viro ja unkari. Taustamuuttujina kerrotaan mm. keräysvuosi, sukupuoli, asuinpaikka, äidinkieli, vanhempien ja opettajan äidinkieli, Suomessa asuminen tai vierailu, oppikirja ja käytetyt apuvälineet, tekstityyppi ja tehtävänanto sekä tekstin tuottotapa, -aika ja -paikka. Aineisto on Eurooppalaisen viitekehyksen³ (2003) mukaan taitotasoarvioitua (A1–C2), mutta annotoimatonta eli lingvistisesti koodaamatonta. Saneita se sisältää n. 970 000. Korpus on karttuva. (ICLFI:n taustatietokansio; Jantunen & Pirkola 2015: 89, 92, 96.)

Natiivikielen aineistona käytän Joensuun yliopiston kansainvälisen viestinnän laitoksen vuosina 1997–2001 kokoamaa verrannollista Käännösuomen korpusta, joka sisältää natiivien alkusuomen (nk. supisuomen⁴) ja suomeksi käännettyjen tekstien osakorpukset. Teen havaintoja supisuomen osuudesta (n. 3,8 milj. sanetta). Tekstilajeina ovat akateemiset tekstit, elämäkerrat, kauno-, viihde- populaari- ja lastenkirjallisuus sekä dekkarit. Tekstit ovat annotoimattomia. (Käännösuomen korpuksen taustatietokansio; Jantunen 2011: 51.)

Olen tässä työssä pitäytynyt näiden kahden aineiston tarkastelussa, vaikka sanomalehtiaineiston mukaan ottaminen olisikin vastannut oppijoiden tehtävänannoista uutisen ja vastineen laatimista sekä täydentänyt natiivikäytöstä saatavaa kuvaa.

1.3. Tutkimusmenetelmä

Tutkimukseni on useampivaiheinen: aluksi sanakirja- ja kielioppitarkastelun (KS ja ISK 2004) avulla luon yleiskuvan ilmiöstä (ks. luku 3), seuraavaksi teen ns. konkordanssilistaukset hakusanoilla tietokoneohjelmaa käyttäen, mistä jatkan laskemalla ilmaisinten ja vierussanojen frekvenssejä manuaalisesti (ks. luku 4.1.). Lopuksi vertailen paitsi ilmaisinten frekvenssejä myös valitsemieni havaintoyksiköiden *lähes* ja *melkein* vierussanoja ja muotteja aineistoista keskenään (ks. luvut 4.2.–4.4.). Vertailtavina ovat paitsi natiiv-

³ Eurooppalainen viitekehys (CEFR) on Euroopan maiden yhteinen kielitaitotasojen kuvaus-asteikko. Se on kehitetty kielenoppijoiden edistymisen seurantaan opetuksessa, mutta myös koulu-tussuunnittelijoille, oppimateriaalin valmistajille ja viranomaisille (Eurooppalainen viitekehys 2003: 19).

⁴ Käännösuomen korpuksen tiedoissa mainitaan nimitys supisuomi alkusuomen rinnalla. Supisuominimitystä on käytetty natiiviaineistosta joissakin aikaisemmissa opinnäytteissä. Jatkan tätä tapaa.

vien ja suomenoppijoiden käyttötapa, myös havaintoyksiköt keskenään. Koska lähtökohtanani on etukäteen valittu kielen ilmiö, tutkimustapani on korpuspohjainen. Korpusvetoisessa menetelmässä lähtökohtana on aineistosta nouseva aihe. (Tognini-Bonelli 2001: 11; Jantunen 2011: 49.)

Korpuslistausta varten valitaan WordSmith Tools (WST) -ohjelmiston Concord-toiminnossa ensin tutkittavat tekstit (Scott 2008). Seuraavaksi teksteistä tehdään konkordanssihakuja merkkijonolla, joksi on kirjoitettu tässä tapauksessa koko ilmaisin tai useampiosaisen ilmaisimen alkuosa. Halutessani tietää, onko astemääritteillä suffikseja, olen liittänyt merkkijonon perään asteriskin. Taulukossa 1 on esimerkki konkordanssihaulla tuotetusta tekstirivistä, josta nähdään (haku)sana eli **noodi** (tässä *lähestulkoon*) eri muotoineen ja siihen yhdistyvät sanat sen tekstuaalisessa lähiympäristössä.

TAULUKKO 1. Otos konkordanssista supisuomen osakorpukselta haulla *lähestulkoon**.

1,035	i määrätietoinen, myötätuntoinen ja <i>lähestulkoon</i> kaikkietävä, kuten
1,036	a heidän kanssaan. Vieraat nähtyäni <i>lähestulkoon</i> peräännyin takaisin n
1,037	yön hetkenä. Jallua pikarissa, hän <i>lähestulkoon</i> itki ajatellessaan ve
1,038	Viljeltyä nurmilaidunta käytettiin <i>lähestulkoon</i> yksinomaisesti vain E
1,039	yrityä väittää, että tämä selittäisi <i>lähestulkoonkaan</i> tietoisuuden syvä
1,040	van siinä, ettei vielä ole keksitty <i>lähestulkoonkaan</i> kaikkia terapiamu
1,041	lakaistu pois. Se ei tietysti ollut <i>lähestulkoonkaan</i> ainoa, joka joutu

Konkordanssit paljastavat toistuvuudet, ja niistä voidaan havaita säännönmukaisuuksia ja syntagmaattisia⁵ rajoituksia, jotka yksittäisiä tekstejä analysoitaessa jäisivät piiloon. Konkordansseista myös nähdään, millaiset kollokaatit esiintyvät taajaan eli ovat frekventtejä. (Jantunen 2009: 358.)

Vierussanavertailuni tukena käytän WST-ohjelmasta automaattisesti saatavaa ja testaukseen perustuvaa kollokaattilistausta (tarkemmin s. 7). WST:n laskemista kollokaateista (liitteet 3 ja 5) voin verrata, ovatko aineistoissa (liitteet 2 ja 4) eniten käytetyt vierussanat myös kollokaatteja. Teen vertailua, vaikka käsitellenkin itse vierussanoja **lemmoina** eli eri sananmuodot sisältävinä perusmuotoina, kun taas ohjelma laskee eri sa-

⁵ Syntagmaattisia suhteita ovat tarkastelemani kollokaation lisäksi mm. yhdyssanat, syntagmaattiset sulautumat (*vaikkei*), muuttumiseltaan jähmeät monisanaiset yksiköt, kuten liittokonjunktiot ja niitä muistuttavat useasanaiset kombinaatiot (*ennen kuin, sillä aikaa kun*) sekä idiomit (*tuottaa harmaita hiuksia*) (Jantunen 2009: 358; ISK 2004 § 125).

nanmuodot erikseen. Käytän vertailuun myös ohjelman muodostamia toistuvia sanaketjuja eli **klustereita**.

Vertailuni on **intralingvaalista**, sillä vertaan saman kielen kahta eri varianttia, natiivi- ja oppijansuomea toisiinsa. Intralingvaalista olisi myös eri lähtökielisten tapa käyttää samaa kohdekieltä. Nämä molemmat vertailuasetelmat ovat tyypillisiä Grangerin edustamassa kontrastiivisessa välikielen analyysissä (*contrastive interlanguage analysis*), joka keskittyy epätyypillisyyksien havainnointiin oppijan tuotoksista. Oppijankielen epätyypillisyyksillä ei tarkoiteta ainoastaan virheitä, vaan myös sanojen, fraasien tai rakenteiden yli- tai aliedustumista oppijan teksteissä. (Granger 2001: 8–9.)

Aiemmin mainitsin Jantusen (2004) tutkimuksesta, jonka havaintoyksikköinä olivat astemääritteet *hyvin*, *kovin* ja *oikein*. Hän on tutkinut korpuslingvistisesti oppijankielestä myös *aivan*-lekseemiä (2009). Vaikka oman tutkimukseni havaintoyksiköitä *lähes* ja *melkein* ei käytetäkään täysin samoin kuin Jantusen havaintoyksiköitä, kaikkia mainittuja lekseemejä yhdistää astemääritteille ominainen valinnaisuus lauseessa (ISK 2004 § 439). Empiirinen tarkasteluni on lähtökohdiltaan kvalitatiivista toisin kuin Jantusen tilastolliseen testaukseen perustuva kvantitatiivinen metodi. Kollokaatti-, kolligaatti- ja klusteritarkastelussani kuitenkin tukeudun hänen tarkastelutapaansa. Tuloksia havainnollistan taulukoimalla esiintymiskertoja ja prosentuaalisia osuuksia.

2. TEOREETTINEN VIITEKEHYS

Tutkin valitsemiani likimäärän ilmaisimia autenttisissa oppimistilanteissa syntyneistä teksteistä. Tällä tavoin osana esiintymisympäristöään kielen ilmiötä tarkastellaan kontekstuaalisessa semantiikassa, jonka ydinajatuksista kerron luvun aluksi. Suuntauksen tunnuslause, Firthin (1957) ”*You shall know a word by the company it keeps*”, mainitaan lähes kaikissa alan lähteissä. Vaikka Firthin merkitysteoriat ovat saaneet kritiikkiäkin, korpuslingvistiikan peruskäsitteistöön ovat jääneet häneltä mm. kollokaatio ja kolligatio, jotka määrittelen seuraavaksi. (Jantunen 2004: 7, 11, 15–16; Sinclair 1987: 319.) Lopuksi luon katsauksen oppijankielen korpustutkimukseen, jossa sovelletaan kontekstuaalista semantiikkaa ja sen käsitteistöä.

2.1. Kontekstuaalinen semantiikka

Korpuslingvistisen tutkimukseni taustalla vaikuttaa niin sanotun brittiläisen koulukunnan (J. R. Firth, M. A. K. Halliday, John Sinclair) käyttämä kontekstuaalinen lähestymistapa ilmaukseen. Kontekstualistien mukaan sanan merkitys syntyy todellisissa käyttötilanteissa yhdistyneenä lähisanoihin. Konseptualismille päinvastaisesti kontekstualistit ajattelevat, että tilannekontekstia tarkastelemalla saavutetaan yleistettävyys, joka ei pohjaudu ainoastaan yksittäisen kielenkäyttäjän mielikuvaan ilmauksen merkityksestä. Kontekstuaalisessa semantiikassa haetaan säännönmukaisuuksia ja toistuvuutta, mutta yksittäistapauksia ei täysin hylätä. (Jantunen 2004: 7–8.)

Tutkimukseni kannalta olennaisia käsitteitä ovat kollokaatti, kolligaatti ja klusteri sekä muotti. Näiden yhtä sanaa laajempien fraseologisten yksiköiden suhteet ”realisoituvat joko tiukkoina kotekstiin suuntautuvina valintarajoituksina tai väljempinä valintapreferensseinä” (Jantunen 2009: 358). Suhteet perustuvat Sinclairin esittämään idiomiperiaateeseen (*idiom choice principle*), jonka mukaan puhuja soveltaa kielenkäyttötilanteessa aikaisemmissa elämänsä toistuvissa kielenkäyttötilanteissa omaksumiaan fraaseja tai malleja. Sanat eivät siis valikoidu käyttöön mielivaltaisesti tai satunnaisesti (*open choice principle* ’avoimen valinnan periaate’). Idiomaattinen valinta ei tarkoita ainoastaan idiomien käyttöä, vaan myös natiivin luontaisesti hallitsemia yksittäisen sanan kontekstuaalisia suhteita. (Sinclair 1987: 319–320; Jantunen 2009: 358.)

2.2. Keskeiset käsitteet

Kollokaatti eli kerasana tarkoittaa kiinteässä syntagmaattisessa suhteessa olevien sanojen (yleensä kahden) toistuvaa yhteisesiintymää, leksikaalista myötäesiintymää. Kollokaatiksi voidaan laskea epätavallisuuden tai uutuusarvon, vuoksi yksittäistapauskin. Kollokaation perusominaisuuksiin kuuluvat toistuvuuden lisäksi tilastollinen merkittävyys ja ulottuvuus. (Jantunen 2004: 16; Sinclair 1987: 325, 1998: 11; Stubbs 2007:178.)

Tilastollisesti merkitsevä on kollokaatti, joka esiintyy aineistossa taajaan ja jonka yhteisesiintymästä noodin kanssa suljetaan tilastollisin testein⁶ pois sattuma. Tilastollisen merkittävyyden laskennassa otetaan huomioon noodin ja sen kollokaatin todelliset frekvenssit ja odotetut frekvenssit tarkastelualueella sekä tarkastelualueen ja aineiston koko. Testein on voitu osoittaa sanoilla olevan sellaisiakin kollokaatiosuhteita, jotka eivät perustu syntaktisiin suhteisiin, joita ovat mm. sana- ja lausekejärjestys sekä lausetyyppi. Testin laskemat kollokaatit voivat perustua, mutteivät välttämättä perustu semanttiseen myötäesiintymään. (Jantunen 2004: 17–18; Stubbs 1995: 31.)

Ulottuvuuteen Jantunen laskee kollokationaaliseen suhteeseen kuuluvien sanojen lukumäärän, noodin ja sen kollokaatin välimatkan toisistaan, noodin ja sen kollokaatin kielipillisen suhteen sekä tarkastelualueen symmetrisyyden. Yleensä kollokaatiota on pidetty kahden sanan suhteena, ja useampisanaiset yhdistelmät (monisanaiset yksiköt, leksikaaliset niput, klusterit) on erotettu kollokaateista omaksi tarkasteluryhmäkseen. Tarkastelualue valikoituu usein noodiin nähden symmetriseksi, mutta eri sanaluokkien sanat tai eri sananmuodot voivat saada kollokaatteja kotekstistaan eri tavoin. Symmetrinen tarkastelualue merkitään esimerkiksi 2:2, jolla tarkoitetaan sitä, että kollokaatteja tarkastellaan kahden sanan etäisyydeltä sekä noodin vasemmalta että oikealta puolelta. Osa tutkijoista pitää virkerajan ylittävää kollokointia heikkona, mutta viittaussuhteiden vuoksi kollokaatteja voi tarkastella virkerajan yli. (Jantunen 2004: 18–21; 2009: 358.)

⁶ Olen verrannut todellisiin frekvensseihin perustuvia kollokaattiehdokkaitani WST-ohjelman kollokaattilistaukseen, jossa on käytetty z-testiä. Tilastolliset testit eivät välttämättä laske merkitseviksi todellisuudessa frekventisti käytettyjä sanoja, minkä vuoksi todellisiin esiintymiskertoihin perustuvien sanojen tulisi olla etusijalla kielenopetuksessa (Kennedy 2008: 39).

Kolligaatio on vakiintunut tarkoittamaan lekseimin sananmuotojen, lausekkeen tai lekseimin eri merkitysten suhdetta johonkin kieliopilliseen kategoriaan, lauseenalkuisuuteen tai lauseenloppuisuuteen kontekstissa. Kolligaatit eivät ole yhtä suoraan havaittavia kuin kollokaatit, sillä ne voivat sulautua merkitykseen, kuten Stubbsin esimerkissä yhteys kielteisyyteen voidaan tulkita erilaisista realistumista: *as not*, *hardly* ja *refuse*. (Sinclair 1998: 11; Stubbs 2007: 178; Jantunen 2004: 22–23.) Jantusen (2009: 359) tutkimuksessa astemäärite KOVIN esiintyi synonyymejään HYVIN ja OIKEIN merkittävästi useammin kvanttorin seurassa: KOVIN < kvanttori: HARVA, MONI, PALJON, VÄHÄN > .

Klusteri eli n-grammi on useammasta kuin kahdesta peräkkäisestä sanasta koostuva toistuva sanaketju, yleensä 3- tai 4-grammi. Kollokaatista klusteri eroaa sanojen peräkkäisyydellä, kun taas kollokaatti voi sijaita jopa eri virkkeessä. Klustereita voidaan tuoda näkyviin aineistosta automaattisesti esimerkiksi WST-ohjelmaketin Concord-toiminnolla. Jantusen *aivan*-adverbia koskevassa tutkimuksessa sananmuotoklustereista < *se oli aivan* > ja < *se on aivan* > muodostettiin lemmaklusteri < SE OLLA *aivan* >. (Jantunen 2004:19; 2009: 359, 368.) **Lemma** on sanakirjan hakusana, joka toteutuu eri sananmuotoina. Se merkitään yleensä isoin grafeemein ja sananmuodot kursivoituina. (Stubbs 1995: 24; Jantunen 2004: 8.)

Tutkimuksessani tarkastelen vierussanoja perustuen niiden frekvensseihin aineistossa. Havaintoyksiköittäni määriteluonteen vuoksi keskityn noodin oikeanpuoleisiin vierussanoihin, jota tarkoitan merkinnällä O1. Tämän vierussanan, kollokaattiehdokkaan, on kuitenkin sijaittava samassa lauseessa. Havaintoyksiköistäni *melkein* voi päättää lauseen, silloin kun se esitetään lausumassa vastauksena tai irrallisena lisäyksenä, eräänlaisena jälkikommenttina asiaan. Tällöin O1-asemainen vierussana jää siis tarkastelun ulkopuolelle. Sen sijaan lauseenloppuisuutta voitaisiin pitää ilmaisimen *melkein* kolligaatitina. Laajennan vaihteluväliä klusterilistauksessa ja silloin, kun *melkein* ja *lähes* näyttävät määrittämänsä sanan kanssa kuuluvan muottiin (supisuomi: *melkein joka kolmannessa uutisessa*). Muodostan myös yleistyksiä sananmuotoklustereista lemmatasolle, silloin kun se vaikuttaa aiheelliselta. Kolligaatilla ilmaisen tässä sanaluokkaa. Sillä voi ilmaista myös lauseen alkuisuutta tai -loppuisuutta, mutta jätän tämän piirteen tässä tarkastelun ulkopuolelle.

2.3. Oppijankielen korpustutkimus

Oppijankielestä on käytetty useita nimityksiä, kuten välivaihekieli, vierassuomi, ulkomaalaissuomi, oppijankieli ja oppijansuomi. Niistä etenkin Selinkerin 1969 lanseeraama **välikieli** (*interlanguage*) on monitulkintaisuudellaan herättänyt keskustelua. Nimityksestä saattaa saada vaikutelman, että oppijankieli on vajavaista tai virheellistä. Toisaalta siitä näkee, että kielenoppiminen on jatkuva prosessi. Nykyisin oppijankieltä pidetään itsenäisenä kielimuotona, jonka epätyypillisyydet natiivikielen normin suhteen nähdään virheiden sijaan ominaispiirteinä. Itsekin käytän nimitystä oppijansuomi. Oppijankielen ominaisuuksia ovat systemaattinen vaihtelu tilanne- ja tehtäväkohtaisesti, yleinen kielenkehitysjärjestys ja äidinkielen ja/tai muiden aiemmin opittujen kielten vaikutus siihen. (Martin & Alanen 2011: 34; Nissilä 2011: 39–40; Jantunen 2008: 2–3.)

Korpuksella on alkuaan tarkoitettu mitä tahansa tutkimusta varten koottua kielenaineskokoelmaa. Nykyään se yleisimmin käsitetään systemaattisesti pysyvään tutkimus- ja opetuskäyttöön kootuksi ja sähköiseen muotoon tallennetuksi teksti-, ääni- tai videoaineiston kokoelmaksi. Korpus voi olla kielen muutoksia seuraava yleiskorpus tai erikoiskorpus, joka rajoittuu vaikkapa yhteen tekstilajiin. Se voi koostua kokonaisista teksteistä tai otoksista. Suljetun tai staattisen korpuksen tekstimäärä on rajattu ennalta sovitusti, kun taas avointa tai muuttuvaa korpusta (esim. ICLFI) kartutetaan. (Jantunen 2011: 87, 90–91; McEnery & Wilson 2001: 29–32; Tarnanen 2007: 198.)

Sähköisiä korpuksia alettiin koota 1960-luvulla. 1980-luvulta lähtien sähköisten korpusien käyttö on laajentunut natiivikielen tutkimuksesta oppijankielen tutkimukseen. (Granger 2002: 1.) Laajimmat oppijankielen korpuksat ovat englannista, mutta nykyisin erikielisiä oppijankielikorpuksia on koottu ns. korpusatlakseen lähes 140. 2000-luvulla toisen ja vieraan kielen korpuksia on koottu useita myös suomen kielestä. Oppijansuomea on kerätty seitsemään korpukseen, joista ICLFI on ainoa suomi vieraana kielenä korpus. AS2-, YKI-, Cefling-, Topling-, Dialuki- ja Long Second -korpuksat ovat suomi toisena kielenä -korpuksia. (Jantunen 2011: 87–88; Jantunen & Pirkola 2015.)

Korpuslingvistiikka määritellään Grangerin (2002: 1) mukaan lingvistiseksi menetelmäksi, joka kehitettiin sellaisten sähköisten kokoelmien käyttöön, jotka koostuvat luonnollisesti syntyneistä teksteistä. Sitä voidaan pitää joko lingvistiikan uutena osa-alueena

tai uutena kieliteorianana. Tietokoneavusteisen menetelmän vahvuus piilee siinä, että erilaisia tilastoja ja todennäköisyyslaskelmia voidaan tehdä aiempaa huomattavasti suuremmasta määrästä kielellistä dataa. Tietokoneavusteisten avulla voidaan saada näkyviin toistuvia rakenteita ja ennen havaitsemattomia kielellisiä ilmiöitä. Menetelmä täydentää tutkijan intuitiivista introspektiota. (Granger 2002: 1–2.)

3. LIKIMÄÄRÄN ILMAISIMET

Tässä luvussa kerron ensin yleisemmin likimäärän ilmaisusta suomen kielessä ja luonnehdin sen jälkeen lekseemejä *lähes* ja *melkein*. Havainnollistan ISK:n selitystä sen omien esimerkkien lisäksi KS:n sana-artikkeleiden ja molempien korpusaineistojen esimerkkilauseilla.

3.1. Yleistä adverbeista ja partikkeleista

Likimäärän ilmaisimia kuvataan ISK:ssa sekä adverbeja että partikkeleita käsittelevissä luvuissa. Ennen ISK:n jaottelua adverbeja sekä post- ja prepositioita pidettiin partikkelien alaryhmänä (Alho & Kauppinen 2009: 42). Partikkelien ja adverbien yhteisenä piirteenä on taipumattomuus, tosin joillakin adverbeilla on osittainen taivutusparadigma ja osa niistä komparoituu. Adverbeilla voi olla täydennyksiä tai määritteitä, joten ne voivat muodostaa lausekkeita. Lauseenjäsenenä ne ovat adverbiaaleja. Partikkelit sen sijaan eivät saa omia määritteitä, eivätkä ne voi muodostaa lausekkeita tai toimia lauseen ytimenä. (ISK 2004 § 793; Alho & Kauppinen 2009: 42, 73.)

Adverbit ovat päätteettömiä (*aina, heti*) tai päätteellisiä (esim. *-sti*) sanoja, jotka ilmaisevat lauseessa esimerkiksi aikaa, paikkaa, tapaa, syytä, olotilaa, määrää tai intensiteettiä. Intensiteettisanoista osaa, kuten *melko* ja *aika*, voi pitää myös partikkeleina. (ISK 2004 § 646–647, 792; Alho & Kauppinen 2009: 73, 85.)

Partikkelit on heterogeeninen ryhmä eri tehtävissä toimivia pikkusanoja. Sama partikkeli voi kuulua useampaan partikkeliryhmään, joita ovat muun muassa interjektiot, dialogipartikkelit, lausumapartikkelit, konjunktiot, modaali- ja sävyartikkelit sekä fokuspartikkelit ja intensiteettisanat. Taulukossa 2 on se osa ISK:n asetelmasta 129, jossa näkyvät likimäärän ilmaisimet ja niihin liittyvät liitepartikkelit. Likimäärän ilmaisimia (*aika, melko, noin, joku, jotain, siinä*) luetellaan omassa ryhmässään intensiteettisanojen joukossa. Lekseemi *melkein* mainitaan sävyartikkeleihin kuuluvana. *Melkein* ja sen merkitysvastine *miltei* voivat saada suffiksikseen sävyartikkelin *-pA(s)*. *Lähes(-tulloon)* ja *liki(-pitäen, -määrin, -main)* puolestaan voivat saada lausumapartikkeleihin kuuluvan kieltoisuuden *-kAAAn*, joka tulkitaan myöntömuodon *-kin* kanssa myös implikaation ’jo-

tain muuta' sisältäväksi inklusiiviseksi fokuspartikkeliksi⁷. (ISK 2004 § 792, 801, 829, 839.)

TAULUKKO 2. Likimäärän ilmaisimia ja niiden mahdollisesti saamia liitepartikkeleita partikkeliryhmissään. (ISK 2004 § 792: osa asetelmasta 129)

MODAALIPARTIKKELEITA tietysti, tietenkin, varmaan(kin), kai, ehkä, kaiketi, kenties, lie(nee), kyllä, toki, totta kai, tuskin, ei; muka(mas), kuu- lemma, tiemmä, näemmä, näet	SÄVYPARTIKKELEITA aina, ny(t), sit(ten), vast(a), viel(ä), -hAn, -pa(s)/pä(s), ihan, kyl(lä), oikein, oikeestaan, sitä, tosiaan, vaan, vähä(n), melkein
FOKUSPARTIKKELEITA myös, -kin, -kAAAn; jopa, peräti, vieläpä, ainoastaan, pel- kästään, vain, etenkin, varsinkin, vallankin, juuri, just; itse, ainakin, edes, saati, hädin tuskin; tasan, tarkalleen, täsmäl- leen, jo, vielä, enää	INTENSITEETTISANOJA aivan, ihan, oikein, niin, (noin, näin), hyvin, kovin, perin, sangen, varsin, ylen, kyllin, tasan, tosi, täys(in), (ei) järin ~ kovin ~ eri- tyisen
	LIKIMÄÄRÄN ILMAISIMIA aika, melko, noin, joku, jotain, siinä

Liittyessään sanaluokasta riippumatta mihin tahansa elementtiin partikkelit voivat nos-
taa jonkin lauseen osan fokukseen ja vaikuttaa lauseen kokonaistulkintaan. Sävy- ja
lausumapartikkelit sekä adverbialikonjunktiot ”ovat liitynnäissuhteessa koko lausee-
seen tai lausumaan. Intensiteettisanat sijaitsevat partikkelien ja adverbien välimaastossa,
samoin modaalipartikkelit.” (ISK 2004 § 794.)

3.2. Likimäärän ilmaisu suomen kielessä

Likipitäisyyttä, epämääräisyyttä tai summittaisuutta ilmaistaan muun muassa leksee-
meillä *lähes* ja *melkein* sekä niiden taulukossa 3 nähtävillä merkitysvastineilla. Iso suo-
men kielioppi (2004 § 661: asetelma 109) käyttää nimitystä likimäärän ilmaisun ja jakaa
likipitäisyyttä kantavat adverbit ja partikkelit neljään luokkaan: alle jäävä arvio, yli me-
nevä arvio, ala- tai yläraja ja summittaisuus. Olen täydentänyt ISK:n asetelmaa 109
muutamalla sellaisella merkitysvastineella, jotka ovat Kielitoimiston sanakirjassa (KS
s.v. *lähes*, *melkein*, *jokseenkin*), mutta mainitsematta ISK:ssa. Lisäykset on taulukossa
alleviivattu.

⁷ Inklusiivinen fokuspartikkeli siis kertoo, että on muitakin vaihtoehtoja kuin jokin tietty. Eksklusiivinen
fokuspartikkeli, kuten *vain*, rajaa pois muut vaihtoehdot, jolloin implikaationa on 'ei muuta kuin'. (ISK
2004 § 839.)

TAULUKKO 3. Likimäärää ilmaisevia adverbeja ja partikkeleita. (ISK 2004 § 661 aselma 109; KS s.v. *lähes*, *melkein*, *jokseenkin*).

<p>ALLE JÄÄVÄ ARVIO alle (~<u>alta</u>) <u>melkein</u>, <u>miltei</u>, <u>liki</u> (-main, -<u>pitäen</u>, -<u>määrin</u>), <u>lähimain</u>, <u>lähes</u>, <u>lähestulkoon</u>, <u>vähää vaille</u> <u>likimainkaan</u>, <u>likikään</u>, <u>läheskään</u>, <u>lähestulkoonkaan</u></p>
<p>YLI MENEVÄ ARVIO yli, <u>päälle</u></p>
<p>ALA- TAI YLÄRAJA al(hais)immillaan, enimmillään, korkeimmillaan, niukimmillaan, runsaimmillaan, vähimmillään, ylimmillään enintään, korkeintaan, vähintään; maksimissaan ja minimissään</p>
<p>SUMMITTAISUUS arviolta, jokseenkin, <u>jotakuinkin</u>, keskimäärin, kutakuinkin, noin, osapuilleen, suunnilleen, suurin piirtein; summassa, summa(n)mutikassa</p>

Allejäävää arviota kuvaava adpositio *alle* (~ *alta*) kuuluu samaan luokkaan lekseemien *lähes* ja *melkein* sekä niiden merkitysvastineiden kanssa. Se eroaa luonteeltaan muista ryhmänsä sanoista, jotka mainitaan täyteen asteen määritteinä. Adposition tehtävää hoitavat myös ylimenevää kuvaavat *yli* ja *päälle*. (ISK 2004 § 661.)

Numeraalien likimääräisyyttä ilmaistaan muodolla *pari*, partikkelien *noin*, *suunnilleen* tms. avulla tai johtimilla *-inen* (*kolmisen*) ja *-kunta* (*tuhatkunta*). Samassa tehtävässä toimii yhdysnumeraalityyppi *toista kymmentä* 'jonkin verran yli kymmenen'. "--suhteellisen suurta mutta tarkkaan ilmaisematonta lukumäärää" tarkoitettaessa käytetään lukusanan edellä lekseemiä *pari* tai kvanttoreita *muutama*, *joku(nen)*, *monta*, *moni*, *usea(mpi)*. Summittainen on myös vaihteluvälin ilmaiseminen elatiivi- ja illatiivimuotoisilla numeraaleilla: *kahdesta viiteen*. (ISK 2004 § 784–786.)

3.3. Havaintoyksiköt *lähes* ja *melkein*

Lekseemit *lähes* ja *melkein* sijoittuvat ISK:ssa *alle jäävä arvio* -ryhmään, mutta ilmaisimen *melkein* merkitysvastineiksi on lueteltu KS:ssa myös *summittaisuus*-ryhmään kuuluvia ilmaisimia: *jokseenkin*, *suunnilleen*, *osapuilleen*, *suurin piirtein*. (vrt. s. 12 ja liite 1.) Molempien ilmaisinten käyttötavat on jaoteltu kahteen tyyppiin. *melkein* on käytössä ensinnäkin määrän, mitan tms. ilmaukseen liittyen, kun taas *lähes* ilmaisee

yleensä pienempää määrää. Toiseksi molempia käytetään laadun ja tavan ilmaukseen liittyen, *melkein*-määritettä myös asteen ilmaisuun. (KS s.v. *lähes, melkein*.)

ISK:n mukaan *lähes* ja *melkein* merkitysvastineineen osoittavat likimäärän lisäksi osittaista täyteen astetta: 'raja tai tavoite on lähellä' (esimerkit 1 a–g). Ne ovat astemääritteitä, jotka voivat määrittää määrän ilmausta (2 a–f), adjektiivia⁸ (3 a–i), laatua ilmaisevaa substantiivina (4 a–f) ja adverbina tai adpositioliusekettana (5 a–g). Verbin edellä ne määrittävät verbin lisäksi koko lausetta (6 a–f). (ISK 2004 § 615, 661.) *Melkein* ja *lähes* esiintyvät ainoastaan myönteisissä lauseissa. Kieltoausevastineina lekseemeille *lähes* ja *melkein* ovat *juuri* ja puhekielessä *paljon*. (ISK 2004 § 661; Palander 2010.) Esitän seuraavassa esimerkkilauseet (1–7) edellä luettelemistani määritetapauksista. Liitin mukaan vastaavanlaisia tapauksia KS:sta (ks. liite 1) ja aineistoistani. ISK:n esimerkeissä käytetään myös havaintoyksiköitten merkitysvastineita.

- (1 a) Hän on **melkein** 18-vuotias. (ISK 2004 § 661)
 (1 b) **Lähes** 700 00 henkeä. **Lähes** kokonaan. (KS s.v. *lähes*)
 (1 c) Luin kirjan **melkein** loppuun. (Palander 2010)
 (1 d) --**melkein** kymmenvuotiaat eivät saaneet vielä päättää -- (SUPISUOMI)
 (1 e) --tulostusliuskoilla luku on **lähes** 1000 kilometrin mittainen. (SUPISUOMI)
 (1 f) Hänellä on yksi poika, joka on **melkein** 4-vuotias ja -- (VI0359⁹) (ICLFI)
 (1 g) Maksaa vähän kalliisti **lähes** 500 yuania. (KI0001) (ICLFI)
- (2 a) **miltei** viisi (ISK 2004 § 661)
 (2 b) **Melkein** sata, 30 000. **Lähes** sata. (KS s.v. *lähes, melkein*)
 (2 c) -- kansanauto saattoi maksaa **melkein** kaksisataatuhatta? (SUPISUOMI)
 (2 d) -- osuus kauppavaihdosta oli ollut **lähes** kolmanneksen-- (SUPISUOMI)
 (2 e) On **melkein** 20 astetta. (VI0057) (ICLFI)
 (2 f) --punaisiin joukkoihin kuului **lähes** 30 000 ihmisiä. (PU0005) (ICLFI)
- (3 a) **lähestulkoon** valmis (ISK 2004 § 661)
 (3 b) --hetken ajan hän oli **melkein** onnellinen. (ISK 2004 § 615)
 (3 c) --kertoi hänestä **melkein** ensimmäisenä asiana äidilleen, (ISK 2004 § 615)
 (3 d) **Lähes** samanlainen, mahdoton. (KS s.v. *lähes*)
 (3 e) **Melkein** valmis. **Melkein** valkoinen. (KS s.v. *melkein*)
 (3 f) Kaikki kolme naista olivat kiihtyneitä, **lähes** hysteerisiä. (SUPISUOMI)
 (3 g) --se oli kosteaa ja lämmintä - **melkein** kuumaa. (SUPISUOMI)
 (3 h) --opiskelun jälkeen itsenäistytään on **lähes** mahdoton. (IT0025) (ICLFI)

⁸ Hakulinen ja Karlsson (1995: 75, 84, 134–137) käyttävät adjektiivina määrittävästä adverbista nimitystä ad-adjektiivi.

⁹ Kirjoittajan yksilöivässä koodissa (VI0359) mainitaan ensin kielenoppijan maa (tässä Viro), sitten kirjoittajan muista samanmaalaisista erottava numero. Jos luvun jäljessä olisi merkittynä esimerkiksi *b*, se viittaisi siihen, että kirjoittajalta on useampia tekstejä, ja tämä on hänen kolmas tekstinsä. (ICLFI.)

(3 i) Hän olin **melkein** yksinäinen. (SA0042) (ICLFI)

(4 a) **likimain** syntiä (ISK 2004 § 661)

(4 b) On jo **melkein** aamu. (KS s.v. *melkein*)

(4 c) --he ovat **lähes** ystäviä,-- (SUPISUOMI)

(4 d) --sellaisena päivänä voi tapahtua **melkein** ihmeitä. (SUPISUOMI)

(4 e) Oli jo **melkein** kesä,-- (SUPISUOMI)

(4 f) On **melkein** myrsky koko ajan,-- (RU0011) (ICLFI)

[**lähes**: puuttuu aineistosta] (ICLFI)

(5 a) **lähes** päivittäin ~ kahden viikon kuluttua (ISK 2004 § 661)

(5 b) **Lähes** kokonaan. **Lähes** lopullisesti. **Lähes** keskellä. (KS s.v. *lähes*)

(5 c) **Melkein** aina. **Melkein** alusta asti. (KS s.v. *melkein*)

(5 d) --kokonaismäärä on pysynyt Suomessa **lähes** ennallaan. (SUPISUOMI)

(5 e) Opettaja on jättänyt minut **melkein** joka päivä laiskaan. (SUPISUOMI)

(5 f) --**lähes** heti toisessa ääripäässä -- (PU0005) (ICLFI)

(5 g) -- nousi ja nousee edelleenkin) **melkein** päivittäin (VI0404) (ICLFI)

(6 a) Oltiinhan jo **melkein** päätetty --. (ISK 2004 § 661)

(6 b) **Melkein** säärittää tuollainen. Tuntui **melkein** siltä, että -- (KS s.v. *melkein*)

(6 c) Hän saattoi **melkein** kuvitella, että myös (SUPISUOMI)

(6 d) -- voi **lähes** kuolla,-- (SUPISUOMI)

(6 e) **Lähes** rupean tuntemaan itseni - - (PU0015) (ICLFI)

(6 f) --jossa Vappu unohti tai **melkein** unohti äidin asian. (KI0016) (ICLFI)

(7 a) Miksi sanaa nainen ei voi käyttää **läheskään** samoin kuin mies-sanaa. (ISK 2004 § 661)

(7 b) Eivät **läheskään** kaikki. Ei riitä **läheskään**. (KS s.v. *lähes*)

(7 c) Ei **läheskään** niin huono kuin luultiin. (KS s.v. *lähes*)

(7 d) Tätä ruokavalio ei pysty **läheskään** aina tarjoamaan-- (SUPISUOMI)

(7 e) Eikä tässä **läheskään** kaikki. (SUPISUOMI)

[puuttuu aineistosta] (ICLFI)

Palanderin (2010) mukaan sekä *lähes-* että *melkein-*lekseemin käyttö kieltolauseessa on leviämässä puhekielestä kirjoitettuun kieleen. Hän arvelee sen johtuvan vieraiden kielten, lähinnä englannin ja ruotsin, vaikutuksesta¹⁰. Esimerkit 8 a–d ovat tästä ilmiöstä.

(8 a) **Ei** voi verrata **melkein** muihin maihin. (TV-ohjelma)

(8 b) Täällä **ei** ole **melkein** ketään. (Puhelinkeskustelu)

(8 c) **Ei** **lähes** muita oireita. (Internet)

(8 d) Teitä **ketään** minä **lähes** en tunne. (Tiedotustilaisuus)

Yleistyvyydestä huolimatta pidän tätä kielenpiirrettä tässä työssä epätyypillisyytenä, jollaisten esiintymistä mainitsin johdantoluvussa havainnoivani aineistoista.

¹⁰ Palanderin (2010) esimerkit englannista: *Almost no-one* came to the party. I haven't spent *nearly* long enough here. Ruotsin *nästan* on käytössä sekä myöntö- että kieltolauseissa: *Nästan* alla såg henne. Sådant hander *nästan* aldrig.

4. ANALYYSI

Tarkastelen analyysini aluksi niiden likimäärän ilmaisinten frekvenssejä, jotka luettelen taulukon 3 (s. 12) ryhmissä *summittaisuus* ja *alle jäävä arvio*. Ryhmästä *alle jäävä arvio* jätin pois lekseemin *alle* (~*alta*), sillä halusin mukaan sellaisia likimäärän ilmaisimia, joiden oletan määrittävän useammanlaisia sanoja. *alle* näyttää eroavan muista ryhmänsä sanoista siinä, että se voi ulottua ylärajasta kauemmas kuin täyteen astetta kuvaava ja rajaa tavoitteleva *melkein* merkitysvastineineen. Poistettujen ryhmien ilmaisut vaikuttavat myös konkreettisemmilta kuin vaikkapa summittaiset *noin* tai *jokseenkin* merkitysvastineineen. Lähempää kollokaatti- ja muottivertailua teen tämän jälkeen lekseemeistä *lähes* ja *melkein*.

4.1. Likimäärän ilmaisinten frekvenssitarkastelu

Taulukossa 4 esitän valitsemieni likimäärän ilmaisinten yleisyyden (F)¹¹ eli kuinka monta kertaa lekseemi on mainittu tutkimusaineistoissa. Taulukosta havaitaan, että kaikki supisuomen aineistossa edustuneet likimäärän ilmaisimet eivät edustu ICLFIssä eivätkä Taajuussanastossa¹². Suomenoppijat ovat kuitenkin käyttäneet useampia ilmaisimia kuin TS:ssa on mainittu ja suunnilleen kahta kolmasosaa (13/21) natiivien käyttämistä ilmaisimista.

Havaintoyksiköiksi valitsemani *lähes* ja *melkein* ovat taajimmin esiintyviä vertailuaineistoissa niihin merkitysvastineisiinsa verrattuina, jotka on lueteltu ryhmässä *alle jäävä arvio* (ks. s. 12). TS:ssa *lähes* on yleisempi kuin *melkein*, mutta supisuomen korpuksessa ja ICLFIssä niiden keskinäinen järjestys on päinvastainen. *noin* on yleisin niin vertailuaineistoissa kuin TS:ssakin. Kaiken kaikkiaan järjestys pääpiirteissään on kaikissa aineistoissa samansuuntainen kahdeksaa yleisintä ilmaisinta tarkasteltaessa. *suunnilleen* on suomenoppijoilla käytetympi kuin *keskimäärin*, kun taas natiiveilla *keskimäärin* on käytetympi. *jotakuinkin* sijoittuu oppijoilla korkeammalle kuin natiiveilla, ja

¹¹ F = sanan esiintymiskertojen määrä aineistossa eli absoluuttinen frekvenssi (Saukkonen ym. 1979: 8)

¹² Taulukoin likimäärän ilmaisinten edustuman Taajuussanastossa supisuomen ja ICLFI:n rinnalle. Lekseemien *lähes* ja *melkein* yleisyys TS:ssa oli yksi perusteistani havaintoyksiköiden valinnalle. TS:a on käytetty opinnäytteissä ja tutkimuksissa, vaikka sen ilmestymisajankohdasta 1979 onkin aikaa ja yleisyysjärjestykset saattavat olla nykyisin muuttuneet.

TS:ssa se on viimeisenä. Muita selkeitä eroja ovat seuraavat: *suurin piirtein* ja *arviolta* edustuvat sekä ICLFIssä että supisuomessa, mutteivät TS:ssa. ICLFIssä yhden osuman saanut *likimäärin* ei edustu supisuomessa eikä TS:ssa. Vähiten supisuomessa esiintyneet ilmaisimet (*osapuilleen, lähestulkoon, likimainkaankaan, lähestulkoonkaan, summa(n)mutikassa, likipitäen, vähää vaille*) eivät edustu lainkaan TS:ssa ja ICLFIssä.

TAULUKKO 4. Likimäärän ilmaisinten edustuminen supisuomen korpuksessa, ICLFIssä ja Suomen kielen taajuussanastossa (Saukkonen ym. 1979).

SUPISUOMI			ICLFI			TS		
Ilmaisim	F	≈ %	Ilmaisim	F	≈ %	Ilmaisim	F	≈ %
<i>noin</i>	1459	31,0	<i>noin</i>	966	60,6	<i>noin</i>	335	43,3
<i>melkein, -pä</i>	1135 (1097/38)	24,2	<i>melkein</i>	504	31,6	<i>lähes</i>	157	20,3
<i>lähes</i>	966	20,5	<i>suunnilleen</i>	30	1,9	<i>melkein</i>	95	12,3
<i>miltei, -pä</i>	310 (302/8)	6,6	<i>lähes</i>	24	1,5	<i>miltei</i>	58	7,5
<i>keskimäärin</i>	255	5,4	<i>suurin piirtein</i>	16	1,0	<i>keskimäärin</i>	53	6,8
<i>jokseenkin</i>	103	2,2	<i>miltei</i>	6	0,4	<i>suunnilleen</i>	31	4,0
<i>suunnilleen</i>	101	2,1	<i>jokseenkin</i>	4	0,3	<i>jokseenkin</i>	25	3,2
<i>suurin piirtein</i>	86	1,8	<i>jotakuinkin</i>	2	0,1	<i>likimain</i>	9	1,2
<i>liki</i>	82	1,7	<i>keskimäärin</i>	2	0,1	<i>liki</i>	4	0,5
<i>läheskään</i>	61	1,3	<i>liki</i>	2	0,1	<i>kutakuinkin</i>	4	0,5
<i>arviolta</i>	44	0,9	<i>arviolta</i>	1	0,06	<i>jotakuinkin</i>	3	0,4
<i>kutakuinkin</i>	30	0,6	<i>kutakuinkin</i>	1	0,06			
<i>likimain</i>	26	0,55	<i>likimain</i>	1	0,06			
<i>jotakuinkin</i>	14	0,3	<i>likimäärin</i>	1	0,06			
<i>osapuilleen</i>	14	0,3						
<i>lähestulkoon</i>	11	0,2						
<i>likimainkaankaan</i>	5	0,1						
<i>lähes-tulkoonkaan</i>	3	0,06						
<i>summa(n)-mutikassa</i>	3	0,06						
<i>likipitäen</i>	2	0,04						
<i>vähää vaille</i>	1	0,02						
Yhteensä	4711	100	Yhteensä	1593	100	Yhteensä	774	100
Sanamäärä korpuksessa	3 773 025		Sanamäärä korpuksessa	969642		Sanamäärä korpuksessa	408301	
Ilmaisinten osuus koko sanamäärästä		≈ % 1,2	Ilmaisinten osuus koko sanamäärästä		≈ % 0,2	Ilmaisinten osuus koko sanamäärästä		≈ % 0,2

Ilmaisinten osuus aineiston koko sanamäärästä on prosenttiyksikön suurempi supisuomessa kuin TS:ssa ja ICLFIssä.

4.2. Tapaus 1: *melkein*

Tarkastelen ensin molemmille aineistoille yhteisiä vierussanoja ja nostan niistä joitakin keskeisiä tapauksia käsittelyyn. Toimin samoin myös erottavien kollokaattiehtokaiden suhteen. Klustereista korostan yhteisiä piirteitä, ja tarkastelen lekseemin *melkein* ja muutamien molemmissa aineistoissa yleisimpiin kuuluvien kollokaattien osallistumista muotteihin.

4.2.1. Kollokaattitarkastelu

Kollokaattitarkasteluun olen ottanut supisuomessa vähintään viisi kertaa edustuvat 30 lekseemiä. Halusin valita ICLFIstä suunnilleen saman verran osumia (31), jolloin minimiesiintymismääräksi tuli kolme. Jos olisin pitänyt esiintymiskertojen vähimmäismääränä viittä, vertailuun olisi tullut ainoastaan ICLFIstä 15 kollokaattiehtokasta. WordSmith Tools -ohjelma laski kollokaateiksi myös yksittäisesiintymiä.

Seuraavassa taulukossa 5 esitän lekseemin *melkein* kollokaattiehtokkaat, ts. O1- asemassa esiintyvät vierussanat, jotka ovat molemmille aineistoille yhteiset (ks. liite 2). Olen liitteeseen 3 koonnut yleisyysjärjestykseen WST-ohjelman *z*-testauksella laskemat kollokaatit molemmista aineistoista. Esitän verrokkeiksi niistäkin tässä yhteiset.

TAULUKKO 5. Supisuomen ja ICLFI:n yhteiset kollokaattiehtokkaat sekä WST-ohjelman *z*-testauksella kollokaateiksi laskemat sanat O1- asemassa. (Liitteet 2 ja 3.)

Yhteiset kollokaattiehtokkaat (lemma ¹³)	Yhteiset kollokaatit WST (sananmuotoina)
aina, heti, joka, JOKAINEN, KAIKKI, KAKSI, koko, kokonaan, KOLMEKYMMEN-TÄ, kuin, MAHDOTON, mitä, PUOLI, SAMA, UNOHTAA, yhtä	kaikki, kuin, aina, joka, koko, mitä, kaksi, mitään, mitä

Taulukosta voi ensinnäkin nähdä, että yhteisissä kollokaattiehtokkaissa on useita universaalisia kvanttoripronomineja¹⁴: *kaikki, koko, jokainen, joka*¹⁵. Niistä yleiskäyttöisin

¹⁴ ”Universaalit kvanttoripronominit ilmaisevat, että sanottu pätee koko puheena olevaan joukkoon” (ISK 2004 § 750).

¹⁵ *jokainen* ja *joka* ovat universaaleista kvanttoripronomineista distributiivisia: vaikka ilmaistu koskeekin kaikkia joukon jäseniä, ”asiaa tarkastellaan yksittäisten jäsenten kannalta” (ISK 2004 § 752).

ISK:n mukaan on *kaikki*, minkä huomaa myös aineistoista, sillä *kaikki* eri sananmuotoineen on selkeästi yleisin sekä natiiveilla että suomenoppijoilla. Suomenoppijat myös käyttävät sitä sijamuodoissa yhtä varioivasti kuin natiivitkin. Toisin kuin muut universaaliset kvanttoripronominit *kaikki* voi kvantifioida myös jaollisia entiteettejä eli ainetta (kiittää *kaikesta*, *kaikki vesi*, *kaikkea häpeää*). (ISK 2004 § 750.)

Selkeästi yhteisiä ovat taulukossa 5 myös ajan adverbit *aina* ja *heti* (§ 649), vertailuun semanttisesti yhdistyvät nomini SAMA (§ 766) sekä konjunktio *kuin* (§ 813, 819) ja *yhtä* (§ 819). Kollokaatteina molemmilla edustuvat peruslukusana KAKSI, yhdysnumeraali KOLMEKYMMENTÄ ja murtolukuinen PUOLI, jotka voitaisiin ilmaista kolligaattina: *melkein* <NUMERAALIT>.

Selkeimmät eroavaisuudet aineistojen välillä kollokaattiehdokkaissa liittyvät verbeihin ja kieltoon. Supisuomessa listalla on viisi verbiä kun taas ICLFIssä vain yksi, joka kyläkin on supisuomen kanssa yhteinen UNOHTAA. Lukusanoja ICLFIssä on puolta enemmän (6) kuin supisuomessa (3). (ks. liite 2.) Kieltohakuiset kvanttoripronominin *mitään* (ISK 2004 § 757) ja ajankohdan adverbien *koskaan* (ISK 2004 § 649), WST laskee ICLFIssä kollokaatiksi. Supisuomessa molemmat esiintyvät kerran, ja kollokaatiksi WST laskee niistä toisen, *mitään*. (ks. liitteet 2 ja 3.)

Taulukosta 5 myös nähdään, että WST-ohjelma vahvistaa yhteisiksi kollokaateiksi vain osan. Liitteistä 2 ja 3 voidaan verrata, kuinka paljon todellisiin vierussanaesiintymiin perustuva kollokaattilistaus eroaa WST:n kollokaattilistauksesta. Eroihin vaikuttaa se, että sananmuototasolla sama lekseemi voi asettua kollokaatiksi useamman kerran. Esimerkiksi yhteiseksi kollokaattiehdokkaaksi esitetty SAMA ei osoittaudu kollokaatiksi testauksessa esiintyessään eri aineistoissa eri sananmuodossa. (liitteet 2 ja 3.)

4.2.2. Klusterit ja muotit

WST listaa automaattisesti aineistosta klusterirakenteita. Esitän seuraavaksi molempien aineistojen listaukset, joista olen lihavoanut yhteiset.

TAULUKKO 6. *melkein*-määrittteen klusterilistaukset supisuomen ja ICLFI-aineistoista

SUPISUOMI	ICLFI	YHTEISET (myös lemmana)
<i>melkein</i> joka päivä oli jo <i>melkein</i> se oli <i>melkein</i> se on <i>melkein</i> oli ollut <i>melkein</i> oli <i>melkein</i> kuin <i>melkein</i> saman tien hän oli <i>melkein</i> niin että <i>melkein</i> on <i>melkein</i> aina on <i>melkein</i> yhtä	<i>melkein</i> joka päivä <i>melkein</i> koko ajan on <i>melkein</i> aina se on <i>melkein</i> hän on <i>melkein</i>	<i>melkein</i> joka päivä se on <i>melkein</i> se oli <i>melkein</i> + se on <i>melkein</i> > se OLLA <i>melkein</i> hän oli <i>melkein</i> + hän on <i>melkein</i> > hän OLLA <i>melkein</i> on <i>melkein</i> aina

Taulukosta 6 hahmottuu selkeästi viisi yhteistä klusteritapausta. Yhteen niistä, <*hän* OLLA *melkein*>, olen lemmannut eri aineistoista <*hän oli melkein*> (supisuomi) + <*hän on melkein*> (ICLFI). Jantusen tutkimuksessa vastaavia koskien lekseemiä *aivan* olivat: *hän oli aivan* ja osittaisen lemmaklusterin:

se on aivan + *se oli aivan* > SE OLLA *aivan* (Jantunen 2009: 368).

Kollokaatteina mainituista universaaleista kvanttoripronomineista *koko*, *jokainen* ja *joka*, toisinaan myös *kaikki* voi päätellä, että ne määrittävät pitempää ilmausta eli ne muodostavat jonkin toisen elementin kanssa lauseessa lausekkeen. Likimäärän ilmaisimen kanssa ne voisivat muodostaa muotin¹⁶ <*melkein* koko/JOKAINEN/joka/KAIKKI x >, jossa x on kaikkien pronomien yhteisesti määrittämä sana.

Taulukosta 7 nähdään, että suomenoppijoilla on samanlaista käyttöä kuin natiiveillakin näissä rakenteissa. Semanttisesti omaan jatkotarkasteluunsa voisivat nousta muoteissa <*melkein* joka/koko/KAIKKI x > erilaiset aikaan liittyvät ilmaisut, mutta myös etenkin oppijoilla ihmistarkoitteiset sanat x:n paikalla lukuun ottamatta lekseemiä *koko*. Natii-veilla ihmistarkoitteiset sanat painottuvat muotin <*melkein* KAIKKI x> osaksi. Univer-
saaleista kvanttoripronomineista ISK:ssa (2004 § 751) mainitaan erityisesti lekseemin *koko* esiintyvän ajan ilmauksissa. Myöskin WST:n klusteriksi laskema <*melkein* *joka päivä*> edustuu molemmista aineistoissa sellaisenaan.

¹⁶ Lisäyksenä aiempaan: Muotti on tietynhahmoinen ja ”se koostuu kiinteistä ja vaihtuvista elementeistä” (Alho & Kauppinen 2009: 23).

TAULUKKO 7. Esimerkkejä eri muoteista, joihin ilmaisin *melkein* osallistuu.

SUPISUOMI	ICLFI
<i>melkein joka</i> < talvi/päivä (10)/kolmannessa uutisessa/yö (2)/aamu/TOINEN (2)/näytöksessä /mänyssä/iltapäivä/ilta/kylässä/sadussa	<i>melkein joka</i> < filmi/päivä (21) /viikko (2)/ lapsella/aikuinen/suurella kaupungissa/aiheessa/ viikonloppuna/maissa/kerta/illalla /aamu/ vuosi/PAIKKA (2)
<i>melkein KAIKKI</i> < eläimillä/POIKA (2)/ keskusteluissa/ elämän perusasioista/ viikonloput/IHMINEN (2) /se/valtiat/ kaksineuvoiset lajit/karjataloustuotteet/ kaupunkilaiset/ huonekalut/ piparit/ pesuaineet/pyykit/hänen luokkansa tytöt/aikoihin	<i>melkein KAIKKI</i> < haastatellut saamelaiset/i ihmiset/ miehet/suomalaiset/kiinalaiset /henkilöä/kaupat/rahat/tavara/ tärkeimmät tapahtumat/paperit/ruokia/ kilpailuissa/muissa suhteissa/tutkituista/eläimistä
<i>melkein koko</i> < yön (3), ajan (4) / kesä/ matkan/ kyljen/1980-luvun/leiriaikansa/Kannas/luokka/ joukkue/räystään/kalan/Ilmarinpuisto/ olemassaolonsa ajan/nokka/	<i>melkein koko</i> < PÄIVÄ (10), AIKA (15), ELÄMÄ (3)/ yön (3), kesän, Suomelle/ matkan ajan/ ryhmäni/historian varrella/KAUPUNKI (2)/loma/heinäkuun
<i>melkein JOKAINEN</i> < meistä/puhelu/kumottu väite/maapallon kolkkiaan/ikkunalla/ kirkkomaille/kadunkulmassa/näkyvästä pesästä	<i>melkein JOKAINEN</i> <ihminen (3), asia, opiskelija/ baarilla/järkevän henkilön/suurella kaupassa/SANA (2)

Oppijoiden kollokaattivalinnoista epätyypillisiltä vaikuttavat *niin* (esimerkit 9 a–b) ja *vain* (esimerkit 10 a–b), joiden yhteyteen natiivit eivät likimäärän ilmaisinta valinneet.

- (9 a) *Tietokone on melkein niin rakas hänelle kuin minä!* (VI0308)
 (9 b) *...melkein niin ruma kuin sinä.* (PU0026)
 (10 a) *Olen melkein vain kiinnostunut suomalaista musiikkia.* (HO0011)
 (10 b) *Hänellä on melkein vain edustavat toimintot.* (SA0104)

Natiivi käyttäisi esimerkkilauseissa 9a ja 9b sanan *niin* sijaan vertaavaa partikkelia *yhtä*. Esimerkkilauseiden 10 a ja 10 b vieras vaikutelma saattaa tulla jo siitä, että natiivi ei vain käytä sanoja *melkein* ja *vain* yhdessä.

4.3. Tapaus 2: *lähes*

Kollokaattitarkastelu lekseemin *lähes* osalta vaikuttaa ongelmalliselta, sillä se saa osumia ICLFIssä vain 24, mikä on 1,5 % kaikista suomenoppijoiden käyttämistä likimäärän ilmaisimista, kun taas natiiviaineistossa *lähes* saa osumia 966, ja prosenttiosuus on n. 20,5 % kaikista supisuomen ilmaisimista. Osaksi tämä selittyy aineistojen sanamäärien epätasapainolla, osaksi sillä, että kaikista käyttämistään ilmaisimista suomenoppijat käyttävät enemmän lekseemejä *noin* ja *melkein*. Supisuomen osalta yli viisi osumaa saaneita kollokaattiehdokkaita on 32, kun taas ICLFIssä olen laskenut 15 ehdokkaan joukkoon yhdenkin osuman aineistossa saaneet vierussanat.

4.3.1. Kollokaattitarkastelu

Määritteellä *lähes* on molemmille aineistoille yhteisiä kollokaattiehdokkaita vain neljä, KAIKKI, JOKAINEN, *koko* ja MAHDOTON (ks. liite 4), jotka olivat yhteisiä vierussanoja molemmille aineistoille määritteen *melkein* tarkastelussakin. WST-ohjelman perusteella lekseemille *lähes* saa ainoastaan yhden yhteisen kollokaatin, joka on *kaikki* (liite 5). Sekä natiivit että suomenoppijat ovat käyttäneet sitä eniten ilmaisimen *lähes* kollokaattina. Kuten aiemmin määritteen *melkein* analyysissä totesin, tämä selittyy *kaikkilekseemin* yleiskäyttöisyydellä. Muoto *jokaisessa*, joka on ainoa osuma lekseemin *jokainen* sananmuodoista ICLFIssä, ei olisi sijoittunut supisuomessa yleisimpien joukkoon lainkaan, sillä *jokaisessa* esiintyi sielläkin ainoastaan kerran.

Yhtäläisyyksistä voidaan havaita aiemmin mainittu adjektiivin määrittäminen, sillä molemmissa aineistoissa kolligaattina on vähintään yksi adjektiivi. Useat lukusanat nousevat natiiviaineistossa kollokaatiksi WST:nkin mukaan. Merkitseväksi kolligaatiksi lukusanat nousevat sanaluokkana numeraali¹⁷. Kaikkiaan lukusanoja supisuomessa oli kaikkiaan 189 ja ICLFIssäkin 4 osumaa. (liitteet 4 ja 5.)

Natiivikäyttöön verrattuna oppijat näyttävät onnistuvan ilmaisimen *lähes* käytössä hyvin. Henkilökohtaisen natiivituntuman ja aineistojenkin perusteella epätavalliselta astemääritteen käyttö kuulostaa esimerkeissä 11–13, jotka ovat eri suomenoppijoilta.

(11) *Presidentti on **lähes kuin** kuningas.* (PU0002)

(12) *Mitä epäoikeudenmukaisuus! **Lähes rupean** tuntemaan itseni* (PU0015)

(13) *tämä kauan odotettu ilta **veti lähes loppuun** --* (IS0004)

Ensimmäisessä lauseessa esitetään vertailu, joka on kieliopillisesti oikein muotoiltu, mutta havaitsemani vierauden tunne tulee useuserosta natiivikäyttöön: supisuomessa *lähes kuin* -yhdistelmä ei saa osumia lainkaan, mutta määritteen *melkein* kanssa *kuin* on natiiveilla toiseksi käytetyin vierussana (liitteet 2 ja 4). *Lähes rupean* on ymmärrettävä, mutta ongelmana vaikuttaa olevan ilmaisimen sijainti lauseessa. Likimäärän ilmaisin voi määrittää verbiä, mutta tässä tapauksessa olisi ollut luontevampaa määrittää olotilaa. Fraasilla *veti lähes loppuun* lienee tarkoitettu samaa kuin fraasilla *veti lähes täyteen*, jolloin oppijan sanavalinta ei aivan ole osunut kohdalleen.

¹⁷ Järjestysluvut lasken kuuluviksi adjektiiveihin (Alho & Kauppinen 2009: 70–71).

4.3.2. Klusterit ja muotit

Klustereita, joissa ilmaisin *lähes* esiintyy, WST muodostaa ainoastaan supisuomesta. ICLFIssä *lähes* saa osumia vähän, ja oppijat ovat käyttäneet määritettä erilaisissa yhteyksissä, minkä vuoksi ohjelma ei toistuvia rakenteita, kuten klustereita, löydä. Supisuomen klustereita vastaavia rakenteita yksittäisosumina ICLFIssä kuitenkin on, ja olen asettanut ne taulukkoon 8 natiiviosumien rinnalle.

TAULUKKO 8. Määritteen *lähes* muotti- ja klusterilistaukset supisuomen ja ICLFI-aineistoista.

SUPISUOMI klusteri	ICLFI klusteria vastaavat osumat	YHTEINEN MUOTTI
<i>lähes</i> kaksi kolmasosaa oli <i>lähes</i> yhtä <i>lähes</i> kymmenen vuotta	<i>lähes</i> 1/3 ihmisten <i>lähes</i> 30 000 ihmisiä <i>lähes</i> viisi minuttia	<i>lähes</i> NUMERAALI X

Vastaavia muotteja kuin ilmaisimen *melkein* yhteydessä, voidaan muodostaa myös ilmaisimen *lähes* vierussanojen ohjaamana, sillä universaalisista kvanttorigonomineista kollokaatteina samat ovat supisuomessa: *lähes* <KAIKKI/koko/joka/JOKAINEN x>. Sen sijaan *joka* puuttuu ICLFIstä kokonaan.

TAULUKKO 9. Esimerkkejä eri muoteista, joihin ilmaisin *lähes* osallistuu.

SUPISUOMI	ICLFI
<i>lähes</i> KAIKKI < SAIRAUS (2)/ alkoholorikkomusten/yrtytien/äidit/nuoret/miehen roolit/ varusteet/astiat/tiedot/valtiot/ rokotetuilla/ tartunnan saaneilla/ mittareilla/puolilla/ työn/aikansa/ <i>lähes</i> koko < juhannuksen/ seinämän/otsan/ väkensä /esihistorian/näkymän/ajan (2)/ omaisuutensa /Suomessa/väestön/elämä/maasta/ kivikauden/alkuvuoden <i>lähes</i> JOKAINEN < puheenvuoro/ vastaanottaja/talollinen/lahjaton raukka/nainen/ kirjoittaja/sivu / <i>lähes</i> joka < viidennellä/toinen/viikonloppu/ seitsemättä/paikassa /kerta/puolella/päivä (3) /talossa /yö /ilta/neljäs/kilpailussa /kymmenennelle /säkeistöltään	<i>lähes</i> KAIKKI < elävää, bussit, KIINALAINEN (2), suomalainen, paikkoihin, pelaajien <i>lähes</i> koko < maailmaa <i>lähes</i> JOKAINEN < artikkelissa

Yhtäläisyydet vaikuttavat ensin rajoittuvan siihen, että samalla kaavalla toimivia muotteja on mahdollista muodostaa molemmille aineistoille yhteisten vierussanojen avulla. Voidaan kuitenkin havaita, että semanttisesti jatkotarkasteluunsa voisivat sekä natiiveil-

la että suomenoppijoilla nousta muotista <*melkein* KAIKKI x> ihmistarkoitteiset sanat. Lekseemin *koko* vaikuttaa liittyvän natiiveilla myös ilmaisimen *lähes* käytössä ajan ilmauksiin, ja WST:n klusteriksi laskema <*melkein joka päivä*> edustuu tällä kertaa natiiveilla: <*lähes joka päivä*>.

4.4. Määritteiden *melkein* ja *lähes* keskinäistä vertailua

Vierussanatarkastelussa *lähes* ja *melkein* saavat natiiviaineistossa useita samoja kollokaattiehdokkaita, joista useimmat vahvistuivat kollokaateiksi WST:n testauksessa. Molemmille yhteisiä vierussanoja ovat: KAIKKI, *aina*, *koko*, *yhtä*, *joka*, SAMA, *kokonaan*, KAKSI, KOLMEKYMMENTÄ, MAHDOTON, *päivittäin*, PUOLI. Ilmaisimen *lähes* vierussanana edustuvat supisuomessa *yksinomaan* ja *poikkeuksetta* sekä lemma TÄYDELLINEN varsin taajaan, kun taas ilmaisim *melkein* saa huomattavasti taajempaan vierussanoikseen *heti*, *kuin* ja *kiinni*. (liitteet 2 ja 4.) Kolligaattierona voi pitää sitä, että *melkein* esiintyy herkemmin lauseenloppuisena ja itsenäisenä lausekkeena kuin *lähes*.

ICLFI:ssä *lähes* saa osumia 24, ja *melkein* 504, minkä vuoksi vertailua ei voi pitää luotettavana. Jos yksittäisosumat kuitenkin lasketaan kollokaattiehdokkaiksi, molempien likimäärän ilmaisinten yhteisiksi valikoituivat KAIKKI, MAHDOTON, *heti*, JOKAINEN, LOPPU, *viisi*, *koko*. Ilmaisimen *melkein* vierussanoina on määrällisen epäsuhtan vuoksi useita, joita ei *lähes*-määritteen kanssa esiinny, kuten *kello*, SAMANLAINEN, KAIKKIALLA. Yhteistä on käyttö kielteisessä yhteydessä.

Palanderin huomaama muutos ilmaisinten *melkein* ja *lähes* kieltokäyttöön natiiveilla näkyi ainoastaan ilmaisimen *melkein* kohdalla. Seuraavat esimerkit 14 a–f ovat supisuomen aineistosta.

- (14 a) Meidän äiti **ei melkein** koskaan paista lettuja.
- (14 b) --hellemekossa, siinä, missä **ei ole** melkein mitään yläosassa.
- (14 c) Mie **en melkein** uskaltanu hengittää—
- (14 d) --jota hän **ei** sitten **melkein** uskaltanut edes tervetä,..
- (14 e) **Ei** vain **melkein** vaan todella minua,--
- (14 f) **Ei melkein** vaan täydellisesti nautiskellen,--

Sen sijaan suomenoppijoilla *ei* kollokoi noodin (tässä *melkein*) kanssa etenkin asemassa O1, mistä ovat seuraavassa esimerkit 15 a–g ICLFistä.

- (15 a) -- *on melkein aina onnellinen, ja **melkein ei** koskaan vihainen* (EN0012)
- (15 b) *Mutta Suomesa **melkein ei** käytetä kehon kieltä.* (VE0204)
- (15 c) -- *mutta nyt **melkein ei** kukaan tuntee--* (SA0075)
- (15 d) --, ***melkein ei** kukaan paikallisista* (VE0126)
- (15 e) *Miksi **melkein ei** kukaan ystäväistäni* (TS0004)
- (15 f) *Rouva Peltonen **melkein ei** itse ole tietoinen* (RU0050)
- (15 g) *Minäkin syön **melkein ei** lihaakaan,--* (SA0070)

Määritteen *lähes* kieltokäytöstä on ainoana esimerkki 16:

- (16) *Mutta että **ei** onnistu **lähes ikinä**.* (HO0010)

Joillakin oppijoilla kieltokäyttö saattaa johtua mainitusta englannin tai ruotsin transferrista tai siitä, että opetuksessa ei erikseen opeteta astemääritteiden käyttöä tai ei opeteta näiden likimäärän ilmaisinten suhdetta myöntölauseeseen.

Vaikka oppijat käyttävät itse ilmaisinta yleensä oikein, seuraavissa esimerkeissä 17 a–d natiivi olisi todennäköisemmin valinnut täyteen asteen ilmaisimen *melkein* sijaan ominaisuuden asteen kohtalaista määrää ilmaisevan *melko* (ISK 2004 § 666).

- (17 a) --*lihasta ja kalasta, mitä syön melkein harvoin,--* (SA0063)
- (17 b) *Puhun vielä hyvää englantia, melkein hyvää ranskaa --* (VI0021)
- (17 c) *Puhun melkein hyvin Englantia,--* (VI0028)
- (17 d) --*viikkon kuluttua jo ymmärsimme melkein paljon--*(VI0449)

5. PÄÄTÄNTÖ

Kandidaatintutkielmassani kuvasin likimäärän ilmaisinten käyttöä natiivi- ja oppijan-suomessa. Suomenoppijoiden tekstit ovat ulkomaisissa yliopistoissa suomea vieraana kielenään opiskelleilta. Tarkensin vierussanatarkasteluni määritteisiin *lähes* ja *melkein*.

Ensimmäistä hypoteesiani vastaavasti suomenoppijat hallitsevat harvempia likimäärän ilmaisimia kuin natiivit teksteissään. Heiltä puuttuvat esimerkiksi kielteiset -kAAAn-litteiset ilmaisimet *läheskään*, *lähestulkoonkaan* ja *likimainkaan* sekä natiivienkin harvimmin käyttämät *summa(n)mutikassa*, *likipitäen* ja *vähää vaille*. *Likimäärin* oli kuitenkin yksittäismainintana oppijoilla, kun taas natiiveilla se ei esiintynyt lainkaan. Ilmaisimet *noin*, *melkein*, *lähes*, *suunnilleen*, *miltei*, *keskimäärin* ja *jokseenkin* ovat paitsi ICLFIssä ja supisuomessa, mutta myös Suomen kielen taajuussanastossa yleisimpiä.

Vierussanoista yleisintä, universaalista kvanttoripronominia *kaikki*, oppijat käyttävät molempien havaintoyksiköitteni määrittämänä yhtä varioivasti eri sananmuodoissa kuin natiivit. Osaksi tätä selittää Ison suomen kieliopin mainitsema yleiskäyttöisyys. Samoja kollokaatteja (KAIKKI, MAHDOTON, JOKAINEN, *koko*) *melkein* ja *lähes* saattoivat saada supisuomen ja ICLFI:n aineistoissa verrattuina myös toisiinsa senkin tähden, että tarkastelin vierussanoja yleistyksenä, lemmatasolla. WST-ohjelmiston tilastollinen testi vahvisti yleisimmät havaintoyksikköjen määrittämät kollokaatit. Liitteitä 2 ja 3 sekä liitteitä 4 ja 5 vertaamalla voidaan nähdä, kuinka WST:n sananmuotoina tehty laskelma osittain eroaa lemmalistauksestani.

Pääsääntöisesti oppijat onnistuvat varsin hyvin astemääritekäytössä, sillä Ison suomen kieliopin esimerkkejä vastaavia tapauksia oli helppo löytää molempien havaintoyksiköiden käytöstä. Toista hypoteesiani vastaavista epätyypillisyyksistä merkittävimpanä pidän kollokointia kieltoon. Palander pitää lekseemien *melkein* ja *lähes* yhdistymistä kieltoon yleistyvästä ilmiönä natiiveilla. Tätä tapaa havaitsin vain muutamina esimerkkeinä natiiviteksteistä, mutta kielteiset *ei* ja *mitään* samoin kuin yksittäistapaukset *ikinä*, *koskaan* ja *mitään* kollokoivat havaintoyksiköitä suomenoppijoiden aineistossa. Epätyypilliseltä vaikuttivat oppijoilla myös sellaiset lauseet, joissa natiivi olisi määritteen *melkein* sijaan valinnut partikkelin *melko*.

Klusteroitumisessa lekseemin *melkein* suhteen natiiveilla ja oppijoilla esiintyi yhtenevää käyttöä, kuten *melkein joka päivä*. Tämä yhdistyy muottirakenteeseen *melkein y x*, jossa *melkein* on kiinteä, *x* vaihtuva elementti. Väliin jäävä määrite *y* voi olla myös osittain vaihtuva, kuten tässä myös universaali kvanttoripronomini KAIKKI/JOKAINEN/koko. Muottitarkastelu jäi lekseemin *lähes* osalta viitteelliseksi, sillä sen esiintymiä ICLFIssä oli vähän. Semanttisesti muoteissa *x:n* tilalla suomenoppijoilla ja natiiveilla vaikuttivat edustuvan usein aikaan liittyvät ja ihmistarkoitteiset sanat. Tämän kollokaattitarkastelun pohjalta voisikin lähteä tarkastelemaan lekseemien *melkein* ja *lähes* kontekstissa edustuvia merkitysryhmiä, samoin kuin havaintoyksiköiden synonyymisyyden astetta.

LÄHTEET JA LYHENTEET

Tutkimusaineisto

Kansainvälinen oppijansuomen korpus (ICLFI). Oppijansuomen sähköinen tutkimusaineisto. ICLFI-hankkeessa Oulun yliopisto ja viisi muuta yhteistyöyliopistoa. (Luettu 13.11.2015)

Käännössuomen korpus. Käännössuomen sähköinen tutkimusaineisto. Alkusuomen osakorpus. Koostettu Käännössuomi ja kääntämisen universaalit: tutkimus korpusaineistolla -hankkeessa. Joensuun yliopiston kansainvälisen viestinnän laitos 1997–2001.

Kirjallisuus

ALHO, IRJA & KAUPPINEN, ANNELI 2009: *Käyttökielioppi*. SKS. Hämeenlinna: Kariston Kirjapaino Oy.

GRANGER, SYLVIAINE 2002: Bird's eye view of learner corpus research. – Granger, S., Hung, J. & Petch-Tyson, S. (toim.), *Computer Learner Corpora Second Language Acquisition and Foreign Language teaching*. Amsterdam & Philadelphia: Benjamins 3-33.

HAKULINEN, AULI & KARLSSON, FRED 1979: *Nykysuomen lauseoppia*. Suomalaisen Kirjallisuuden Seuran toimituksia 350. Helsinki: Suomalaisen Kirjallisuuden Seura.

ISK = HAKULINEN, AULI – VILKUNA, MARIA – KORHONEN, RIITTA - KOIVISTO, VESA – HEINONEN, TARJA RIITTA – ALHO, IRJA 2004. *Iso suomen kielioppi*. SKST 950. Hämeenlinna: Kariston kirjapaino Oy.

JANTUNEN, JARMO H. 2004: *Synonymia ja käännössuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin*. Joensuun yliopiston humanistisia julkaisuja n:o 35. Joensuu: Joensuun yliopisto paino.

JANTUNEN, JARMO H. 2008: Haasteita oppijankielen korpusanalyysille: oppijankielen universaalit. – Pille Eslon (toim.) *Õppijakeele analüüs: võimalused, probleemid, vajadused*. Tallinna ülikooli eesti filoloogia osakonna edetised 10, 67–92. Tallinn, TLÜ Kirjastus.

- JANTUNEN, JARMO H 2009: ”Minulla on aivan paljon rahaa” – Fraseologiset yksiköt suomen kielen opetuksessa. *Virittäjä* 109, s. 356–381.
- JANTUNEN, JARMO H. 2011: Kansainvälinen oppijansuomen korpus (ICLFI): typologia, taustamuuttajat ja annotointi.
<http://www.rakenduslingvistika.ee/ajakirjad/index.php/lahivordlusi/article/view/LV21.04/23>.
- JANTUNEN, JARMO H. & PIKOLA, SILJA 2015: Oppijansuomen sähköiset aineistot. Nykytilanne. *Virittäjä* 119, s. 88–103.
- KENNEDY, GRAEME 2008: Phraseology and language pedagogy. – Meunier, F. & Granger, S. (toim.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- KS = *Kielitoimiston sanakirja*. Kotimaisten kielten keskuksen julkaisuja 166. Helsinki: Kotimaisten kielten keskus ja Kielikone 2015.
<http://mot.kielikone.fi/mot/OUYO/netmot.exe> . (27.10.2015)
- MARTIN, MAISA 1999. Suomi toisena ja vieraana kielenä. – Sajavaara, K. & Piirainen-Marsh, A. (toim.), *Kielenoppimisen kysymyksiä*. Jyväskylän yliopisto. Soveltavan kielenoppimisen keskus. Jyväskylä: Jyväskylän yliopistopaino.
- MARTIN, MAISA & ALANEN, RIIKKA 2011: Oppijankieli – kehitystä ja virheitä. – Kalaja, P., Alanen, R. & Dufva, H. (toim.), *Kieltä tutkimassa – tutkielman laatijan opas*. Finn lectura. Tampere: Tammerprint.
- MCENERY, TONY & WILSON, ANDREW 2001: *Corpus linguistics*. Edinburgh: Edinburgh university press.
- NISSILÄ, LEENA 2011: *Viron kielen vaikutus suomen kielen verbien ja niiden rektioiden oppimiseen*. Väitös. Oulun yliopisto.
<http://herkules oulu.fi/isbn9789514296161/isbn9789514296161.pdf>
- PALANDER, MARJATTA 2010. Palauttaa ja melkein. ”*Ei melkein mitään*”. *Kielikello* 4/2010. <http://www.kielikello.fi/index.php?mid=2&pid=11&aid=2348>.
 (27.10.2015)
- SAUKKONEN, P., HAIPUS, M., NIEMIKORPI, A. & SULKALA, H. 1979. *Suomen kielen taajuussanasto*. Porvoo: WSOY.
- SCOTT, MIKE 2008: *WordSmith Tools version 4*. Oxford: Oxford University Press.
- SINCLAIR, JOHN 1987: Collocation: a progress report. – Steel, R. & Threadgold, T. (toim.), *Language Topics. Essays in Honour of Michael Halliday*. Val 2. Amsterdam: John Benjamins.

- SINCLAIR, JOHN 1998: The Lexical Item. – Weigand, Edda (toim.), *Contrastive Lexical Semantics*. Current Issues in Linguistic Theory 171. Amsterdam: John Benjamins.
- STUBBS, MICHAEL 1995: Collocations and semantic profiles. – *Functions of Language* 2, 1: 23–55. Amsterdam: John Benjamins.
- STUBBS, MICHAEL 2007: Quantitative data on multi-word sequences in English: the case of the word *world*. – Hoey, M., Mahlberg, M., Stubbs, M. & Teubert, W. (toim.), *Text, Discourse and Corpora. Theory and Analysis*. Bloomsbury Publishing.
- TARNANEN, MIRJA 2007: Testiaineistosta kielenoppijakorpukseksi. – Salo, O.-P., T. Nikula & P. Kalaja (toim.), *Kieli oppimisessa – Language in Learning*. s. 197–213. AFinLAn vuosikirja 2007. Suomen soveltavan kielitieteen yhdistyksen julkaisuja no. 65. Jyväskylä.
- TOGNINI-BONELLI, E. 2001: *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

LIITE 1. Kielitoimiston sanakirjan sana-artikkelit s.v. *melkein* ja *lähes*

KS (s.v. *melkein*)

melkein vain myönteisissä lauseyhteyksissä: vähää vaille, miltei, lähes, liki, likimain, -määrin, -pitäen, jokseenkin, suunnilleen, osapuilleen, suurin piirtein.

1. määrän, mitan tms. ilmaukseen liittyen. *Melkein sata, 30 000. Kello on melkein viisi. Melkein puoli vuotta. Melkein puolet luokasta. Melkein kaikki, jokainen. Melkein isänsä pituinen. Melkein yhtä iso kuin isänsä. Melkein aina. Melkein ahusta asti.*

2. laadun, asteen, tavan tms. ilmaukseen liittyen. *Melkein sama, samanlainen. Melkein yhtä suuret. Melkein valmis. Melkein valkoinen. Melkein lapsi vielä. Melkein heti. Melkein poikkeuksetta. On jo melkein aamu. Lumi sulaa melkein silmissä. Loma oli melkein lopussa. Melkein säälittää tuollainen. Tuntuu melkein siltä, että – –.*

Sana on taipumaton tai vaillinaisesti taipuva.

KS (s.v. *lähes*)

lähes (kielteisissä yhteyksissä: läheskään) *melkein*, *miltei*, *liki*(pitäen); vrt. *likimain*, *likimäärin*, *noin*, *jokseenkin*, *suunnilleen*, *osapuilleen*, *suurin piirtein*.

1. ilmaisemassa hieman jtk pienempää määrää. *Lähes sata. Lähes 700 000 henkeä. Lähes joka vuosi. Lähes kokonaan, puolillaan. Lähes isänsä mittainen. Eivät läheskään kaikki. Ei läheskään aina. Ei riitä läheskään.*

2. laadun, tavan tms. ilmaukseen liittyen. *Lähes samanlainen, mahdoton. Ei läheskään niin huono kuin luultiin. Lähes lopullisesti. Lähes keskellä. Lähes romuksi ajettu auto.*

Sana on taipumaton tai vaillinaisesti taipuva.

LIITE 2. Lekseemin *melkein* vierussanat (O1) supisuomessa ja ICLFI:ssä. Vierussanat olen käsitellyt lemmoina, mutta silloin kun niillä on ollut edustuma kollokaattiehdoikkaana vain yhdessä sananmuodossa, olen kirjoittanut sen pienin grafeemein.

	O1-kollokaattiehdoikkaat			
	SUPISUOMI	F	ICLFI	F
<i>melkein</i>	KAIKKI	69	KAIKKI	73
<i>melkein</i>	KUIN (kuin, ku)	44	koko	49
<i>melkein</i>	aina	29	joka	40
<i>melkein</i>	yhtä	25	aina	34
<i>melkein</i>	joka	23	SAMA	15
<i>melkein</i>	koko	23	JOKAINEN	14
<i>melkein</i>	SAMA	19	KAKSI	11
<i>melkein</i>	MIKÄ	19	kuin	9
<i>melkein</i>	kokonaan	18	mitään	9
<i>melkein</i>	heti	15	ei	7
<i>melkein</i>	MAHDOTON	12	MAHDOTON	7
<i>melkein</i>	HUUTAA	9	mitä	6
<i>melkein</i>	JOKAINEN	9	SAMANLAINEN	6
<i>melkein</i>	kiinni	9	kolme	5
<i>melkein</i>	päivittäin	9	niin	5
<i>melkein</i>	UNOHTAA	9	KAIKKIALLA	4
<i>melkein</i>	SE	8	kello	4
<i>melkein</i>	KUULLA	8	koskaan	4
<i>melkein</i>	MUSTA	7	LOPPU	4
<i>melkein</i>	KOLMEKYMMENTÄ	7	PUOLI	4
<i>melkein</i>	LIKAA	7	vain	4

<i>melkein</i>	TYHJÄ	7	VALMIS	4
<i>melkein</i>	KAKSI	6	viisi	4
<i>melkein</i>	TÖRMÄTÄ	6	yhtä	4
<i>melkein</i>	AIKUINEN	6	heti	3
<i>melkein</i>	VARMA	6	20	3
<i>melkein</i>	itketti	5	keskustassa	3
<i>melkein</i>	KYMMENEN	5	kokonaan	3
<i>melkein</i>	PIMEÄ	5	KOLMEKYMMENTÄ	3
<i>melkein</i>	PUOLI	5	SATA	3
			TUNTI	3
			UNOHTAA	3

LIITE 3. Taulukko lekseimin *melkein* kollokaateista asemassa O1 WordSmith Tools -ohjelman z-testauksen mukaan. Kollokaatit ovat sananmuotoina.

	O1-kollokaatit WST-ohjelmassa (z-testi) sananmuotoina			
	SUPISUOMI	F	ICLFI	F
<i>melkein</i>	kaikki	53	koko	49
<i>melkein</i>	kuin	43	joka	40
<i>melkein</i>	aina	29	kaikki	37
<i>melkein</i>	yhtä	25	aina	34
<i>melkein</i>	joka	23	kaksi	9
<i>melkein</i>	koko	23	mitään	9
<i>melkein</i>	kokonaan	18	kuin	8
<i>melkein</i>	heti	15	ei	7
<i>melkein</i>	mitä	10	kaiken	7
<i>melkein</i>	kiinni	9	sama	6
<i>melkein</i>	päivittäin	9	jokainen	5
<i>melkein</i>	huusi	8	jokaisessa	5
<i>melkein</i>	mahdotonta	7	kolme	5
<i>melkein</i>	saman	6	mitä	5
<i>melkein</i>	itketti	5	niin	5
<i>melkein</i>	kaksi	5	kello	4
<i>melkein</i>	kuollut	5	koskaan	4
<i>melkein</i>	kuulin	4	vain	4
<i>melkein</i>	liian	4	puoli	3
<i>melkein</i>	samaan	4	kaikilla	2
<i>melkein</i>	valmis	4	lopussa	2
<i>melkein</i>	varma	4	että	1
<i>melkein</i>	ilmaiseksi	3	hyvin	1
<i>melkein</i>	mikä	3	kun	1
<i>melkein</i>	samassa	3	minun	1
<i>melkein</i>	uusi	3	ole	1
<i>melkein</i>	heidän	2	olisi	1
<i>melkein</i>	juoksi	2	paljon	1
<i>melkein</i>	kolme	2	suomalaiset	1
<i>melkein</i>	neljä	2	vanha	1
<i>melkein</i>	vaan	2		
<i>melkein</i>	vieressä	2		
<i>melkein</i>	mitään	1		

LIITE 4. Lekseemin *lähes* vierussanat (O1) supisuomessa ja ICLFIssä. Vierussanat olen käsitellyt lemmoina, mutta silloin kun niillä on ollut edustuma kollokaattiehtokkaana vain yhdessä sananmuodossa, olen kirjoittanut sen pienin grafeemein.

	O1-kollokaattiehtokkaat			
	SUPISUOMI	F	ICLFI	F
<i>lähes</i>	KAIKKI	109	KAIKKI	9
<i>lähes</i>	kokonaan	30	MAHDOTON	2
<i>lähes</i>	KAKSI	30	heti	1
<i>lähes</i>	yhtä	26	ikinä	1
<i>lähes</i>	PUOLI	24	jokaisessa	1
<i>lähes</i>	aina	23	koko	1
<i>lähes</i>	koko	22	kuin	1
<i>lähes</i>	poikkeuksetta	21	loppuun	1
<i>lähes</i>	SAMA	21	rupean	1
<i>lähes</i>	joka	19	suosituimmaksi	1
<i>lähes</i>	SATA	16	unohdettu	1
<i>lähes</i>	TÄYDELLINEN	16	1/3	1
<i>lähes</i>	yksinomaan	16	30 000	1
<i>lähes</i>	JOKAINEN	15	500	1
<i>lähes</i>	AINOA	14	viisi	1
<i>lähes</i>	MAHDOTON	14		
<i>lähes</i>	KYMMENEN	14		
<i>lähes</i>	OLEMATON	13		
<i>lähes</i>	täysin	13		
<i>lähes</i>	KAIKKIALLA	12		
<i>lähes</i>	päivittäin	11		
<i>lähes</i>	KOLMEKYMMENTÄ	10		
<i>lähes</i>	KAKSIKYMMENTÄ	9		
<i>lähes</i>	TYHJÄ	9		
<i>lähes</i>	TUHAT	7		
<i>lähes</i>	KOLMINKERTAINEN	6		
<i>lähes</i>	pelkästään	6		
<i>lähes</i>	YHDEKSÄNKYMMENTÄ	6		
<i>lähes</i>	NELJÄKYMMENTÄ	5		
<i>lähes</i>	TÄYSI	5		
<i>lähes</i>	VIISIKYMPPIINEN	5		

LIITE 5. Taulukko lekseemin *lähes* kollokaateista asemassa O1 WordSmith Tools -ohjelman z-testauksen mukaan. Kollokaatit ovat sananmuotoina.

	O1-kollokaatit WST-ohjelmassa (z-testi) sananmuotoina			
	SUPISUOMI	F	ICLFI	F
<i>lähes</i>	kaikki	55	kaikki	6
<i>lähes</i>	kokonaan	30		
<i>lähes</i>	yhtä	26		
<i>lähes</i>	aina	23		
<i>lähes</i>	koko	22		
<i>lähes</i>	poikkeuksetta	21		
<i>lähes</i>	joka	19		
<i>lähes</i>	kaksi	18		
<i>lähes</i>	yksinomaan	15		
<i>lähes</i>	täysin	13		
<i>lähes</i>	kaikilla	12		

<i>lähes</i>	kymmenen	11		
<i>lähes</i>	puolet	11		
<i>lähes</i>	päivittäin	11		
<i>lähes</i>	jokainen	10		
<i>lähes</i>	100	9		
<i>lähes</i>	mahdotonta	9		
<i>lähes</i>	kaikissa	8		
<i>lähes</i>	täydellinen	8		
<i>lähes</i>	20	7		
<i>lähes</i>	kaikkialla	7		
<i>lähes</i>	kaikkien	7		
<i>lähes</i>	kolme	7		
<i>lähes</i>	sama	7		
<i>lähes</i>	kahden	6		
<i>lähes</i>	kaiken	6		
<i>lähes</i>	pelkästään	6		
<i>lähes</i>	ainoa	5		
<i>lähes</i>	ainoana	5		
<i>lähes</i>	10	5		
<i>lähes</i>	30	5		
<i>lähes</i>	60	5		
<i>lähes</i>	90	5		
<i>lähes</i>	puolella	5		
<i>lähes</i>	sata	5		
<i>lähes</i>	40	4		
<i>lähes</i>	miljoona	4		
<i>lähes</i>	tyhjä	4		
<i>lähes</i>	2	3		
<i>lähes</i>	vuosittain	3		
<i>lähes</i>	1	2		
<i>lähes</i>	500	2		