

»MINULLA ON AIVAN PALJON RAHAA» — FRASEOLOGISET YKSIKÖT SUOMEN KIELEN OPETUKSESSA

Kun vieraan kielen opiskelija tuottaa opittavaa kieltä, häneltä vaaditaan melkoista kykyä arvata oikein: vaikka kielioppi alkaisikin olla hallussa, erilaisiin opituihin rakenteisiin sijoitettavan sanaston käyttöön ei juurikaan ole tarjolla ohjeita. Kielenoppija joutuu aivan kuin kokeilemaan, käykö hänen tuntemansa tai sanakirjasta löytämänsä sana yhtäältä rakenteeseen ja toisaalta toisten sanojen yhteyteen. Erityisesti tämä koskee suomea vieraana kielenä opiskelevia, joilla ei ole tukena ympäröivän kieliyhteisön tarjoamaa mallia. Usein oppijan apuvälineenä toimivat sanakirjatkaan eivät tuo apua tähän ongelmaan, sillä lukuisat sanastot ja sanakirjat, mukaan lukien oppikirjojen sanastot, esittävät kielen sanat kontekstistaan irrallisina yksikköinä (poikkeuksena ovat englannin oppijoille tarkoitetut kollokaatiosanakirjat, kuten *Oxford Collocations Dictionary for Students of English*, 2002). Tämä on kuitenkin vastoin kielen olemusta: sanat eivät ole kontekstistaan irrallisia, vaan niillä on usein hyvinkin tarkkoja »käyttösääntöjä» tai ainakin vahvoja preferenssejä esiintyä tietyn kaltaisessa ympäristössä. Natiivit kielenkäyttäjät hallitsevat nämä usein hyvinkin hienovaraiset säännöt opittuaan ne kielenomaksumisen aikana, mutta vieraan kielen oppija joutuu opiskelemaan säännöt sana sanalta osana kielenoppimisprosessia.¹

¹ Artikkelin perustuu Suomi toisena kielenä ja monikielisyys -konferenssissa Jyväskylässä 1.–2.2.2008 pitämäni esitelmään. Kiitän Virittäjän arvioijia hyödyllisistä kommentteista.

Natiivi kielenkäyttäjät tietää, että pilalle mennessään *voi yleensä härskiintyy* mutta harvoin *hapantuu*. Samoin *oikein mukava ilta* on hänen kielenkäytössään tavallisempi ja odotuksenmukaisempi kuin *hyvin mukava ilta*. Tällaiset melko arbitraariset leksikaaliset yhteisiintymät ovat hankalia oppimisen näkökulmasta, sillä ne eivät perustu kielenulkoiseen todellisuuteen, kuten vaikkapa *kana kotkottaa* (ei *hirnu*) ja *kissa maukuu* (ei *hauku*), jotka ovat äänneellisesti motivoituneita ja joiden muistaminen voi siksi olla helpompaa. Näissä tapauksissa on kyse sanojen kollokoinnista eli leksikaalisesta myötäesiintymisestä toisten sanojen kanssa, mikä on useasti luettu yhdeksi sanan osaamisen kriteeriksi (ks. Puro 1999: 5–9). Kollokaatioiden huomioon ottaminen sekä opetuksessa että oppimisen, opetuksen ja oppijankielen tutkimuksessa on jo tavallista englannin kielessä, muissa kielissä myötäesiintymisen tarkastelu on vielä melko vähäistä. Tällä hetkellä oppijankielen tutkimuksessa ja kielenopetuksessa vallitsee kuitenkin jonkinlainen fraseologinen buumi: sähköisten konelukuisten aineistojen eli korpusien myötä on alettu tarkastella leksikkoo entistä enemmän yhtä sanaa laajempina kokonaisuuksina, samoin kuin opetukseen on otettu korpuksat ja fraseologia avuksi. Tämäkin pätee kuitenkin toistaiseksi pääasiassa jälleen englantiin, tosin korpusavusteinen ja laajojen aineistojen tutkimusta hyödyntävä opetus on hiljalleen leviämässä myös muihin kieliin.

Tarkastelen tässä artikkelissani leksikaalista yksikköä fraseologisesta näkökulmasta. Fraseologisuudella en kuitenkaan tarkoita pelkästään erilaisten fraasien, sanontojen, idiomien ja keskustelurutiinien (kuten *Hyvää päivää*, *Kiitos paljon*, *Mitäs tässä*) muodostamaa ilmiötä, vaan lasken fraseologiaan kuuluvaksi Sylviane Grangerin (mm. 1998, 2005) tavoin myös vakiintuneet monisanaiset leksikaaliset yksiköt. Pyrin analysoimaan sanastoa vahvasti kontekstisidonnaisina leksikaalisina yksikköinä, ja laajennan tarkastelua kollokationaalisesta suhteesta abstraktimpiin kontekstuaalisiin suhteisiin. Käytän esityksessäni apuna *aivan*-adverbia ja synonyymiparia HIEVAHTAA–LIIKAHTAA²; näistä *aivan* on kielessä frekventti ja kontekstuaalisilta suhteiltaan varioiva, mikä luo analyysille ja käytönsääntöön kuvaukselle haasteita, HIEVAHTAA ja LIIKAHTAA ovat puolestaan harvinaisia ja paljastavat yllättävänkin selvärajaisia eroja synonyymisten sanojen käytöstä.³ Vaikka synonyymipari saattaa harvinaisuutensa vuoksi olla etenkin alkeis- ja keskitason kielenopetuksen kannalta marginaalinen, on sen avulla hyvin kuvattavissa opetuksen näkökulmasta vaikeita synonyymisyyteen liittyviä piirteitä — erityisesti sitä, miten synonyymit monella tapaa eroavat toisistaan. Selvitettyään englannin kielen opettajien hankalasti opetettavaksi kokeimia kielenpiirteitä Tsui (2004, 2005) raportoi erityisesti synonyymisten sanojen erojen ja kollokationaalisten suhteiden selittämisen olevan huomattavan haastavaa.

Luon tässä artikkelissani katsauksen siihen, millaista tietoa korpus tarjoaa opetuksen tueksi. Kun fraseologiaa tarkastellaan korpuslingvivistisistä lähtökohdista, yksiköiden frekvenssit ja tilastollinen merkitsevyys ovat vahvasti mukana määritelmiä luotaessa

² Kapiteeli-merkintä tarkoittaa lemmaa, eli se sisältää kaikki taivutusmuodot. Kursiivi puolestaan merkitsee kyseistä muotoa.

³ Tieteen tietotekniikan keskuksen (CSC) 90-luvun sanomalehtikielen taajuussanastossa *aivan* on sijalla 275, *liikahtaa* ja *hievahtaa* eivät mahdu 10 000 yleisimmän sanan joukkoon; sanasto on osoitteessa www.csc.fi/tutkimus/alat/kielitiede/taajuussanasto-B9996/view. Suomen kielen taajuussanastossa (Saukonen ym. 1979) *aivan* on sijalla 124, *liikahtaa* 7844, mutta *hievahtaa* ei ole mahtunut vajaan 13 000 yleisimmän sanan joukkoon.



(ks. Ellis 2008: 2). Myös tässä esityksessä havainnot perustuvat esiintymistajuuksiin ja tilastollisuuteen, ja kvalitatiivisilla havainnoilla täydennetään kvantitatiivisia tuloksia. Ei kuitenkaan tarvitse lähteä siitä oletuksesta, että esimerkiksi kielenopettajan olisi hallittava tilastollisia menetelmiä: aineistoa voi tarkastella ilman tilastollisiakin menetelmiä, jolloin voi tehdä jo vahvasti suuntaa antavia havaintoja. Menetelmiä voivat kuitenkin hyödyntää aineistoa analysoivat tutkijat, joiden tuottamia havaintoja voidaan käyttää hyväksi ope-
tustyössä muun muassa erilaisten sanastojen, sanakirjojen ja oppikirjojen muodossa.

FRASEOLOGINEN YKSIKKÖ – YHTÄ SANAA LAAJEMPI ELEMENTTI

Sanastoa on totuttu tarkastelemaan etupäässä paradigmaattisina suhteina, kuten taivutusparadigmoina, synonymiana ja antonymiana. Syntagmaattisten suhteiden tarkastelu on harvinaisempaa, joskin yhdyssanojen ja syntagmaattisten sulaumien (*miksette, tajuutsä*, VISK §125) analysointi korostaa sanaston syntagmaattista ulottuvuutta. Syntagmaattisesta ulottuvuudesta kertovat myös niin sanotut monisanaiset yksiköt (*multi-word units*), jotka ovat suhteellisen jähmeitä muuttumaan mutta joissa saatetaan tavata kuitenkin jonkin verran morfologista tai leksikaalista variaatiota. Tällaisia ovat mm. liittokonjunktiot ja niiden kaltaiset usean sanan yhdistelmät (*ennen kuin, sen jälkeen kun, sillä aikaa/välin kun*) ja idiomit (*aiheuttaa/tuottaa harmaita hiuksia, saada turpiin/turpaan*, Muikku-Werner ym. 2008 s.v. *harmaa, saada*). Sanoilla on kuitenkin myös monia muita syntagmaattisia suhteita, jotka realisoituvat joko tiukkoina kotekstiin eli läheiseen tekstiympäristöön suuntautuvina valintarajoituksina tai väljempinä valintapreferensseinä. Näiden suhteiden taustalla on lekseemien idiomaattisen valikoitumisen perussääntö: lekseemit eivät esiinny tavallisessa kielenkäytössä satunnaisesti eivätkä mielivaltaisessa järjestyksessä ja suhteessa (= vapaan valinnan perussääntö), vaan niiden käyttöä ohjaavat (kieliopillisten rajoitusten lisäksi) lukuisat idiomaattiset esiintymisrajoitukset ja -preferenssit. Natiivilla kielenkäyttäjällä on siis hallussaan runsaasti valmiita rakenteita. (Ks. käsitteistä *idiom principle* ja *open choice principle* mm. Sinclair 1991: 109–115.) Tässä yhteydessä on tärkeä huomata, että idiomaattista valintaa ei pidä ymmärtää yksinomaan idiomien käyttöön liittyvänä periaatteena, vaan se koskee myös yksittäisen sanan kotekstuaalisia suhteita, jotka natiivi kielenkäyttäjällä hallitsee luonnostaan. Näitä syntagmaattisia suhteita ovat seuraavat (ks. tarkemmin Sinclair 1996; Stubbs 2001; Jantunen 2004):

kollokaatio

Kahden sanan muodostama kiinteä suhde eli **leksikaalinen** myötäesiintymä. Yleensä kollokaatio määritellään toistuvaksi, tilastollisesti merkitseväksi ja ulottuvuudeltaan rajatuksi syntagmaattiseksi suhteeksi. Kollokaatio voi motivoitua selvästi kielenulkoisesta maailmasta (KANA <KOTKOTTA>, KUKKO <KIEKUA>), mutta varsin usein suhteelle ei ole esitettävissä tämänkaltaista syytä. Esimerkiksi KOVIN kollokoi kieltosanan EI kanssa, mutta sille synonyymiset OIKEIN ja HYVIN eivät saa kieltosanaa kollokaatukseen (tarkemmin Jantunen 2004: 86–92).

klusteri ~ n-grammi

Useamman kuin kahden peräkkäisen sanan muodostama toistuva **sanaketju**. Klusterin ei tarvitse olla kieliopillisesti motivoitu, eikä sitä välttämättä voida korvata yhdellä sanalla kuten idiomeja. Tällaisia voivat olla esimerkiksi 3-grammit (EI OLLA KOVIN) tai 4-grammit (PER SENT OF THE; Biber ja Conrad 1999: 186–187).

kolligaatio

Sanan ja kieliopillisen kategorian (tai lauseaseman) välinen syntagmaattinen suhde eli **leksikaalis-kieliopillinen** myötäesiintymä. Esimerkiksi KOVIN esiintyy synonyymisiä astemääritteitä HYVIN ja OIKEIN useammin kvanttorien seurassa: KOVIN <kvanttori: HARVA, MONI, PALJON, VÄHÄN>. Niin sanottua (syntagmaattista) morfologista preferenssiä voidaan pitää osana kolligatiionaalista suhdetta; sillä tarkoitetaan lekseemin ja kontekstissa esiintyvän morfologisen ominaisuuden suhdetta (Jantunen 2004: 27–29). KOVIN-määritteestä esimerkiksi käyvät mm. seuraavat verbin taivutusmuotoja koskevat morfologiset preferenssit: KOVIN <akt. yks. 3> ja <akt. ind. imperf. yks. kielto>, joista ensimmäinen on usein *ei* ja jälkimmäinen esimerkiksi *ei juossut* (*Mies ei juossut kovin kovaa*). Kolligaatioon kuuluu myös suomen kielen opetuksesta tuiki tuttu ja oppijalle vaikea rektio-ilmiö eli sanan täydennysten sijavalinta (vrt. *Tykkään sinusta* vs. *Rakastan sinua*).⁴

semanttinen preferenssi

Lekseemin ja kontekstissa esiintyvän merkityspiirteen syntagmaattinen suhde eli **leksikaalis-semanttinen** myötäesiintymä. Semantiselle preferenssille on tyypillistä, että merkityspiirre voi ilmetä kontekstissa vaihtelevissa positioissa ja eri sanaluokan sanoissa. OIKEIN-astemääritteellä on mm. selvä semanttinen preferenssi 'arvoa, hyvyttä' merkitseviin sanoihin: OIKEIN <'hyvyys': HYVÄ, HYVIN, MAINIO>. Sinclairin (1996) antamassa *naked eye* -esimerkissä puolestaan kyseisen ilmauksen semanttiseksi preferenssiksi paljastuu 'näkyminen, näkyvyys', joka ilmenee verbeissä (SEE ja VIEW) ja adjektiiveissa (VISIBLE ja INVISIBLE) eri kohtaa kontekstia. Vaikka positio ja sanaluokkavaihtelu kuuluu tähän kontekstuaaliseen piirteeseen, merkityspiirteen esiintymisen on oltava säännönmukaista lekseemin kontekstissa.

semanttinen prosodia

Semanttista preferenssiä abstraktimpi **leksikaalis-semanttinen** myötäesiintymä, niin sanottu merkitysaura (Louw 1993: 157), joka lekseemiä ympäröi. Semanttinen prosodia määritellään positiiviseksi tai negatiiviseksi (tai neutraaliksi) esiintymisympäristöksi tai Sinclairin (2004a: 140–146) tapaan ilmauksen pragmaattiseksi merkitykseksi, diskurssifunktioksi, joka on perimmäinen syy, miksi ilmaus on ylipäättään valittu käytettäväksi. Esimerkiksi Stubbsin (1995: 42–48) mukaan englannin CAUSE-verbillä ('aiheuttaa') on negatiivinen käyttöympäristö eli semanttinen

⁴ Tämän artikkelin analyyseissa rektio ohitetaan, koska tarkasteltavana olevilla leksikaalisilla yksiköillä ei ole tietynsijaisia täydennyksiä. Rektion ja kolligaation suhdetta olisi kuitenkin tärkeä analysoida syvemmin, erityisesti S2-opetuksen näkökulmasta.

prosodia, joka ilmenee mm. seuraavina merkitykseltään negatiivisina kollokaatteina CAUSE <DEATH, PROBLEM, TROUBLE>.⁵

Näiden syntagmaattisten suhteiden lisäksi lekseemeillä on myös ns. paradigmaattisia preferenssejä eli taipumuksia tulla käytetyksi tietyissä taivutusmuodoissa. Tällainen paradigmaattinen morfologinen preferenssi (Jantunen 2004: 29) ilmenee muun muassa adjektiiveilla TÄRKEÄ ja KESKEINEN sekä verbeillä VOIDA ja PYSTYÄ seuraavasti: TÄRKEÄ-adjektiivia käytetään tyypillisesti partitiivimuotoisena, KESKEINEN on usein nominatiivimuotoinen (Jantunen 2001); PYSTYÄ puolestaan esiintyy usein imperfekti- ja partisiippimuodoissa, VOIDA taas on taajaan passiivi- ja konditionaalimuodoissa (Arppe 2001; ks. myös Arppe 2002). Kaiken kaikkiaan lekseemeillä on siis runsaasti erilaisia leksikaalisia, kieliopillisiä ja semanttisia valintoja suhteessa kotekstiinsa ja omiin taivutusmuotoihinsa. Näiden lisäksi tulevat luonnollisesti vielä lukuisat tyyli- ja rekisteri- ja tekstilajipiirteet, jotka myös ohjaavat sanojen käyttöä. Lekseemeihin on koodattuna siis hyvin runsaasti ko- ja kontekstuaalisia »käyttöohjeita», joista etenkin kontekstuaaliset ominaisuudet ovat hyvin abstrakteja. Tognini Bonellin (2008) sanoin jokaisella sanalla on oma lokaalinen kielioptionsa, ja samoin niillä on omat lokaaliset semanttis-pragmaattiset suhteensa.

FRASEOLOGINEN BUUMI OPIJANKIELEN JA KIELENOPETUKSEN TUTKIMUKSESSA

Suurten yleiskorpusten ja 90-luvun alusta kerättyjen oppijankielen korpusten myötä oppijankielen — ja myös yleensä kielen — tutkimuksessa on meneillään eräänlainen fraseologinen buumi (ks. Granger ja Meunier 2008a: XX). Näin on asianlaita erityisesti niin sanotun brittiläisen, firthiläis-sinclairiläisen koulukunnan (tutkijat kuten J.R. Firth, John Sinclair, Elena Tognini Bonelli, Michael Stubbs) korpuspohjaisessa tutkimuksessa; vastaava trendi on nähtävissä voimakkaasti myös Grangerin alulle panemassa korpuksiin pohjautuvassa oppijankielen tutkimuksessa. Yhtenä syynä fraseologisen tutkimuksen lisääntymiseen voidaan luonnollisesti pitää suurten aineistojen saatavuutta mutta myös näkökulman muutosta fraseologian tutkimuksessa: kun aiemmassa tutkimuksessa ilmaukset on pyritty luokittelemaan ennalta formuloituihin luokkiin, rinnalle on tullut aineistosta nouseva tutkimus, jossa yksittäisiä ilmauksia ryhmittämällä luodaan uusia luokituksia (Granger 2005: 166). Fraseologinen näkökulma korostaa nimenomaan kielen **käytön** tutkimusta ja kielen elementti- tai valmisrakenteisuutta (*prefabricated patterns, prefabs*) — lähtökohdana on ajatus siitä, että hyvin suuri osa kielenkäytöstä muodostuu nimenomaan fraseologisesta kielenkäytöstä. Elementtirakenteiksi Granger (mm. 1998) lukee kollokaatiot ja rutiiniformulat, kuten tervehdykset ja anteeksipyyntöt, mutta myös eräänlaiset vakiintuneet diskurssimuotit (esim. *it + modaaliverbi + passiivi + that*-lause: *it can be claimed/assumed that* –). Kielessä ilmenevää fraseologiaa ei pitäisikään tarkastella vain keskittyen pelkästään idiomeihin (*mennä kuuroille korville*), rutiineihin (*Hyvää päivää!*, *Ei kestä*) tai muihin perinteisesti fraaseiksi luettuihin, suhteellisen varioimattomiin

⁵ Käsitteen problemaattisuudesta ks. Whitsitt (2005) ja Hunston (2007).

ilmauksiin⁶, vaan fraseologiseen tarkasteluun on otettava mukaan Grangerin tapaan myös muut vakiintuneet rakenteet. Grangerin ja muidenkin (esim. Nesselhauf 2003, 2005) analyysit jättävät kuitenkin lähes huomiotta tässä artikkelissa käsiteltävät kollokointia tai klustereita abstraktimmat paradigmaattis-syntagmaattiset suhteet.

Oppijankielen tutkimuksessa etenkin kollokaatiot ovat olleet merkittävä analyysin kohde kontekstuaalisten piirteiden tarkastelussa (ks. korpustutkimuksista esim. Granger 1998 ja Nesselhauf 2003, 2005; oppijansuomesta manuaalisin aineistoin Grönholm 2000, 2001). Laajoihin aineistoihin pohjautuen on pystytty entistä vakuuttavammin todistamaan, että sanaston oppimisessa kollokaatiot ovat arbitraarisen suhteensa vuoksi yksi vaikeimmin opittavissa oleva asia, samaan tapaan kuin esimerkiksi suomen kielen rektiosäännöt, joiden oppimisessa ei pysty käyttämään tukena tietoa kielen järjestelmällisestä rakenteesta vaan oppiminen on pitkälti muistamisen varassa (ks. Jönsson-Korhola ja White 1997: 5). Kuten idiomit, kollokationaaliset suhteetkin on yleensä opeteltava ulkoa. Siksi edistynytkin opiskelija tai kielenkäyttäjät luo epätyypillisiä kollokaatioita tuottaessaan kohdekielisiä ilmauksia, mikä kertoo ilmiön haasteellisuudesta oppimisen kannalta. Onkin todennäköistä, että epätyypilliset kontekstuaaliset suhteet — laajemmin kontekstuaalinen epätyypillisuus — on yksi oppijankielen yleisistä, universaaleista piirteistä (Jantunen 2008)⁷. Oppijan tuottama kieli saattaakin kuulostaa epämääräisesti »väärältä» huolimatta ilmauksen kieliopillisuudesta. Nesselhaufin (2005: 70) tutkimuksesta käy ilmi, että tutkimuksessa mukana olleista 207 edistyneestä äidinkieleltään saksalaisesta englanninoppijasta lähes 90 % tuottaa kohdekielelle epätyypillisiä kollokationaalisia rakenteita; erityisen haastavia ovat sellaiset kollokaatiot, joissa kollokaation muodostavissa sanoissa on jonkin verran valinnan vaihtoehtoja, helpommin osataan sellaiset kollokaatiot, jotka ovat suhteellisen jähmeitä varioimaan leksikaalisesti, sillä ne omaksutaan idiomien tapaan kokonaisina leksikaalisina yksikköinä (Nesselhauf 2003: 233; 2005: 203–204). Oppijan äidinkielellä on suuri vaikutus kollokaatioiden muodostamiseen (Nesselhauf 2003: 235–236).

Opetuksessa ja oppijankielen tutkimuksessa fraseologinen lähestymistapa on kuitenkin vinoutunut vahvasti pelkästään kollokationaalisten suhteiden tarkasteluun. Huolimatta siitä, että laajojen korpusten avulla on tehty jo melkoisen paljon tutkimusta myös kollokaatiota käsitteellisemmistä kontekstuaalisista suhteista, ei tällainen tutkimus ole yltänyt oppijankielen tutkimukseen ja opetuksen hyödyksi. Aivan viimeaikaisessakin tutkimuksessa oppijankielen fraseologiasta ja sen hyödyistä kielenopetuksessa (esim. Nesselhauf 2005; Kennedy 2008) pitäydytään pelkästään kollokaatioissa; samoin on laita, kun tarkastelussa ovat opiskelun tueksi tarkoitetut sanakirjat ja niiden kehittäminen (esim. Moon 2008; Siepmann 2008). Seuraavaksi pyrin laajentamaan näkökulmaa kollokaatioista klustereihin ja näitä abstraktimpiin paradigmaattis-syntagmaattisiin suhteisiin.

⁶ Ks. fraasien luokittelusta mm. Kuusi (1971) ja Nenonen (2002). Fraaseihin liitetään tyyppisesti syntaktinen, morfologinen ja leksikaalinen jähmeys, mutta esimerkiksi idiomeissa vaihtelua voi esiintyä yllättävänsä paljon (ks. Nenonen 2002; Muikku-Werner ym. 2008).

⁷ Oppijankielen universaaleiksi piirteiksi on esitetty epäkonventionaalisuutta (rekistereiden sekoittuminen ja kontekstuaalinen epäkonventionaalisuus), kieltenvälistä vaikutusta, kielenainesten epätyypillisiä frekvenssejä ja yksinkertaisuutta (Jantunen 2008).



TAPAUS 1: »MINULLA ON AIVAN PALJON RAHAA»

Otsikon virke on aito esimerkki suomen kielen oppijan tuottamasta tekstistä.⁸ Se on ymmärrettävä, mutta epätyypillinen *aivan paljon* -määritys-suhteen vuoksi. Virke on kieliopilliselta rakenteeltaan hyväksyttävä (vrt. *Minulla on hyvin paljon rahaa*), mutta leksikaalis-grammaattiselta rakenteeltaan ei: ilman korpusanalyysiakin natiivi kielenpuhujat pystyvät sanomaan, että *aivan* ei tyypillisesti määritä *paljon*-kvanttorin. Oppitunnilla keskustelunaiheeksi saattaisikin luultavasti nousta, miten *aivan*-sanaa käytetään tai mitä sanoja käytetään *paljon*-kvanttorin määrittäen. Etenkin ensimmäisessä kysymyksessä opettajan intuitio voi osoittautua riittämättömäksi, joskin intuitionkin perusteella selitysmalleiksi nousisivat varmankin käyttö dialogipartikkelina ja jonkinlaisena astemäärittäen. Luotettavaa tietoa ilmausten käytöstä, kuten frekvenssistä ja variaatiosta, on vaikea saavuttaa pelkästään intuition avulla (Sinclair 2004c: 2; Mauranen 2004: 96). Pikemminkin intuitio ja introspektio kertovat, mitä kielenkäyttäjät tietävät kielestä tai millaiseksi hän ymmärtää sen (Tsui 2004: 39; ks. myös Jantunen 2009). Tarkastelen seuraavaksi, millaisia selitysmalleja opettajalle on tarjolla.

Luonnollinen vaihtoehto etsiä tukea intuitiolle on tarkistaa ilmausten käyttö sanakirjoista ja kieliopeista. Kielitoimiston sanakirja (MOT KTS) määrittelee astemäärittäen synonyymien avulla, kuten tyypillistä on. *Aivan* tarkoittaa sanakirjan mukaan siis (liki-pitäen) samaa kuin *ihan*, *vallan*, *erittäin*, *hyvin*, *kovin*, *perin*; *juuri*, *tarkalleen*, *täysin*, *täydellisesti*, *kokonaan*, *tykkänään*; *suorastaan*, *kerrassaan* (s.v. *aivan*). Esimerkistö tarjoaa jonkin verran myös käyttömalleja: *Aivan pieni* (ei: *aivan suuri*), *aivan kevyt* (ei: *aivan raskas*). Lisäksi esimerkeissä luetellaan joukko muita määritettäviä adjektiiveja ja adverbeja ja mainitaan *aivan* (*niin*) *kuin* -rakenne sekä pari tapausta, jossa *aivan* esiintyy verbin määrittäenä (*Olin aivan läikähtyä*). Erikseen mainitaan vielä käyttö vahvistavana vastauksena: *Hänhän on sinun veljesi. Aivan* (*niin*). Tällainen dialogipartikkelimainen käyttö on mainittu myös Isossa suomen kieliopissa (VISK § 863). Kielioppi määrittelee *aivan*-adverbin hyvin monikäyttöiseksi (VISK § 665) täyteyden asteen määritteeksi (VISK § 615, § 660) ja täsmentimeksi (VISK § 663). Kieliopista ei saa tietoa tyypillisimmistä syntagmaattisista suhteista, tosin sen käyttöä valaistaan useassa luvussa lukuisin esimerkein. Taajakäyttöisyytensä vuoksi *aivan*-lekseemistä on siis saatavana melko paljon tietoa, jota opettaja voisi käyttää hyväkseen; toisaalta tyypilliset käyttöyhteydet eivät varsinaisesti selviä oppaista ja oppaiden tieto voi olla myös osin ristiriitaista.

Tarkastelen seuraavaksi, millaista tyypillisyyteen liittyvää tietoa kyseisen sanan käytöstä on saatavissa laajan sähköisen tekstiaineiston avulla. Käytän analyysissäni hyväksi Käännösuomen korpuksista (Mauranen 2000) koostettua Natiivisuomen korpusta, joka sisältää sekä tietokirjallisuutta (akateemiset ja populaaritieteelliset tekstit sekä muistelmat) että kaunokirjallisuutta (romaanit, mukaan lukien viihde- ja salapoliisiromaanit sekä lastenkirjat) 3,8 miljoonan saneen verran. Kuvausta olen täydentänyt paikoin Suomen kielen tekstipankin sanomalehtiaineistosta saaduilla frekvenssitiedoilla. Analyysityökaluna olen käyttänyt WordSmith Tools -ohjelmapakettia (Scott 2006).

⁸ Virke on peräisin Cimon suomen kielen ja kulttuurin kesäkurssilta Savonlinnasta vuodelta 1999, missä työskentelin opettajana.

Korpusaineistosta sanojen käyttöä voi tarkastella niin sanotun konkordanssin avulla, joka saadaan ohjelmapaketin Concord-ohjelmaa käyttäen. Esimerkki ohjelman tuottamasta konkordanssista on liitteessä 1. *Aivan*-lekseemin frekvenssi on aineistossa 2708, eli se esiintyy keskimäärin 72 kertaa 100 000 saneen otoksessa (*aivan* näyttää aineiston valossa olevan yleisempi kaunokirjallisissa teksteissä ($f/100\ 000 = 85$) kuin tietoteksteissä ($f/100\ 000 = 59$). (*Tekstipankin* sanomalehtiteksteissä [*Kalevan* vuosikerrat 1998 ja 1999] vastaava suhdeluku on 46). Tarkastelen seuraavaksi korpukselta saatavaa tietoa *aivan*-lekseemin kollokaateista, kolligaaateista ja klusteroinnista. Taulukkoon 1 on lueteltu *aivan*-lekseemin vasemman- ja oikeanpuoleiset vieruskollokaatit; vähimmäisfrekvenssiksi kollokaatiolle on asetettu 10 esiintymää. Taulukossa on esitetty todellinen frekvenssi (f) ja tilastollinen merkitsevyys t -testin avulla (merkitsevyysraja-arvo $t=2.00$).

Taulukko 1. *Aivan*-lekseemin kollokaatit ($f \geq 10$) Natiivisuomen korpuksessa.

V1-kollokaatti	f	t	O1-kollokaatti	f	t
olla	628 (23,2 %)	19,35	kuin	189 (7,0 %)	12,66
jo	35	4,90	liian	75	8,55
tulla	35	4,82	uusi	71	8,02
jokin	31	4,49	varma	62	7,79
näyttää	23	4,27	kuten	64	7,76
nyt	27	4,11	yhtä	49	6,79
kuitenkin	25	4,11	sama	53	6,61
mennä	17	2,90	toinen	56	6,52
vielä	15	2,59	varmasti	38	6,04
saada	15	2,10	eri	36	5,71
myös	13	1,12	oikein	34	5,63
tehdä	10	0,54	lähellä/e	31	5,44
niin	13	0,44	tavallinen	31	5,43
sanoa	13	0,14	muu	42	5,40
voida	13	0,11	samoin	30	5,38
ei	37	-0,03	hyvin	32	5,16
mutta	15	-0,10	selvästi	28	5,15
minä	21	-0,87	erilainen	28	5,01
hän	17	-3,19	ilmeisesti	26	5,01
se	22	-4,35	hiljaa	26	5,00
ja	58	-6,14	toisenlainen	20	4,43
			erityisesti	21	4,42
			mahdoton	20	4,38
			oikea	24	4,35
			erityinen	20	4,32



V1-kollokaatti	f	t	O1-kollokaatti	f	t
			hiljainen	19	4,27
			pieni	23	4,18
			selvä	19	4,11
			samanlainen	16	3,88
			viime	15	3,67
			väärä	14	3,58
			yksin	13	3,42
			uudenlainen	11	3,29
			tyhjä	12	3,28
			kaikki	19	3,27
			välttämätön	11	3,26
			liikaa	11	3,24
			turha	11	3,20
			sattumalta	10	3,14
			yllättäen	10	3,12
			luonnollinen	10	3,08
			niin	26	2,86
			oma	17	2,85
			----- hän	17	-3,19
			se	13	-8,16

Jo silmämääräisestäikin yli kymmenen kertaa esiintyvien kollokaattien määrästä on nähtävissä, että *aivan* kollokoi voimakkaammin oikeanpuoleisen kontekstin kanssa. Kun asiaa tarkastellaan tilastollisten tunnuslukujen avulla, kollokoinnin painottuminen oikealle on entistä selvempää: vain kahta O1-kollokaattia lukuun ottamatta (*hän*, *se*) kaikki yli kymmenen kertaa esiintyvät kollokaatit ovat myös tilastollisesti merkitseviä ($t > 2.00$), kun taas vasemmalla puolella vain puolet ylittää merkitsevyysrajan. Merkitsevyys tarkoittaa *aivan*-adverbin kannalta siis sitä, että merkitseviksi osoittautuvat kollokaatit eivät ole sattumalta sen läheisyydessä ja että ne ovat sen läheisyydessä useammin kuin yleensä tasaisesti jakautuneena tekstimassassa.⁹ Kaikkiaan yli kymmenen kertaa tavattavat kollokaatit muodostavat vasemmalla 40,5 % kaikista kollokaateista, oikealla osuus on selvästi suurempi, 51,2 %¹⁰. Vasemmassa kontekstissa OLLA-verbi on muita huomionarvoisempi kollokaatti sekä frekvenssinsä että merkitsevyytensä vuoksi: se kattaa kaikista vasemmanpuoleisista vieruskollokaateista lähes neljäsosan. Rakenteen kuvauksen ja opetuksen kannalta tätä

⁹ Luonnollisestikaan sanat eivät esiinny teksteissä tasaisesti jakautuneena, vaan niiden sijoittumista ohjaavat monet suhteet, kuten tässäkin artikkelissa on todettu.

¹⁰ Ero on myös tilastollisesti merkitsevä, $z = 7,8807$, $p < 0.001$.

voisikin pitää tärkeänä havaintona. Oikeassa kotekstissa frekventein ja merkitsevin kollokaatti on *kuin* 7 prosenttia peitollaan; *aivan* ja sen vieruskollokaatti muodostavat tällöin liittokonjunktion *aivan kuin* ’ikään kuin, niin kuin, siten kuin, samalla tavalla’:

- (1) a. Ossi *aivan kuin* aavisteli jotain kohtalokasta olevan tulossa.
- b. Karhu murahteli *aivan kuin* olisi keskustellut.
- c. Tuo kuva on *aivan kuin* Uudesta-Seelannista.

Kuten on myös taajaan esiintyvä vieruskollokaatti, ja *aivan kuten* -rakenne on merkityksensä puolesta läheinen *aivan kuin* -liittokonjunktioille: sitä on käytetty merkityksissä ’samoin kuin, samalla tavalla kuin, niin kuin’ esimerkiksi seuraavissa tapauksissa:¹¹

- (2) a. Ennustaminen epäonnistui *aivan kuten* hyökkäämispäätöskin.
- b. *Aivan kuten* jo vuonna 1918 Suomea ryhdyttiin siis jälleen »demokratisoimaan».
- c. Mutta Avellan taipui, *aivan kuten* von Haartman oli arvannut.

Frekventtien ja tilastollisesti merkitsevien kollokaattien joukosta on löydettävissä runsaasti myös muita vertailusanoja, kuten adjektiivin luonteiset pronominit TOINEN, SAMA, MUU; (pro)adjektiivit ERI, ERILAINEN, SAMANLAINEN, TOISENLAINEN, UUDENLAINEN; liiallisuutta ilmaiseva astemäärite *liian* ja kvanttoriadverbi *liikaa*, proadverbi *samoin* sekä samaa ominaisuuden astetta ilmaisevat adverbit *yhtä* ja *niin*. Näiden kollokaatioiden lisäksi *aivan* myötäesiintyy muunkinlaisten adjektiivien (kuten UUSI, VARMA, MAHDOTON) ja adverbien (*varmasti*, *oikein*, *selvästi*) kanssa. Taulukko 1 paljastaa, että pelkästään yli kymmenen kertaa esiintyvien kollokaattien määrä on suurehko, joten *aivan*-adverbin kollokationaalista valintaa ei voi pitää kovin rajoittuneena. Tämä onkin tavallista frekventeille astemääritteinä käytetyille adverbeille.

Vasemman- ja oikeanpuoleisen vierusposition erona on myös se, että suuri osa vasemmalla olevista kollokaateista on kielessä esiintyvää frekventeimpää sanastoa; sanat eivät esiintymistiheydensä vuoksi läheskään aina ole tilastollisesti merkitseviä.¹² Monet oikeanpuoleiset kollokaatit motivoituvat puolestaan *aivan*-adverbin ominaisuuden astetta ilmaisevan merkityksen ja mitattavan adjektiivin merkityksen suhteesta.

Adverbin kotekstista voi saada lisätietoa tarkastelemalla kotekstin semanttisia ja kielipiillisiä ominaisuuksia — eli semanttisia preferenssejä ja kolligaatioita. Semanttisen preferenssin analyysissa kotekstin kollokaatteja ryhmitellään merkityspiirteidensä perusteella ryhmiksi, jolloin päästään tarkastelemaan sitä, esiintyykö ilmaus järjestelmällisesti merkitykseltään tietynlaisten sanojen yhteydessä, vaikka kollokationaalinen kirjo näyttäisikin hyvin varioivalta. Ennako-oletuksena voidaan kuitenkin pitää, että semanttisten luokkien määrä ei ole niin suppea kuin aiemmin mainitussa Sinclairin (1996: 83–89) *naked eye* (’paljas silmä’) -esimerkissä: ilmaus esiintyy lähes yksinomaaisesti kotekstissa, jossa ilmenee semanttinen preferenssi ’näkyvyys, näkyminen’. Stubbsin (2001: 64)

¹¹ *Aivan kuten* -rakenne on selvästi rajoittunut (populaari)tieteellisiin teksteihin, sillä 64 esiintymästä vain 4 on kaunokirjallisuudesta.

¹² Tästä OLLA on kuitenkin poikkeus, sillä yleensä näin frekventti sana ei ole tilastollisesti merkitsevä. Merkityksensä johtuu rakenteista, joissa *aivan*-adverbia taajaan käytetään: mm. omistusrakenteista ja kopulalauseista.

esittelemä COMMIT-esimerkki (mm. 'tehdä, harjoittaa') osoittaakin, että jokin preferenssi tiettyyn semanttiseen piirteeseen voi olla vain osittaista: kyseisellä verbillä on rikollista tai ei-suotavaa toimintaa merkitsevä preferenssi (esim. kollokaatit CRIME 'rikos', MURDER 'murha'), mutta preferenssi ei ole yksinomainen vaan kattaa kollokaattivariaatiosta vain noin 30 %.

Frekventtien astemääritteiden semanttisten preferenssien variaatio onkin yleensä laaja. Esimerkiksi astemääritteillä HYVIN, KOVIN ja OIKEIN on laaja kirjo semanttisessa preferenssissään, joskin näistä OIKEIN-määritteellä on voimakas preferenssi 'arvoa, hyvyyttä' merkitseviin ilmauksiin (mm. OIKEIN <HYVÄ, HYVIN, MAINIO, PAHA>) (Jantunen 2004: 138–139). Oletettavaa siis on, että kirjoitetussa kielessä pääasiassa astemääritteenä käytettävällä *aivan*-adverbilläkin on laaja semanttisen preferenssin variaatio. Näin onkin, joskin jo kollokaatioanalyysissä esiin nouseva 'vertailu' on selvästi ominainen kontekstuaalinen piirre tällä adverbillä: kaikista oikeanpuoleisista kollokaateista vertailua merkitseviä on peräti neljännes, 25,1 %. Semanttinen preferenssi *aivan* <'vertailu'> muodostuu pääasiassa taulukon 1 vertailua osoittavista kollokaateista, mutta myös niihin rinnastuvista alle 10 kertaa esiintyvistä kollokaateista, kuten *päinvastoin*, *näin* ja VASTAKKAINEN:

- (3) a. Todennäköisesti meidän seuraajiemme mielestä asia olisi *aivan päinvastoin*.
- b. Kuva ei kuitenkaan ole *aivan näin* yksinkertainen.
- c. Härkönenhän asui *aivan vastakkaisella* suunnalla.

Määritteelle on löydettävissä myös muita preferenssejä, mutta ne ovat selvästi vähemmän yleisiä kuin *aivan* <'vertailu'>. Näitä ovat *aivan* <'paikka' 10,3 %; 'määrä' 8,3 %> sekä 3–5 prosentin osuuksillaan yleisyysjärjestyksessä *aivan* <'aika, oikeellisuus, varmuus, erikoisuus, selvyys'>. Kaikkiaan nämä semanttiset preferenssit yhdessä 'vertailun' kanssa kattavat esiintymistä 64,4 %, eli käytetyn aineiston perusteella voidaan sanoa, että kaksi kolmesta *aivan*-lekseemin esiintymästä on kontekstissa, jossa ilmaistaan 'vertailua, paikkaa, määrää, aikaa, oikeellisuutta, varmuutta, erikoisuutta' tai 'selvyyttä'. Näistä on lueteltu esimerkkitapauksia konkordanssitaulukossa 2. Konkordansseista on nähtävissä, että semanttinen preferenssi voi näkyä myös muista asemista kuin O1, joka on tässä yhteydessä ollut pääasiallisena tarkastelukohteena. Tälle kontekstipiirteelle on ominaista, että se voi löytyä sanan molemmin puolin ja ilmetä eri sanaluokkaan kuuluvissa sanoissa ja rakenteissa. Esimerkiksi 'selvyys' voi ilmetä sanoista SELVÄ, *selvästi* ja KIRKASTUA 'selkiytyä, ymmärtää', 'paikka' puolestaan postpositiorakenteesta *oven vieressä* ja sanoista *lähelle*, *sängyssä* ja *rannoilla*. Taulukon ensimmäiseltä riviltä huomaa myös sen, että useat adverbille tyypilliset kontekstuaaliset semanttiset piirteet voivat esiintyä yhtä aika lähiympäristössä: kontekstissa on 'vertailua' (*toisennäköisten*), 'aikaa' (*nyt*) sekä 'paikkaa' (*puiden katveessa*) merkitseviä ilmauksia. Semanttisen preferenssin analyysi poikkeaa mm. kollokaatioanalyysistä siten, että tutkijan intuitio ja havaintokyky saavat suuremman roolin, sillä merkitykset eivät ole luonnollisestikaan yhtä läpinäkyviä ja yksiselitteisiä kuin kollokaatit.

Taulukko 2. Esimerkkejä *aivan*-adverbien semanttisista preferensseistä.

'vertailu'		
välikohtausta laivalla istuisin nyt	aivan	toisennäköisten puiden katveessa.
ja nokkeluus saattavat siis merkitä	aivan	muuta kuin mihin olemme tottuneet.
nimi koulun poissaolovihkossani on	aivan	yhtä tyyni ja eteerinen kuin
'paikka'		
läpsytti rappuset alas ja tuli kohti,	aivan	lähelle , hermostunut hymynhäive
Yksi oli tyhjä ja toisessa sängyssä	aivan	oven vieressä makasi nainen, joka
sijaitsevat lähes poikkeuksetta	aivan	rannoilla . Tosin ranta-asumiseen
'määrä'		
täällä jää vähälle , pahoittelin. Se on	aivan	olematonta , sanoi leijona. Vanki
sanoin koivule, että mie tykkään siitä	aivan	kauheasti . Ilosena muistin miten
tutkimuksen mukaan symboloivan	aivan	kaikkea ekstaasista melankoliaan.
'aika'		
rivejä jossakin määrin koko ajan . Jo	aivan	1930-luvun alussa vangittiin ja
en tulisi tarvitsemaan niitä jatkossa .	Aivan	kohta kertoisin Dorikselle uusista
kahvia. Sen on täytynyt olla aamulla	aivan	väsyksissä. Kun äiti sai Nipen
'oikeellisuus'		
vettä tuottava sähköboileri. - Muistat	aivan	oikein , mutta mitä merkitystä sillä on?
ei saa tuoduksi Suomeen. Tämä ei ole	aivan	totta . Esimerkiksi suomalaiset ostavat
sijaan aivoissa lähti pyörimään	aivan	väärä sana: suoni. Rajjan tilaama
'varmuus'		
niitä pitäisi juhlia. »Oletko todella	aivan	varma , että juhannuksena virvotaan?»
puhuttamisen sitten sieltä. - Olen	aivan	saletti että se on punkannut joko
komisario Kosminen. Minä tiedän	aivan	varmasti ettet sinä ole tehnyt mitään
'erikoisuus'		
korvaansa. - Mutta Korvatunturilla on	aivan	erityinen tehtävä ja sen sinä tulet
syntipukeiksi ei ole	aivan	epätavallista . Samankaltainen ilmiö
ja nuhruihin hytti alkoi kohta näyttää	aivan	oudolta . Lakanat muuttuivat
'selvyys'		
selitti. - Alkujaan maastossa olivat	aivan	selvät pyykit, mutta Kaupinkylän
uudessa sävyssä 1880-luvulta alkaen.	Aivan	selvästi maanvuokrakysymys oli
Asia ei näyttänyt hänelle vielääkään	aivan	kirkastuneen . - Näin ollen,

Aivan-adverbien kotekstia voidaan tarkastella myös kolligaatioiden avulla. Vasemmalla V1-positiossa on odotuksenmukaisesti eniten verbejä (taulukko 3), tai sitten positiossa on yleensä adverbi, substantiivi tai lauseen alku. O1-positiossa on puolestaan taajimmin adjektiiveja ja adverbija. Adverbikolligaateista viidesosa on astemääritteitä tai niiden

▷

tavoin käytettyjä genetiivisijaisia intensiteettisanoja, jotka määrittävät joko adjektiivia tai adverbia (*aivan erityisen tehokas*, *aivan hirvittävän kallis*, *aivan niin viaton*, *aivan liian harvoin*). Lauseenloppu esiintyy O1-positiossa vain harvoin (1 %); tällöin adverbia on käytetty tekstissä dialogipartikkelin funktiossa (*»Siis Dorotea Kiss...» »Aivan. Juuri hän», Louner kiirehti innokkaana.*) On luultavaa, että *aivan* on dialogipartikkelina huomattavasti tavallisempi puhekielessä, mutta tämä olisi tutkittava puhekielen aineistoista.

Taulukko 3. *Aivan*-adverbin vieruskolligaatit.

V1		O1	
verbit	46 %	adjektiivit	43 %
adverbit	15 %	adverbit	31 %
lauseen alku (LA)	14 %	konjunktiot	10 %
substantiivit	13 %	adpositiorakenteet	8 %
pronominit	5 %	substantiivit	3 %
konjunktiot	4 %	pronominit	2 %
adpositiorakenteet	2 %	verbit	2 %
adjektiivit	1 %	lauseen loppu (LL)	1%

Kun tavoitteena on tarkastella sanastoa fraseologisina yksikköinä eli valmiina ja toistuvina yhtä sanaa pitempinä kontekstuaalisia kytkentöjä sisältävinä rakenteina, on syytä käsitellä myös klustereita. Klusterit eli n-grammit ovat useamman kuin kahden peräkkäisen sanan muodostamia toistuvia lekseemikombinaatioita, kuten 3-grammeja ja 4-grammeja. Klustereita voi hakea tekstiaineistoista automaattisesti muun muassa WordSmith Tools -ohjelmopakettin Concord-ohjelman avulla; tämän esityksen aineistona olevasta tekstimassasta tehty automaattinen klusterianalyysi on liitteessä 2. Liitteen klusterilistauksessa on sekä 4- että 3-grammeja, jotka muodostuvat sananmuodoista, kuten *aivan varma että, hän oli aivan* (ei siis lemmoista HÄN OLLA *aivan*). Klusterilistasta huomaa helposti yleisimmät *aivan*-adverbin sananmuotoklusterit, ja näitä yhdistelemällä voi luoda ainakin osittaisia lemmaklustereita: *se oli aivan + se on aivan > SE OLLA aivan*. Jo tämän kaltainen tieto lekseemin klustereista voi olla hyödyllistä lekseemin käytön kuvauksessa, mutta konkordanssia tarkemmin analysoimalla voi kuvata klusteroitumisesta edelleen tarkentaa. Olen koonnut taulukkoon 4 taajimpia *aivan*-lekseemin 3-grammeja. Grammeissa on mukana sekä lemmoja että kolligaaatteja, jotta rakenteet voidaan kuvata mahdollisimman yleisellä tasolla. Grammeista on annettu esiintymisfrekvenssit. Esitys etenee vasemmasta kontekstista oikeaan. Positiotaulukoissa on tummennettu kolligaatin (sanaluokan) sisältävä 3-grammi, ja sen alla on esitetty grammeihin sisältyvät tavallisimmat rakenteet lemma- tai sananmuototasolla.

Taulukko 4. *Aivan*-adverbin kotekstin 3-grammit.

V2	V1	noodi	O1	O2	f
PRO	OLLA	<i>aivan</i>			160
SE	OLLA	<i>aivan</i>			46
HÄN	OLLA	<i>aivan</i>			38
MINÄ	OLLA	<i>aivan</i>			17
TÄMÄ	OLLA	<i>aivan</i>			17
	OLLA	<i>aivan</i>	ADJ		402
	OLLA	<i>aivan</i>	VARMA		49
	OLLA	<i>aivan</i>	HILJAINEN		16
	OLLA	<i>aivan</i>	SELVÄ		15
	OLLA	<i>aivan</i>	MAHDOTON		13
	OLLA	<i>aivan</i>	ERITYINEN		11
	OLLA	<i>aivan</i>	UUSI		10
	OLLA	<i>aivan</i>	TAVALLINEN		10
	OLLA	<i>aivan</i>	<i>eri</i>		10
	OLLA	<i>aivan</i>	ADV		172
	OLLA	<i>aivan</i>	<i>liian</i>		16
	OLLA	<i>aivan</i>	<i>yhtä</i>		16
	OLLA	<i>aivan</i>	<i>oikeassa</i>		12
		<i>aivan</i>	PRON/ SUBST	POSTPO	212
		<i>aivan</i>	SUBST	VIERESSÄ	24
		<i>aivan</i>	SUBST	EDESSÄ	14

Edellä olevasta taulukosta näkee, että niinkin taajaan käytetyllä ja kollokaatioiltaan varioivalla adverbilla kuin *aivan* näyttää olevan joitakin toistuvia 3-grammeja kotekstissaan. Vasemmanpuoleisessa kotekstissa tyypillisin 3-grammi on <PRO OLLA *aivan*>, josta tavallisimmat ovat <SE OLLA *aivan*> ja <HÄN OLLA *aivan*> (esim. 4). Keskikotekstin tavallisimmat grammit muodostavat OLLA-verbin kanssa yhdessä esiintyvät adjektiivit tai adverbit (esim. 5). Oikeanpuoleisessa koteksissa tyypillisin on postpositiorakenne, jossa on tavallisesti substantiiveja (esim. 6).

- (4) a. Sitä ei tarvitse pelätä, vaan *se on aivan* luonnollista ja ymmärrettävää.
 b. Tyttö muistaa kun *hän oli aivan* pieni ja iso piti häntä kädestä kun he hakivat keväällä sinivuokkoja.
- (5) a. Hän *oli aivan varma*, että myös mäyräkoira ja lammas olivat menossa niitylle.
 b. Esimerkiksi Klaaralla ei ole mitään tekemistä Liisa Voiman kanssa, joka *on aivan liian* hieno daami minun henkiselle tasolleni.



- (6) a. Ruumisjunat pysäytettiin pohjoiselle ratapihalle *aivan hautausmaan viereen*.
 b. Luentosali tunki itsensä väkisin Gretan aivoihin: nukkuva Edith *aivan luennoitsijan edessä*, miehen joka ei piitannut Edithin unesta.

Jos edellä kerrotun korpusanalyysin *aivan*-adverbista haluaisi tiivistää yhteen taulukoon, voisi adverbia kuvata fraseologisena yksikkönä taulukon 5 tapaan. Siihen on kerätty kuitenkin vain kaikkein tavallisimmat ja merkitsevimmät kontekstuaaliset rakenteet, eli adverbien käytön kannalta kyseessä on selkeästi pelkistys. Tällaista kuvausta voi pitää eräänlaisena lähtökohtana, jota voi syventää yksityiskohtaisemmillä kontekstuaalilla valintaa koskevilla havainnoilla.

Taulukko 5. *Aivan* fraseologisena yksikkönä kuvattuna frekventeimpien ja merkitsevimpien kontekstuaalisten valintojen avulla.

	V2	V1	noodi	O1	O2
kolligaatio		VERBI	<i>aivan</i>	ADJ	
klusterit	PRO (SE)	OLLA	<i>aivan</i>		
		OLLA	<i>aivan</i>	ADJ (VARMA)	
			<i>aivan</i>	PRON/SUBST	POSTPO
kollokaatio		OLLA	<i>aivan</i>	<i>kuin</i>	
sem. pref.			<i>aivan</i>	'vertailu'	

Otsikon virkkeessä *Minulla on aivan paljon rahaa* adverbi *aivan* on omistusrakenteessa. Käsillä olevan aineiston perusteella *aivan* on erittäin marginaalinen omistuslauseissa, sillä aineiston esiintymistä vain 16 (0,6 %) tapausta on sellaista, jossa adverbi tavataan yhdessä habitiivisen NP:n kanssa. Näissä *aivan* määrittää tavallisesti astemäärätteen tai vertailevan adjektiivin sisältävää NP-lauseketta:

- (7) a. Sitä paitsi *tällä naisella oli aivan käsittämättömän* ihana, käheän matala ääni, josta muistui mieleen vanhat kunnan amerikkalaiset elokuvat ja erityisesti Lauren Bacall.
 b. *Sillä oli aivan samanlaiset* eleet kuin sinulla.

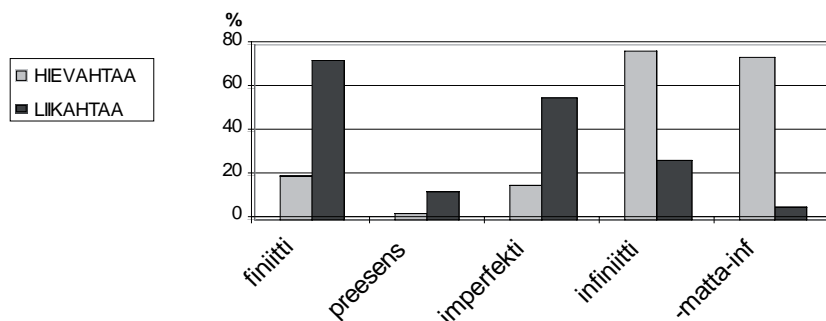
Todennäköisesti opetustilanteessa tulisi esiin myös kysymys *paljon*-kvanttorin saamista määritteistä, jos on tuotu ilmi, että *aivan* ei yleensä tätä kvanttoria määritä. Samaisen korpusaineiston perusteella sen tavallisimmat määritteet ovat demonstratiiviset *niin* (f=407; josta 3-grammit *niin paljon että* 77, *niin paljon kuin* 75) ja *yhtä* (3-grammi *yhtä paljon kuin* 59), interrogatiiviset *kuinka* (f=119) ja *miten* (f=69) sekä *aika* (f=58), *kovin* (f=53) ja *liian* (f=30). Suomen kielen tekstipankin mukaan (Kalevan vuosikerrat 1998 ja 1999) tyypilliset määritteet ovat samoja; kymmenen kärki tilastollisen merkitsevyyden mukaan näyttää seuraavalta: *niin, kuinka, yhtä, kovin, liian, hyvin, aika, mahdollisimman, miten ja erittäin*.¹³

¹³ *Paljon*-kvanttorin kollokoinnista ja klusteroitumisesta myös Jantunen (2004: 91, 95, 103, 176, 180) ja kollokoinnista oppijansuomessa Kallioranta (2009).

TAPAUK 2: SYNONYIMIT »LIIKAHTAA» JA »HIEVAHTAA»

Tsuin (2004, 2005) tekemästä kyselytutkimuksesta selviää, että kollokaatioiden lisäksi opettajat tarvitsevat usein tietoa myös synonyymisten ilmausten merkitys- ja käyttötapajeroista. Jo oman äidinkielenäkin perusteella kielennoppija oivaltaa, että (likipitäen) samaa merkitsevät ilmaukset eivät ole yleensä täysin synonyymisiä (täydellisen synonyymian mahdottomuudesta ks. mm. Lyons 1968: 447; Cruse 1986: 268, 270). Sen lisäksi että samaa merkitsevät ilmaukset eroavat toisistaan tyyppillisesti rekisteri-, tekstilaji- ja tyyli- ja -piirteiltään sekä konnotaatioiltaan (ks. esim. Cruse 2000: 158; Saukkonen 1984: 112), täydellistä synonyymiaa rikkoo myös monipolvinen leksikaalis-kieliopillis-semanttinen eli fraseologisen yksikön rakenteeseen liittyvä variaatio (ks. Jantunen 2001, 2004). Tarkastelen seuraavaksi, miten synonyymiaa ja kahden synonyymisen ilmauksen eroja voi selittää korpuustiedon avulla. Keskityn *aivan*-esimerkistä poiketen morfologiseen preferenssiin ja semanttiseen prosodiaan, joita en vielä adverbin yhteydessä käsitellyt esimerkkien avulla. Käytän havainnollistavana tapauksena *liikahtaa–hievahtaa*-synonyymiparia.

Käytettävän korpusaineiston perusteella on ensinnäkin havaittavissa, että LIIKAHTAA on synonyymiaan selvästi frekventimpi: se esiintyy aineistossa 128 kertaa, kun taas HIEVAHTAA 43 kertaa.¹⁴ Käyttöyleisyys siis jo erottelee lekseemit toisistaan. Lekseemien paradigmaattinen morfologinen preferenssi eli eri taivutusmuotojen yleisyys on kuvattu kuviossa 1. Taivutusmuodoista on otettu huomioon finiitti- ja infiniittimuotojen osuudet (kuviossa preesensin ja imperfektin osuudet sisältyvät finiittimuotojen, MATTA-infinitiiviin infiniittiuviin osuuteen). Taivutusmuotojen osuudet ovat selvästi erilaiset, ja sanojen muotojen määrien ero on tilastollisesti erittäin merkitsevä¹⁵. Merkitsevimmät erot ovat finiittimuotojen osuuksissa (19 % vs. 73 %)¹⁶, imperfektimuotojen osuuksissa (16 % vs. 56 %)¹⁷, infiniittimuotojen osuuksissa (81 % vs. 27 %)¹⁸ ja MATTA-infinitiivin osuuksissa (76 % vs. 5 %)¹⁹.



Kuvio 1. Synonyymien HIEVAHTAA ja LIIKAHTAA paradigmaattinen morfologinen preferenssi.

¹⁴ Myös Suomen kielen tekstipankin koko suomenkielisessä sanomalehtiaineistossa LIIKAHTAA on tavallisempi: LIIKAHTAA f=398, HIEVAHTAA f=96.

¹⁵ $\chi^2=118.1$, $df=5$, $p<0.001$.

¹⁶ $z=6.34$, $p<0.001$.

¹⁷ $z=4.55$, $p<0.001$.

¹⁸ $z=5.83$, $p<0.001$.

¹⁹ $z=9.32$, $p<0.001$.

Tavallista siis on, että LIIKAHTAA esiintyy finiittimuotoisena (usein imperfektissä) ja HIEVAHTAA puolestaan MATTA-infinitiivissä, kuten esimerkeissä 8 ja 9.

- (8) a. Että kinkku, äiti sanoi ja hänenkin suupielensä *liikahtivat* vähän.
b. Kun puoli vuotta oli kulunut syksyllä 1996, vain muutama indikaattori oli *liikahtanut*.
- (9) a. Suljin stereot ja istuin nojatuolissa *hievahtamatta*. Pitkään oli hiljaista, mutta sitten ääni alkoi kuulua uudelleen.
b. Poika istui keulassa ja tapitti minua *hievahtamatta*.

Näiden taivutusmuotopreferenssien lisäksi synonyymit poikkeavat toisistaan myös sen mukaan, miten niistä tehdään johdoksia: aineiston perusteella LIIKAHTAA on selvästi useammin kantasananä käytetty. Sen johdoksia on aineistossa yhteensä 61 (LIIKAHDELLA, LIIKAHDUS, LIIKAHTAMATON, LIIKAHTAMINEN), mutta HIEVAHTAA-kantasanan johdoksia ei löydy ainuttakaan.²⁰

HIEVAHTAA löytyy usein kontekstista, jossa on jotain negatiivista: 91 % esiintymistä on sellaisia, joissa kontekstista ilmenee jotain uhkaavaa, pelottavaa tai ei-toivottavaa. Samoin on LIIKAHTAA-lekseeminkin kanssa, sillä 84 % esiintymistä on samantapaisessa kontekstissa. Esimerkit 10, 11 ja 12 kuvaavat tällaisia esiintymiä; kyseessä on siis molemmilla sanoilla negatiivinen semanttinen prosodia.

- (10) Elena Wekmann oli saanut odottaa siinä kuolemaansa aivan rauhassa. Veterinäkuljettaja oli kertonut naisen seisseen kasvot junaan päin tyynenä ja *hievahtamatta*, kädet uumalla ristikkäin kuin tanssiin kutsua odottava neito.
- (11) Mutta tiheä oksaverkko sulki kaikki ovet ja ikkunat niin tiukkaan, ettei niitä saanut *hievahtamaankaan*. Koska koneet eivät olleet enää aikoihin käyneet, olivat myös valot käytävissä ja hyteissä sammuneet. Aavemaisessa hämärässä lojuivat valkotukkaiset tytöt ja pojat, raukeina happinaamareittensa takana, pitkin penkkejä ja seinustoja.
- (12) Mä romuttaisin kaikkien suunnitelmat. Mä en *liikahtaisi* huoneestani vanhojen päivänä. Mä varasin huoneeseeni kaksi termoskannullista kahvia, valmiita leipiä ja omenoita, hyviä kirjoja ja lukitsin oven.

Kotekstianalyysi siis paljastaa, että molemmat lekseemit ovat pragmaattiselta funktioltaan »kieltohakuisia» eli tekstuaalinen ympäristö, jossa niitä käytetään, on tyypillisesti jotain kielteistä sisältävää. Varsinainen kieltohakuisuus näkyy selvästi etenkin HIEVAHTAA-verbin käytössä: sen 43 esiintymästä 9 kollokoi EI-kieltoverbin kanssa ja loput ovat MA-infinitiivin abessiivimuotoja (ks. kyseisen verbin kieltohakuisuudesta myös VISK §1637), näistä on esimerkkejä yllä. LIIKAHTAA ei ole yhtä ehdottomasti kieltohakuinen, sillä vain puolet (f=65) on kiellon vaikutuksen alaisia ja MA-infinitiivin abessiivimuotoja. Näin ollen HIEVAHTAA-verbin merkitys onkin oikeastaan liikkumattomuus. Näistä kahdesta HIEVAHTAA-sana valitaan kontekstiin siis juuri siksi, että se funktionaalisenä fraseologisena yksikkönä kertoo sen, että **mikään ei liiku**. LIIKAHTAA-verbiä käytetään yhtä hyvin kuvaamassa sitä,

²⁰ Google-hakukoneella tehdyssä internethaussa kyseisiä (perusmuotoisia) LIIKAHTAA-lekseemin johdoksia löytyi 6625 ja vastaavia HIEVAHTAA-lekseemin johdoksia 152.

että jokin liikkuu, kuin sitä, että liikettä ei synny. Liikkuminen ei kuitenkaan ole yleensä pragmaattiselta funktioltaan positiivista, kuten verbin semanttinen prosodia paljastaa.

HIEVAHTAA saa tavallisesti kotekstiinsa 'seisomista', 'istumista' tai 'katsomista' tarkoittavia verbejä: näitä on 72 %:ssa aineiston kaikista verbin esiintymistä. Myös LIIKAHTAA-lekseemin kotekstissa on tällaisia verbejä, mutta niiden osuus on huomattavasti pienempi: 13 %. Näistä semanttisista preferensseistä on esimerkkejä alla.

- (13) Iso tarjoili punaista Jaffaa ja munkkipossua, hän *istui hievahtamatta* tuolillaan ja iso istui hänen vieressään sinisellä jakkaralla.
- (14) Tämä seiso i hajasäärin, hikinen yläruumis paljaana, voimaa uhoten ja *tuijotti hievahtamatta* tyttöä. Veljen kasvot loistivat.
- (15) Mies *seisoi* kuljettajan oven edessä eikä *liikahtanu*.

Synonyymeistä olisi mahdollista tehdä *aivan*-adverbin tapaan huomattavasti tarkempikin analyysi, mutta tässä yhteydessä olen havainnollistanut lähinnä synonymiaa rikkovaa morfologista ja semanttista variaatiota. Molempia verbejä käytetään tyypillisesti sellaisissa yhteyksissä, joissa liike ei ole diskurssissa toivottavaa tai tilanne, jossa liike tai liikkumattomuus ilmenee, on uhkaava tai pelottava. Synonyymit puolestaan eroavat selvästi morfologisilta ja kollokaattien merkitykseen liittyviltä piirteiltään. Niinpä HIEVAHTAA valitaan synonyyminsä sijaan tyypillisesti silloin, kun momentaanista liikettä merkitsevää verbiä halutaan käyttää infinitiivimuotoisena, erityisesti MATTA-infinitivissä, kun kotekstissa on 'istumista', 'seisomista' ja 'katsomista' merkitseviä sanoja ja erityisesti silloin kun halutaan ilmaista, että tilanteessa ei liikettä tapahdu.

FRASEOLOGISET YKSIKÖT, KORPUKSET JA OPETUS

Edellä olen käsitellyt lekseemejä fraseologisina yksikköinä ja tuonut esiin sen, miten korpuksat ja korpuksista havaittava toistuvuus paljastavat »piilossa olevia» rakenteita. Korpusten olemassaolo yksistään ei kuitenkaan riitä siirtämään niiden sisältämää tietoa opetukseen, vaan sanaston fraseologiselle opetukselle on olemassa lukuisia haasteita ja avoimina olevia kysymyksiä. Ensiksikin olisi pystyttävä vastaamaan kysymyksen, missä vaiheessa sanastoa ruvetaan opettamaan yhtä sanaa laajempina kokonaisuuksina käyttösääntöineen. Onko tämä hyödyllistä aivan kielenoppisen ensimetreillä vai vasta myöhemmin kielitaidon jo edistyttyä? Jos pidämme paralleelina ilmiönä kielenomaksumista, voisimme olettaa, että tyypillisten fraseologisten rakenteiden pitäisi olla läsnä kielenopetuksen alusta alkaen aivan samoin kuin lapsi kuulee kieltä omaksuessaan tyypillisiä rakenteita omaksumiskonteksteissaan. Toiseksi pohdintaa vaatii näin opettavien leksikaalisten yksiköiden valinta. On todennäköistä, että kielen yleisempien sanojen osalta tällainen lähestymistapa on tärkeää, mutta se on samalla myös haastavaa: taajaan käynteillä sanoilla on lukuisia merkityksiä ja käyttötapoja, joiden kuvaaminen on haasteellista — mutta kielenoppijan kannalta myös toivottavaa. Selvää kuitenkin on, että esimerkiksi erittäin taajakäyttöisiä kieliopillisia sanoja, kuten konjunktioita ja pronomineja, ei voi esittää tarkkarajaisina fraseologisina yksikköinä. Synonyymien ja luultavasti antonyymienkin opettamisessa lähestymistavasta on kuitenkin suuri hyöty. Nesselhauf (2003:

▷

238) painottaa, että opettavien kollokationaalisten yksiköiden olisi esiinnyttävä taajaan ns. neutraaleissa rekistereissä ja myös sellaisissa rekistereissä, joita kielenoppijat tulevat käyttämään opiskellessaan ja sen jälkeen.

Haasteita on myös siinä, että vielä ei ole tarpeeksi tietoa siitä, millä tavalla fraseologiaa yksiköitä tulisi opettaa, tai ylipäätään siitä, kuinka suuri osa kielestä on niin sanotusti fraseologista (Kennedy 2008: 38–39; Granger ja Meunier 2008b). Koska lähestymistapa on kielenopetuksen saralla melko uusi (ks. myös Mauranen 2004: 100), ei esimerkiksi opettajilla ole toistaiseksi valmiuksia yhtäältä analysoida aineistoja kontekstuaalisia piirteitä esiin kaivaakseen, toisaalta käyttää tätä tietoa hyväksi opetuksessa.²¹ Ylipäätään koko fraseologisuus-ilmiö kaipaisi enemmän huomiota ja se pitäisi tuoda opetuksessa eksplisiittisesti esille (Nesselhauf 2005: 252–254). Myöskään korpuksia hyödyntävä tietokoneavusteinen oppiminen, jossa oppija etsii itse tietoa sanaston ja rakenteiden käytöstä, ei ole käsittääkseni suomen kielen opetukseen rantautunut (ks. aiheesta mm. Bernardini 2004). Heikkoutena on siis tradition puute. Korpuksia ja niiden analysoimiseen vaadittavat menetelmät voidaan kokea jopa jonkinlaiseksi uhkaksi opetustilanteessa, koska niiden tarjoama tieto voi rikkoa hyväksi havaittuja rutiineja (Sinclair 2004c: 271). Lisäksi analyysit voivat olla monimutkaisia; sääntöjä ja kuvauksia olisi pystyttävä yksinkertaistamaan tarpeeksi opetuksen ja oppijan tarpeisiin soveltuvaksi (ks. Granger ja Meunier 2008b: 249). Tradition puuttuminen ei kuitenkaan johdu pelkästään opetusmenetelmän vieraudesta: myös tiedon saaminen kielen fraseologisista yksiköistä on monesta syystä hankalaa (mts. 248). Nykyiset tietokonelukuiset tekstiaineistot ovat hankalasti saavutettavissa, sillä niiden käyttöluvut on määritelty yleensä vain tutkimustarkoituksiin. Aineistoja eivät yleensä pääse käyttämään muut yliopistojen ja tutkimusyksiköiden henkilökuntaan kuuluvat, mikä jättää ulkopuolelle suuren joukon kielenopettajia. Lisäksi suomalaiset korpusaineistot ikääntyvät nopeasti, sillä esimerkiksi Suomen kielen tekstipankin tuoreimmatkin tekstit ovat vuodelta 2000, mikä tarkoittaa sitä, että kielen uusimpia ilmauksia ei sen avulla tavoiteta. Opetuksen (ja tutkimuksenkin) näkökulmasta olisi tarvetta niin koulutukselle kuin aineistojen laajamittaisemmalle hyödynnettävyydelle ja ajantasaisuudelle esimerkiksi niin sanottujen monitorikorpuksien avulla.

Opetuksen tueksi olisi saatava myös enemmän todellisen kielenkäytön tutkimukseen perustuvaa materiaalia, kuten oppikirjoja ja sanakirjoja. Tähän mennessä on tehty havaintoja siitä, että esimerkiksi oppikirjoissa esitetty kieli voi olla yllättävänkin paljon todellisesta kielenkäytöstä poikkeavaa (ks. aiheesta esim. Tanner (2008) ja siinä mainitut lähteet) ja että opetuksessa saatetaan opettaa ilmauksia, joita kielessä ei esiinny tai joiden frekvenssi tai variaatio on erilainen kuin opetuksessa tuodaan esille (Mauranen 2004: 96; ks. myös Tsui 2004: 40–41). Oppikirjoille onkin monia vaatimuksia, eikä todellisen kielenkäytön (mukaan lukien ilmausten frekvenssi- ja kontekstitieto) huomioon ottaminen ole niistä kaikkein helpoimpia. Se on kuitenkin tärkeää, sillä kielen käyttöä, esimerkiksi fraseologiaa yksiköitä, opitaan runsaasti myös implisiittisesti oppimateriaalin avulla (ks. Kennedy 2008: 38–39). Suomen kielen sanakirjat eivät puolestaan yleensä anna kovinkaan paljon vinkkejä leksikaalisten yksiköiden fraseologisuudesta, vaan ilmaukset selitetään

²¹ Oulun yliopiston suomi toisena ja vieraana kielenä -oppiaineessa opiskelijoita on alettu kouluttaa käyttämään kieliaineistoja oman opetustyönsä tukena.

merkityksenselityksen, synonyymien ja esimerkkilauseiden avulla, tai pelkästään jollakin näistä. Tilanne on tällainen etenkin pienissä opiskelijasanakirjoissa. Kieliopillisten piirteiden opettaminen onkin saanut paljon enemmän huomiota, ja sanaston opettamisen metodit ovat jääneet osittain tämän jalkoihin (ks. Aalto, Latomaa ja Suni 1997: 536, 539; Niiranen 2004: 81). Opetuksen tueksi tarvitaan yhä tarkemmin todellista kielenkäyttöä kuvaavia oppikirjoja ja sanakirjoja tai sanastoja, joissa sanastoa esitellään sanojen käyttökontekstia ja kotekstuaalista sidoksisuutta kuvaten. Yksi tällainen yritys on vuonna 2008 alkanut Conlexis-hanke, jossa tuotetaan pilottiluontoisesti suomen kielen verkkosanasto opetuskäyttöön (Jantunen ja Tammimies 2008).

Korpuspohjaisessa opetuksessa ja tutkimuksessa korostuu hyvin usein toistuvuus ja variaation esittäminen jää vähäiselle osalle. Vaikka alkeisopetuksessa onkin kiinnitettävä huomio toistuviin ja suuritajuisiin kielenpiirteisiin (ks. sanaston osalta Puro 1999, myös Kennedy 2008), opetus on vääristynyttä, jos siinä keskitytään vain yleisiin ilmiöihin ja jätetään huomiotta ilmausten variaatio (Conrad 2004: 68–70). Opetuksessa ja oppikirjatutonnossa lienee kuitenkin mahdollista löytää keskitie toisaalta tyypillisyyden ja toisaalta variaation välillä. Lisäksi korpuksot ovat toisinaan myös rajallisia ja vinoutuneita, mikä vaikuttaa niistä saatavaan tietoon kielen rakenteista — mutta eivätkö sellaisia ole myös oppikirjat ja sanakirjat?

LÄHTEET

TUTKIMUSAINESTOT

- Käännössuomen korpus.* Käännössuomen sähköinen tutkimusaineisto. Koostanut Käännössuomi ja kääntämisen universaalit: tutkimus korpusaineistolla -hankkeessa Joensuun yliopiston kansainvälisen viestinnän laitos 1997–2001. Saanti: Joensuun yliopisto, vieraat kielet ja käännöstiede.
- Suomen kielen tekstipankki.* Koostanut Kotimaisten kielten tutkimuskeskus ja Tieteen tietotekniikan keskus CSC. Saanti: <http://www.csc.fi/>.

KIRJALLISUUS

- AALTO, EIIJA – LATOMAA, SIRKKU – SUNI, MINNA 1997: Suomi toisena ja vieraana kielenä: tutkittua ja keskusteltua. – *Virittäjä* 101 s. 530–562.
- ARPPE, ANTTI 2001: Corpus-based observations on the use and occurrence of inflected word forms in Finnish: The case of synonyms. – Jussi Niemi & Janne Heikkinen (toim.), *Nordic and Baltic morphology: Papers from a NorFA course. Tartu, June 2000* s. 1–14. Joensuu: University of Joensuu.
- 2002: The usage patterns and selectional preference of synonyms in a morphologically rich language. – Annie Morin & Pascal Sébillot (toim.), *Proceedings of the 6th international conference on textual data statistical analysis (JADT 2002)*. Vol. 1 s. 21–32. Rennes: Inria.
- BERNARDINI, SILVIA 2004: Corpora in the classroom: An overview and some reflections on future developments. – John McH. Sinclair (toim.), *How to use corpora in language*

- teaching* s. 15–36. Studies in corpus linguistics 12. Amsterdam: Benjamins.
- BIBER, DOUGLAS – CONRAD, SUSAN 1999: Lexical bundles in conversation and academic prose. – Hilde Hasselgård & Signe Oksefjell (toim.), *Out of corpora: Studies in honour of Stig Johansson* s. 181–190. Amsterdam: Rodopi.
- CONRAD, SUSAN 2004: Corpus linguistics, language variation, and language teaching. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 67–85. Studies in corpus linguistics 12. Amsterdam: Benjamins.
- CRUSE, D. ALAN 1986: *Lexical semantics*. Cambridge: Cambridge University Press.
- 2000: *Meaning in language: An introduction to semantics and pragmatics*. Oxford: Oxford University Press.
- ELLIS, NICK C. 2008: Phraseology: The periphery and the heart of language. – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in foreign language learning and teaching* s. 1–13. Amsterdam: Benjamins.
- GRANGER, SYLVIANE 1998: Prefabricated patterns in advanced EFL writing: Collocations and formulae. – Anthony P. Cowie (toim.), *Phraseology: Theory, analysis and applications* s. 145–160. Oxford: Oxford University Press.
- 2005: Pushing back the limits of phraseology: How far can we go? – Christelle Cosme, Céline Gouverneur, Fanny Meunier & Paquot Magali (toim.), *Pre-proceedings of Phraseology 2005. The many faces of Phraseology. An interdisciplinary conference. Louvain-la-Neuve Belgium, 13–15 October 2005* s. 165–168.
- GRANGER, SYLVIANE – MEUNIER, FANNY 2008a: Introduction. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspective* s. XIX–XXVIII. Amsterdam: Benjamins.
- 2008b: Phraseology in language learning and teaching: Where to from here? – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in foreign language learning and teaching* s. 247–252. Amsterdam: Benjamins.
- GRÖNHOLM, MAIIA 2000: *Pörröinen nalle söi herkullisen jäätelön*: Kollokaatiotaidoista suomen kielessä (L2). – Eeva Lähdemäki & Yvonne Bertills (toim.), *Femistica fausta in honorem Mauno Koski septuagenarii. Juhlakirja Mauno Kosken 70-vuotispäivän kunniaksi 21.2.2000* s. 115–135. Åbo: Åbo Akademi.
- 2001: Idiomien ja kollokaatioiden oppiminen suomenkielisessä kielikylvyssä. – Helena Sulkala & Leena Nissilä (toim.), *XXVII kielitieteen päivät Oulussa 19.–20.5.2000* s. 55–62. Acta Universitatis Ouluensis B 41. Oulu: Oulun yliopisto.
- HUNSTON, SUSAN 2007: Semantic prosody revisited. – *International journal of corpus linguistics* 12 (2) s. 249–268.
- JANTUNEN, JARMO H. 2001: *Tärkeä seikka ja keskeinen kysymys*: Mitä korpuslingvistinen analyysi paljastaa lähisyronnyymeista? – *Virtittäjä* 105 s. 170–192.
- 2004: *Synonymia ja käännessuomi: Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin*. Joensuun yliopiston humanistisia julkaisuja 35. Joensuu: Joensuun yliopisto.
- 2008: Haasteita oppijankielen korpusanalyysille: oppijankielen universaalit. – Pille Eslon (toim.), *Õppijakeele analüüs: võimalused, probleemid, vajadused* s. 67–92. Tallinna ülikooli eesti filoloogia osakonna toimetised 10. Tallinn.
- 2009: Ei pelkästään mielikuvituksen puutteen vuoksi – kieliaineistojen systemaattinen käyttö kielentutkimuksessa. – *Virtittäjä* 113 s. 101–113.

- JANTUNEN, JARMO H. – TAMMIMIES, TANJA 2008: *Conlexis*. Hankesuunnitelma. Mimeo. Oulun yliopiston suomi toisena ja vieraana kielenä -oppiaine ja Oulun ammatillinen aikuiskoulutuskeskus.
- JÖNSSON-KORHOLA, HANNELE – WHITE, LEILA 1997: *Tarkista tästä: Suomen sanojen rektioita suomea vieraana kielenä opiskeleville*. Helsinki: Finn Lectura.
- KALLIORANTA, OTTO 2009: *Paljon*-adverbin kollokointi oppijansuomessa: korpusvetoinen tutkimus. Pro gradu -tutkielma. Oulun yliopisto, suomen kieli. <http://www oulu.fi/hutk/sutvi/oppijankieli/tutkimus>.
- KENNEDY, GRAHAM 2008: Phraseology and language pedagogy: Semantic preference associated with English verbs in the British National Corpus. – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in foreign language learning and teaching* s. 22–41. Amsterdam: Benjamins.
- KUUSI, ANNA-LEENA 1971: *Johdatusta suomen kielen fraaseologiaan*. Suomi 115 (4). Helsinki: Suomalaisen Kirjallisuuden Seura.
- LOUW, BILL 1993: Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. – Mona Baker, Gill Francis & Elena Tognini Bonelli (toim.), *Text and Technology. In Honour of John Sinclair* s. 157–176. Amsterdam: Benjamins.
- LYONS, JOHN 1968: *Introduction to theoretical linguistics*. Cambridge University Press.
- MAURANEN, ANNA 2000: Strange strings in translated language. A study on corpora. – Maeve Olohan (toim.), *Intercultural faultlines. Research models in translation studies I: Textual and cognitive aspects* s. 119–141. Manchester: St. Jerome.
- 2004: Spoken corpus for an ordinary learner. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 89–105. Studies in corpus linguistics 12. Amsterdam: Benjamins.
- MOON, ROSAMUND 2008: Dictionaries and collocation. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspective* s. 313–336. Amsterdam: Benjamins.
- MOT KTS = *MOT Kielitoimiston sanakirja 2.0*. Toimittaneet Eija-Riitta Grönros, Minna Haapanen, Tarja Heinonen, Leena Joki, Liisa Nuutinen, Marjatta Vilkkamaa-Viitala, Riitta Eronen & Outi Lehtinen. Kotimaisten kielten tutkimuskeskuksen julkaisuja 149. Helsinki: Kotimaisten kielten tutkimuskeskus ja Kielikone Oy 2008.
- MUIKKU-WERNER, PIIRKKO – JANTUNEN, JARMO H. – KOKKO, OSSI 2008: *Suurella sydämellä ihan sikana. Suomen kielen kuvaileva fraasisanakirja*. Helsinki: Gummerus.
- NENONEN, MARJA 2002: *Idiomit ja leksikko: Lausekeidiomien syntaktisia, semanttisia ja morfologisia piirteitä suomen kielessä*. Joensuun yliopiston humanistisia julkaisuja 29. Joensuu: Joensuun yliopisto.
- NESSSELHAUF, NADJA 2003. The use of collocations by advanced learners of English and some implications for teaching. – *Applied Linguistics* 24 (2) s. 223–242.
- 2005: *Collocations in a learner corpus*. Amsterdam: Benjamins.
- NIIRANEN, LEENA 2004: Oppimisen kohteena suomen verbit: Kaksikielisten oppilaiden ja luokkahuoneoppilaiden verbisanaston ja verbintaivutuksen vertailua – Sirkkua Latomaa (toim.), *Äidinkieli ja toiset kielet: Pohjoismainen kaksikielisyytyöpaja Tampereella 18.–20.10.2002* . 79–96. Tampere Studies in Language, Translation and Culture. Series B 1. Tampere: Tampere University Press.

▷

- PURO, TARJA 1999: Sanastollinen tieto ja suomen kielen oppikirjojen sanasto. – *Virittäjä* 103 s. 2–26.
- SAUKKONEN, PAULI 1984: *Mistä tyyli syntyy?* Helsinki: WSOY.
- SAUKKONEN, PAULI – HAIPUS, MARJATTA – NIEMIKORPI, ANTERO – SULKALA, HELENA 1979: *Suomen kielen taajuussanasto*. Helsinki: WSOY.
- SCOTT, MIKE 2006: *WordSmith Tools online manual, version 4.0*. Oxford: Oxford University Press.
- SIEPMANN, DIRK 2008: Phraseology in learners' dictionaries: What, where and how? – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in foreign language learning and teaching* s. 185–202. Amsterdam: Benjamins.
- SINCLAIR, JOHN 1991: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- 1996: The Search for Units of Meaning. – *Textus IX* s. 75–106.
- 2004a: *Trust the text: Language, corpus and discourse*. London: Routledge.
- 2004b: Introduction. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 1–10. Studies in corpus linguistics 12. Amsterdam: Benjamins.
- 2004c: – New evidence, new priorities, new attitudes. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 271–299. Studies in corpus linguistics 12. Amsterdam: Benjamins.
- STUBBS, MICHAEL 1995: Collocations and semantic profiles: On the cause of the trouble with quantitative studies. – *Functions of Language 2* (1) s. 23–55.
- 2001: *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- TANNER, JOHANNA 2008: S2-oppikirjojen dialogien tarkastelua. – Johanna Tanner & Marja Kokkonen (toim.), *Suomenoppiminen, kielitaito ja tutkimus* s. 163–189. Kakkoskieli 6. Helsingin yliopisto: Suomen kielen ja kotimaisen kirjallisuuden laitos.
- TOGNINI BONELLI, ELENA 2008: Phraseology as the starting point in a corpus-driven perspective. Esitelmä konferenssissa International seminar on corpus linguistics: New trends in language teaching and translation studies. In honour of John Sinclair. Granada, Spain 22.–24. syyskuuta 2008.
- TSUI, AMY B. M. 2004: What teachers have always wanted to know - and how corpora can help. – John Sinclair (toim.), *How to use corpora in language teaching* s. 39–61. Studies in corpus linguistics 12. Amsterdam: Benjamins.
- 2005: ESL teachers' questions and corpus evidence. – *International journal of corpus linguistics* 10 (3) s. 335–356.
- VISK = HAKULINEN, AULI – VILKUNA, MARIA – KORHONEN, RIITTA – KOIVISTO, VESA – HEINONEN, TARJA RIITTA – ALHO, IRJA 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoversio, viitattu 20.8.2008. <http://scripta.kotus.fi/visk>.
- WHITSITT, SAM 2005: A critique of the concept of semantic prosody. – *International journal of corpus linguistics* 10 (3) s. 283–305.

Liite 1. Korkordanssiesimerkki *aivan*-lekseemistä tuotettuna WordSmith Tools -ohjelmapaketin Concord-ohjelmalla.

N	Concordance	File
11	varo niitä, jotka kehuvat varallisuudellaan. He ovat aivan varmasti tyhjätskuja. Rikkaat yrittävät	spo006.txt
12	sairaalaa, jossa hän vietti yön. Hän olisi aivan varmasti välittänyt koko sairaalakeikalta, jos hän	spo010.txt
13	itkevä, naurava ja tunteva mies? Heitä on aivan varmasti, tiedän sen. Monet naisetkin esiintyvät	spo006.txt
14	kävelyä päivässä. Se ei ole paljon, mutta sillä on aivan varmasti merkitystä painonhallinnassa, jos	spo007.txt
15	Pikkukuu istahti murheissaan katkivelle. Hän oli aivan varma, että myös määräkoira ja lammas olivat	lsia003.txt
16	jossain metsän sydämessä. Kahdeksalta hän oli aivan varma, että oli hiihtänyt Liittokodalle menevään	lsia001.txt
17	ei kuitenkaan enää ollut näkemässä, ja nyt hän on aivan varma että... Nikon jalka alkoi taas vatkata	sde005.txt
18	kiven katveessa? Seuraavana aamuna hän oli aivan varma. Sillä aivan sortuneen laiturin eteen,	lsia009.txt
19	ja jos lääkkeitä ei olisi ollut, hän olisi kuollut, hän oli aivan varma, että olisi kuollut. Elisa ihmetteli, kuinka oli	lsvi003.txt
20	kääntyi ja kietoi kätensä Lindan ympärille. - Minä olen aivan varma, että Amir tuo hänet takaisin! hän sanoi.	sde004.txt
21	äänestä oli jo äskeinen pehmeys kaukana. - Se on aivan varma että tässä talon alla kulkee joku suoni. Ei	sde001.txt
22	Possu ei sanonut enää mitään. Mutta se oli aivan varma, että noita teräviä kiviä ei vielä eilen ollut	lsia009.txt
23	minne. Jossain ulkomaillaakin he kävivät ja minä olin aivan varma, että heistä tulisi pari. Että Johannes oli	lsvi005.txt
24	kuinka totta Anneli näkee. - Kyllä minä olen aivan varma. - Jos sitten näkyy kasvot kuvassa ja	lska010.txt
25	näköinen, mutta suomalainen hän oli, siitä olen aivan varma. - Nyt sitten soittamaan poliisille, että	sde003.txt
26	liikaa. Jostus rahalla voi paikata orkesterin, joka on aivan valmis hautaan. Tästä on olemassa yksi hyvä	lska009.txt
27	aina selvää, oliko hän tosissaan vain ei. Hän oli aivan vakava, mutta odotin hänen purskahtavan	lska017.txt
28	vilkaisi Kirikkää huikan ihmeissään, mutta tämä oli aivan vakava, ei nähnyt mitään outoa siinä, että hän	lsia018.txt
29	töihin, hän tunsii tämmöiset ihmiset... - Ne ovat aivan vailla tuntoa ja ne kyllä nakertavat sinut hengiltä,	lska010.txt
30	ja katsoi tutkivasti taikuria. - Kuulin että teillä oli aivan uusi tempu työn alla. - Ei enää työn alla, vaan	sde001.txt
31	Kisu Pikkukuu sydän pamppaili oudosti. Tämä oli aivan uusi tunne. Jotain erityistä oli tapahtunut hänelle.	lsia003.txt
32	Mitä hän tiesi isän elämästä? Ehkä hänellä oli aivan uusi naisystävä, josta hän oli niin innoissaan,	lsvi004.txt
33	olet tiennyt jo pitkään, mutta joka minulle on aivan uusi. Sofi kävi täällä torstaina kuten ehkä jo	lska011.txt
34	kadulla olin sylkäissyt hänet nähdessäni. Sehän oli aivan uskomaton. Minä en koskaan syljeskellyt, en	lska027.txt
35	työn edessä, Lissu katsoi selkäni taakse. - Hän on aivan upea. Käännyn hetken kuluttua ympäri ja	lsvi005.txt
36	ei ole tänä talvena aiettu kertaakaan millään. Se oli aivan umpista, mutta sinne se vain vritti kääntvä.	lsia001.txt

concordance
collocates plot patterns clusters filenames source text notes
160 Set

Liite 2. Esimerkki klusterianalyysistä tuotettuna WordSmith Tools -ohjelmapaketin Concord-ohjelmalla. Klusterista oikealle on mainittu klusterin lukumäärä, pituus saneina ja siihen liittyvät muut klusterit.

Concord					
File Edit View Compute Settings Windows Help					
N	Cluster	req.	ngth	Related	
1	AIVAN VARMA ETTÄ	28	3	oli aivan varma että (7),olin aivan varma että (6)	
2	HÄN OLI AIVAN	25	3	hän oli aivan varma (5)	
3	AIVAN SAMOIN KUIN	21	3		
4	SE OLI AIVAN	21	3		
5	EI OLE AIVAN	19	3		
6	AIVAN KUIN OLISI	17	3		
7	OLI AIVAN HILJAISTA	14	3		
8	EI OLLUT AIVAN	14	3		
9	SE ON AIVAN	12	3		
10	OLI AIVAN VARMA	10	3	oli aivan varma että (7),hän oli aivan varma (5)	
11	AIVAN MUUTA KUIN	10	3		
12	AIVAN SAMALLA TAVALLA	9	3		
13	JOTAKIN AIVAN MUUTA	9	3		
14	OLI OLLUT AIVAN	8	3		
15	TÄMÄ ON AIVAN	8	3		
16	AIVAN ERI ASIA	8	3		
17	JOTAIN AIVAN MUUTA	8	3		
18	OLEN AIVAN VARMA	8	3		
19	KUIN JOKU OLISI	7	3		
20	OLI AIVAN VARMA ETTÄ	7	4	aivan varma että (28),oli aivan varma (10)	
21	OLI JO AIVAN	7	3		
22	MINÄ OLEN AIVAN	7	3		
23	AIVAN KUIN HÄN	7	3		
24	AIVAN KUIN JOKU	7	3		
25	ON AIVAN LIIAN	7	3		
26	JA AIVAN ERITYISESTI	7	3		
27	HÄN ON AIVAN	7	3		
28	OLIN AIVAN VARMA	7	3	olin aivan varma että (6)	
29	ON AIVAN ERILAINEN	6	3		
30	OLIN AIVAN VARMA ETTÄ	6	4	aivan varma että (28),olin aivan varma (7)	
31	OLLUT AIVAN VARMA	6	3		
32	KUIN HÄN OLISI	6	3		
33	AIVAN TOISELLA TAVALLA	6	3		
34	AIVAN ERI TAVALLA	6	3		
35	OLI AIVAN IIIISI	6	3		

concordance collocates plot patterns clusters filenames source text notes

2 708 Set hän ei osannut soittaa. Hän tunsu olonsa typeräksi. Typeräksi ja avuttomaksi. Mitä hän tsesi isän elämästä? Ehkä hänellä oli aivan uusi naisystävä, josta hän oli niin innoissaan, että unohti kutsua

»MINULLA ON AIVAN PALJON RAHAA»
‘I HAVE REALLY LOTS OF MONEY’
— PHRASEOLOGICAL UNITS IN THE TEACHING OF
FINNISH

In teaching a language as morphologically rich as Finnish, much attention is paid to the teaching of grammar. Despite the recent publication of numerous new textbooks and dictionaries, which give a far more up-to-date picture of the language, the teaching of vocabulary has still not been paid enough attention: words are generally taught without regard to the contextual information they contain, resulting in that the student does not have the ability to choose the word most appropriate to a given situation from the vocabulary at their disposal. In this article, vocabulary is examined as *phraseological units*. Vocabulary consists of numerous paradigmatic-syntagmatic relationships, the only one of which to have received significant analysis in language learning studies to date is that of collocation. Based upon corpus data, this analysis seeks to examine the other more abstract relationships that words express. These include clustering (n-grams), colligation, semantic preference and semantic prosody.

With regard to the teaching of vocabulary, the cotextual connections and synonymy of words present significant challenges for the process of language teaching. In this paper, these problems and their potential solutions are exemplified using the adverb *aivan* ‘really’ and the pair of synonyms *hievahtaa – liikahtaa* ‘budge’, ‘move’. Analysis of the corpus data reveals that the use of extensive material containing real examples of language use helps to provide information about the lexicon that is currently not presented in textbooks or dictionaries. The article posits that it is important that systematic information about the use of words be available to support students’ learning, and that this information could have a wide range of uses, notably in the production of new teaching materials. However, much research must yet be conducted and many challenges overcome before this phraseological approach can become an integral part of the teaching of vocabulary. ■

Kirjoittajan yhteystiedot (address):

Suomi toisena ja vieraana kielenä

PL 1000

90014 Oulun yliopisto

Sähköposti: *etunimi.sukunimi@oulu.fi*