

»TÄRKEÄ SEIKKA» JA »KESKEINEN KYSYMYKSI»: MITÄ KORPUSLINGVISTINEN ANALYYSI PALJASTAA LÄHISYNONYMEISTA?

JOHDANTO

Synonyymien hallintaa pidetään yhtenä tyylikkään ja vivahteikkaan kielenkäytön ominaispiirteenä: kielenkäyttäjää ohjataan valitsemaan usein toistuvan sanan tilalle jokin toinen ilmaus, jolla on sama merkitys. Tämä pyrkimys näkyy muun muassa taajaan ilmestyvinä synonyymisanakirjoina ja synonyymisanalistoina eri kielenoppaissa. Synonymia on ilmiö, joka kiinnostaa myös tavallista kielenkäyttäjää — ei pelkästään kielentuottamisprosessin vuoksi, vaan myös sanastollisen variaation tähden. Siitä kertovat lukuisat sanomalehdissä ja artikkelikokoelmissa ilmestyvät tekstit ja radion kielipakinat, joissa pohditaan vaikkapa sitä, millä alueella isoisää on kutsuttu *ukoksi*, *äijäksi* tai *vaariksi* (Vehmaskoski 1995).

Miten synonyymien käyttöä sitten ohjataan? Mitä sanakirjat, synonyymisanastot ja laajalle yleisölle tarkoitetut artikkelit kertovat synonyymien käytöstä? Yleiskielen sanakirjat, kuten Suomen kielen perussanakirja (CD-Perussanakirja 1997), antavat hakusanan lisäksi käyttöyhteydesimerkkejä, joiden tarkoituksena on selventää synonyymien käyttöä ja ehkä myös niiden merkityseroja. Varsinaisia ohjeita synonyymisten sanojen käytöstä annetaan Perussanakirjassa vain silloin tällöin — esimerkiksi isoisän synonyymi *ukko* mainitaan arkiseksi tai halventavaksi nimitykseksi, jota tulisi välttää yhteyksissä, joissa

tällaista merkitystä ei haluta tuoda julki. Sen sijaan synonyymisanastoista käyttöyhteys-esimerkit ja käyttöäännöt puuttuvat kokonaan — sanamääritelmät on annettu samaa tarkoittavien sanojen listana, kuten A. Leinon ja P. Leinon Synonyymisanastossa (1990). Synonyymisia sanoja käsittelevät artikkelit sen sijaan kuvaavat yleensä yksittäisten synonyymipesyeiden jäsenten levikkiä, syntyä ja joskus myös niiden käyttöä. Kaikista edellä mainituista puuttuvat kuitenkin selvitykset synonyymeiksi ilmoitettujen sanojen syntaktisesta, semanttisesta ja sanastollisesta ympäristöstä. Tällaista tietoa synonyymisten sanojen eroista kaipaisivat erityisesti ei-äidinkieliset kielenoppijat (Partington 1998: 29). Äidinkielenkin kielenkäyttäjä saattaa joutua ongelmatilanteen eteen silloin, kun hän joutuu selittämään, millaisissa (kieliopillisissa) rakenteissa jotakin synonyymia käytetään. Kielenkäyttäjän onkin intuitionsa perusteella vaikea ennustaa tarkasti, millaiset kombinaatiot ovat kielessä mahdollisia tai mahdottomia; intuition väitetään pikemminkin vahvistavan jonkin aineiston perusteella saatuja tuloksia (de Beaugrande 1999: 247–248). Kiel-tä havainnoiva kiinnittääkin helpommin huomiota poikkeuksiin ja erikoisuuksiin kuin kiel-tä tavallisiin ilmiöihin (Biber ym. 1998: 3).

Parin viime vuosikymmenen aikana kootut laajat tietokonemuotoiset tekstikorpuksset ovat osoittaneet hyötynsä merkityksen tutkimuksessa: sanojen tarkkoja merkityksiä ja merkityksiin sisältyviä leksikaalisia ja syntaktisia käyttöohjeita on pystytty selvittämään laajojen elektronisten tekstimateriaalien ja tietokonemuotoista tekstiä ymmärtävien ohjelmien avulla. Korpuslingvististä merkitysentutkimusta on harjoitettu viime päiviin asti erityisesti anglistien keskuudessa, mikä johtuu osittain tietysti siitä, että laajat korpuksset olivat pitkään juuri englannin kielestä koottuja (ks. mm. Biber ym. 1998; Kennedy 1998; Sinclair 1991). Tilanne on kuitenkin muuttunut: tietokonemuotoisia korpuksia on kerätty runsaasti jo muistakin kielistä, ja samalla myös merkityksen tutkimus on saanut lisää tutkimusaineistoa ja -tietoa.

Tässä artikkelissa tarkastelen synonymiaa ja korpuslingvistiikan keinoja, joilla synonyymisyyttä voidaan tutkia¹. Päähuomioni kohdistuu synonyymien *kotekstiin* (engl. *co-text*), jolla tarkoitan tarkasteltavan lekseemin välittömässä läheisyydessä olevaa tekstiyhteyttä². Tarkoitukseni on tutkia ohjeita ja rajoituksia, joiden mukaan synonyymipesyeiden jäsenet valikoituvat paradigmaattisesta merkityssuhteesta syntagmaattiseen suhteeseen. Analyysissäni en kuitenkaan tyydy tarkastelemaan aineiston tuottamia tuloksia sellaisinaan, vaan pyrin myös etsimään yhtäläisyyksiä ja selityksiä ilmiöille vertailemalla tuloksia keskenään. Havainnollistan synonymiaa ja sen tutkimista kahden synonyymisen adjektiivin — *tärkeän* ja *keskeisen* — avulla. Molemmat ovat etenkin akateemiselle tekstilajille tyypillisiä adjektiiveja ja merkitykseltään melko abstrakteja. Analyysi perustuu tietokonemuotoiseen korpukseen, joka on koottu Joensuun yliopistossa (tästä tarkemmin luvussa Esimerkkitapauksena *tärkeä* ja sen lähisyntonyymit). Tarkastelussa tarvittavat frek-

¹ Esitän kiitokseni professori Anna Mauraselle tätä artikkelia koskevista arvokkaista kommentista ja paranehdotuksista.

² Hakulinen ja Ojanen (1976) määrittelevät kotekstin lauseyhteydeksi. Korpuslingvistiikassa kotekstilla tarkoitetaan myös lause- tai virkerajan ylittävää tekstiympäristöä: Sinclair (1991: 172) määrittelee kotekstin hakusanan molemmilla puolilla esiintyviksi sanoiksi ilman lauseyhteyksirajoitusta. Kotekstin laajuutta on pyrkinyt määrittämään esimerkiksi Rébék-Nagy (1997), jonka mukaan koteksti on alue, joka kattaa 4–8 sanaa hakusanan molemmiin puoliin. Tämän mukaan koteksti voisi olla laajempi kuin Sinclairin (1991: 170) määrittelemä kollokaatiöväli (ks. alaviite 7).

venssisanalistat ja konkordanssit eli listat, jotka esittävät lekseemin koteksteineen, on tehty WordSmith Toolsin sisältämällä WordList- ja Concord-ohjelmilla.

LÄHISYNONYMISYYS

Termiä *synonymi* käytetään yleisesti viittaamassa sellaisiin muodoltaan erilaisiin sanoihin, joiden denotatiivinen merkitys on identtinen. Jotta synonymia kahden lekseemin välillä olisi täydellistä, synonymien olisi oltava vaihdettavissa keskenään kaikissa konteksteissa ilman muutosta denotaatiossa ja ilmauksen totuusarvossa sekä esimerkiksi tyyliarvossa tai affektiivisessä merkityksessä (Karlsson 1999: 219; Lyons 1977: 202). Tällainen kahden lekseemin vastaavuus on kielessä hyvin harvinaista, ja on epätodennäköistä, että kielessä säilyisi kahta täydellistä synonymia rinnakkain (Itkonen 1966: 365; Palmer 1976: 60). Semanttisessa analyysissä käytetäänkin usein termiä *lähisynonymi* erotuksena mahdollisista täydellisistä synonymeista (mm. Lyons 1981: 50).

Merkitykseltään samankaltaisista sanoista toiset ovat synonymisempia kuin toiset (Cruse 1986: 265; ks. myös Partington 1998: 29–32). Crusen (mts. 268) mukaan yhtenä ääripäänä ovat sanat, jotka ovat absoluuttisia synonymeja, ja toisessa päässä ovat sanat, jotka eivät ole synonymisia missään kontekstissa. Synonymisiksi lukeutuvat sanat, joiden keskeiset semanttiset piirteet ovat samat mutta joilla on myös sanat toisistaan erottavia lisämerkityksiä tai käyttötapoja. Erottavat piirteet voivat olla joko kielen sisäisiä tai kielen ulkoisia valintakriteereitä (de Jonge 1993: 525–526). Ensimmäisenä mainitut kriteerit ovat kielen sisäiseen kontekstiin eli kieleen, tekstilajiin ja esimerkiksi tyyliin liittyviä piirteitä. Synonymipeseyden sanoista osa onkin hyvin usein juuri tyyliväritteisiä, jolloin niitä voidaan käyttää vain tietyissä yhteyksissä. Hyvänä esimerkkinä tästä on pari *haaksi* – *laiva*, jonka jäsenistä *haaksi* on selvästi runollinen (Itkonen 1966: 354). Toisaalta osa lähisynonymieista voi kuulua vain tietyn tekstilajin ominaispiirteisiin: *ekvaattori* lienee enemmänkin eksaktissa tieteellisessä tekstissä käytetty, kun taas *päiväntasaaja* löytynee edellisistä useammin vaikkapa kaunokirjallisuudesta tai matkaesitteistä. Aiemmin mainitut kielen ulkoiset valintakriteerit ovat taas sidoksissa kielen ulkoiseen kontekstiin, kuten viestintätilanteeseen ja puhujan intention.

Seuraavassa tarkastelen erilaisia synonymien kontekstissa esiintyviä valintasääntöjä, jotka ohjaavat synonymien paradigmaattista valintaa. Käsittelen *kollokaatiota* ja *kolli-gaatiota*, joiden tutkimukseen korpuslingvistiikka on luonut oivat menetelmät. Korpus-tutkimus mahdollistaa hyvin myös hakusanan ja sen kontekstin *morfologisen analyysin*, joka saattaa olla yksi synonymeja erotteleva piirre useita taivutusmuotoja sisältävässä suomen kielessä. Hakusanalla tarkoitin jatkossa tutkimuksen kohteena olevaa lekseemiä, jota käytetään konkordanssiohjelmissa hakuterminä (vrt. englannin *node*).

KOLLOKAATIOT

Synonymiatutkimuksessa kollokaatioanalyysija on tehty runsaasti esimerkiksi englannista (ks. mm. Biber ym. 1998, Kennedy 1991), mutta suomen kielen osalta tutkimus on ollut vähäistä. Kollokaatioita on tarkasteltu jossain määrin kyllä Suomessakin; esimerkiksi Forsgren (mm. 1997) ja Westerlund (2000) ovat tarkastelleet niitä kääntämisen ja kään-

nöskielen kannalta. Synonyymeja tutkittaessa kollokaatiosuhteiden selvittäminen on oleellista, sillä kollokationaalista rajoitusta pidetään yhtenä synonyymeja toisistaan erottavana tekijänä (Cruse 1986: 279). Kollokaatio on määritelty yleensä joko assosiatiiviseksi, tekstuaaliseksi tai puhtaasti tilastolliseksi ilmiöksi. Assosiatiivinen määritelmä painottaa merkitykseltään yhteen liittyvien sanojen luontaista taipumusta esiintyä yhdessä ja siten muodostaa merkityspohjaisia kokonaisuuksia (Leech 1990: 17; ks. myös Karlsson 1999: 232). Esimerkiksi sanat *hevonon* ja *hinnua* liittyvät merkityksiltään yhteen, ja esiintyvät siten tekstissä toistensa kollokaatteina. Tämä tulkinta, joka liittyy kielenkäyttäjän kieli-tietoon merkityskokonaisuuksista, on lähellä käsitystä, jonka mukaan kielenkäyttäjä pystyy kompetenssinsa avulla muodostamaan kontekstin tai kielenulkoisen kontekstin laukaiseman käyttöyhteyteen sopivan tavallisen kollokaation (Partington 1998: 18). Kielenkäyttäjä voi myös yllätyksellisyyttä tavoitellessaan rikkoa tavallisia kollokaatorakenteita ja valita synonyymiparadigmasta sellaisen lekseemin, joka ei normaalisti muodosta kollokaatiota kontekstissa olevan sanan kanssa (mp.). Tekstuaalisen määritelmän mukaan kollokaatio muodostuu kahdesta tai useammasta sanasta, jotka esiintyvät tekstissä toistensa läheisyydessä (Sinclair 1991: 170). Mikä tahansa sana, joka esiintyy toisen sanan yhteydessä ennalta määrätyn välimatkan — kollokaatiovälin — sisällä, voi siis tämän määritelmän mukaan olla hakusanan kollokaatti. Tämä tarkoittaa myös sitä, että kollokaatioparin jäsenten ei tarvitse sijaita tekstissä vierekkäin (Kennedy 1998: 114; Sinclair 1991: 117). Edellisistä poiketen kollokaation määrittelyssä voidaan korostaa myös puhtaasti tilastollisen merkitsevyyden asemaa: »collocation has long been the name given to the relationship a lexical item has with items that appear with greater than random probability in its (textual) context» (Hoey 1991: 7–8). Määritelmä korostaa erityisesti kollokaation toistuvuutta: jotta kollokaatio olisi tilastollisesti merkitsevä, sen on toistuttava korpuksessa riittävän usein. Myöhempänä tässä artikkelissa olevassa esimerkkitapauksessa käsitellään kollokaatiota tilastollisen merkitsevyyden kannalta.

Vain harvat kollokaatiosuhteet ovat kiinteitä eli sellaisia, että ne säilyisivät teksti- ja käyttöyhteydestä toiseen täysin samanlaisina, ilman syntaktista tai leksikaalista variaatiota (Partington 1998: 25–26). Tällaiset sidotut kollokaatiot (Cruse 1986: 40–41) heijastavat hyvin usein idiomien ominaispiirteitä mutta ovat kuitenkin leksikaalisesti läpinäkyviä. Sidotuista kollokaatioista poiketen kollokaatioissa on yleensä kuitenkin variaation mahdollisuus. Vaihtelu on usein mahdollista vain rajatun joukon kesken. Esimerkiksi sopii hyvin substantiivi *toimenpide*, joka kollokoi ainakin suomalaisessa lakikielessä verbien *ryhtyä*, *suorittaa*, *toteuttaa* ja *tehdä* kanssa (Westerlund 2000: 404–405). Ääripäänä ovat sellaiset kollokaatiot, joiden valinta ei ole rajoitettu. Tällaisesta suhteesta pyrkii antamaan esimerkin Carter (1987: 63), jonka mukaan englannin verbi *run* »collocates relatively unrestrictedly with a range of – – entities» (korostus omani). Kuten lainauksesta näkyy, aivan täydellisestä vapaasta valinnasta ei kuitenkaan liene kyse; sellaista esiintyneekin lähinnä vain kieliopillisilla sanoilla, kuten konjunktiolla (esimerkiksi *ja*, *mutta*). Laajan variaation sallivat kollokaatiot selviävät vasta suuren tekstimateriaalin avulla. Sen sijaan kiinteät leksikaalistuneet kollokaatiot on yleensä helppo tunnistaa toisen tai toisten jäsenten ollessa tiedossa, vaikka kollokaation esiintymisfrekvenssi ei olisikaan taaja (ks. esim. Kennedy 1998: 121). Kollokaatiot vaihtelevat myös tekstilajeittain, tyyleittäin ja rekisterittäin. Kollokaatio, jota tavataan yhdessä tekstilajissa tai rekisterissä, saattaa olla toisessa harvinainen tai sitä ei esiinny siinä ollenkaan (Partington 1998: 17). Jo Firth (1957:

195) on jaotellut kollokaatioita »yleisiin» tai »tavallisiin» sekä tekstilaji- tai kirjoittajaspesifeihin kollokaatioihin.

KOLLIKAATIOT JA HAKUSANAN KIELIOPILLINEN TEHTÄVÄ

Leksikaalisen sidoksisuuden lisäksi useilla leksemeillä on myös taipumus esiintyä tietyn kielioppikategorian yhteydessä; näitä voivat olla esimerkiksi sanaluokka tai lauseenjäsen. Tällaisia hakusanan ympäristössä olevia kategorioita kutsutaan kolligaaiteiksi³ (ks. Firth 1968; Barnbrook 1996; Tognini Bonelli 1996). Hakusanojen kotekstin kieliopillisen rakenteen tarkastelu on osoittautunut hyväksi apukeinoksi lähisyronyymien erotte- lussa (Biber ym. 1998: 84). Jo kollokaatioanalyysin perusteella voi tehdä alustavia ha- vaintoja siitä, millaisissa rakenteissa hakusanat esiintyvät, mutta koska etenkin tilastolli- seen merkitsevyyteen perustuva kollokaatioanalyysi nostaa esiin vain osan hakusanan ym- päristössä olevista sanoista ja antaa siten vain viitteitä hakusanaa ympäröivistä kieliopil- lisistä kategorioista, voi hakusanan kieliopillisen ympäristön tarkastelu jäädä vaillinaiseksi. Kolligatioanalyysi täydentää siten merkittävästi kuvaa hakusanojen kotekstista. Tämän lisäksi myös itse hakusana on syytä ottaa tarkasteluun nimenomaan kieliopillisen tehtä- vänsä kannalta: analyysin avulla on mahdollista selvittää, toimivatko lähisyronyymit lau- seissa erilaisissa kieliopillisissa funktioissa. Kielioppikategorioihin liittyvät piirteet voi- vat siis löytyä kahtaalta: kolligaaiteista ja hakusanoista.

MORFOLOGISET OMINAISUUDET

Morfologinen analyysi ei ole kollokaatioanalyysin tavoin ollut kovin suosittua korpus- pohjaisessa synonymiatutkimuksessa. Joitakin tutkimuksia on kuitenkin tehty, esimerkiksi Kennedy (1991) on tarkastellut englannin prepositioiden *between* ja *through* ympäristös- sä olevien kollokaattien morfologista variaatiota ja havainnut morfologiset erot (esimer- kiksi kollokaattien yksikkö- ja monikkomuotojen erilaiset määrät) yhdeksi lähisyrony- ymeja erottavaksi tekijäksi. Voitaneen olettaa, että suomen kielen synonymiapeseiden tarkastelussa morfologiaa ei kannata sivuuttaa — onhan suomen morfologinen variaatio huomattavasti runsaampaa kuin esimerkiksi englannin. Tässäkään yhteydessä ei kuiten- kaan kannata pitäytyä pelkästään kollokaattien tarkastelussa, vaan aivan kuten edellä, myös morfologisia ominaisuuksia tutkittaessa on hyvä tutkia myös itse hakusanoja.

ESIMERKKITAPAUKSENA »TÄRKEÄ» JA SEN LÄHISYRONYIMIT

Tässä esitettävä analyysi perustuu Joensuun yliopiston kansainvälisen viestinnän laitok- sessa kootun elektronisen Käännösuomen korpuksen alun perin suomeksi kirjoitettujen

³ Hakusanan kotekstissa säännöllisesti esiintyvää kieliopillista kategoriaa on voitu kutsua myös kieliopilli- seksi (grammatical) kollokaatioksi ja tässä aiemmin esillä ollutta kollokaatiota leksikaaliseksi (lexical) kol- lokaatioksi (ks. esim. Carter 1987: 47–50).

tekstien osakorpukseen (korpuksesta tarkemmin Mauranen 1998, 2000)⁴. Tutkimuskorpuksen laajuus on noin 1 150 000 sanaa, jotka jakautuvat suunnilleen puoliksi akateemisiin teksteihin ja kaunokirjallisuuteen kuuluviksi⁵. Akateemisia tekstejä on yhteensä 12 ja kaunokirjallisia yhteensä 14 kappaletta; tekstit ovat kokonaisia teoksia, eivät tekstifragmentteja. Kukin tutkimusmateriaalin teksti on peräisin eri kirjoittajalta, ja tekstit on pyritty valitsemaan siten, että laajuutensa vuoksi yli- tai aliedustuvat tekstit on jätetty pois. Näin on pyritty takaamaan korpuksen tasapaino.

Olen valinnut tutkimusaineistostani tarkasteltavaksi yhden synonyymipesyeen, jonka jäsenille on yhteistä tekstissä kulloinkin käsiteltävänä olevan asian 'tärkeiden' ilmaiseminen. Pesyeen frekventin jäsen on *tärkeä*, joka on etenkin tietokirjallisuudessa yleinen adjektiivi: Oulun korpuksen (ks. Saukkonen ym. 1979) mukaan se on 15:nneksi yleisin adjektiivi koko korpuksessa ja tietokirjallisuudessa 10:nneksi yleisin. Sen synonyymit olen koonnut CD-Perussanakirjan (1997) ja Leinojen Synonyymisanaston (1990) sanaselityksistä. *Tärkeä*-adjektiivin synonyymeiksi on teoksissa mainittu seuraavat (suluissa on ilmoitettu, kuinka monta kertaa sana on mainittu synonyymeiksi ilmoitettujen sanojen sanaselityksissä): *huomattava* (13), *merkittävä* (9), *tähdellinen* (9), *arvokas* (7), *merkityksellinen* (7), *olennainen* (7), *keskeinen* (5), *oleellinen* (4) ja *kallisarvoinen* (3). *Tärkeä* esiintyi näiden sanojen sanaselityksissä 17 kertaa. Otin aluksi tutkittavaksi nämä kymmenen lähisyronymia ja laskin niiden frekvenssit tutkimuskorpuksessa. Taulukossa 1 näkyvät *tärkeän* ja edellä mainittujen lähisyronymien todelliset ja suhteelliset frekvenssit akateemisissa teksteissä ja kaunokirjallisuudessa.

Taulukko 1. *Tärkeä* ja sen lähisyronymit akateemisissa ja kaunokirjallisuudessa. Taulukossa on ilmoitettu todellinen frekvenssi (f) ja frekvenssi 100 000 sanaa kohti.

	Akateeminen kirjallisuus		Kaunokirjallisuus	
	f	f/100 000	f	f/100 000
<i>tärkeä</i>	602	102,6	160	28,4
<i>keskeinen</i>	559	95,3	26	4,6
<i>merkittävä</i>	166	28,3	30	5,3
<i>olennainen</i>	157	23,8	11	2,0
<i>arvokas</i>	54	9,2	23	4,1
<i>huomattava</i>	44	7,5	5	0,9
<i>oleellinen</i>	41	7,0	7	1,2
<i>merkityksellinen</i>	37	6,3	6	1,1
<i>tähdellinen</i>	4	0,7	7	1,2
<i>kallisarvoinen</i>	1	0,2	6	1,1

Taulukosta näkyy selkeästi, että akateemisissa tekstilajissa *tärkeä* ja *keskeinen* erottuvat omaksi parikseen ja muiden lukumäärät jäävät huomattavasti alhaisemmiksi. Kaunokirjallisuuden korpuksessa vain *tärkeää* voi pitää selvästi muuta yleisempänä. Jos *tärkeän* yleisintä lähisyronymia olisi päätelty Synonyymisanastossa ja CD-Perussanakir-

⁴ Käännösuomen korpus on niin sanottu verrannollinen korpus (Baker 1995: 234–235): se jakautuu kahteen osakorpukseen, joista toisen muodostavat suomennetut tekstit ja toisen alun perin suomeksi kirjoitetut tekstit.

⁵ Tässä yhteydessä tarkoitan kaunokirjallisuudella vain suorasanaista kertomakirjallisuutta; toisin on esimerkiksi Oulun korpuksessa, jossa kaunokirjallisuus kattaa mm. runot ja näytelmät (Saukkonen ym. 1979: 14).

jassa esiintyvien sana-artikkelien sanaselityksissä olevien lähisynonyymien määrän perusteella, olisi huomattava ollut kolmellatoista esiintymällään todennäköisimmin yleisimpänä pidettävä lähisynonyymi. Tässä korpuksessa huomattava on kuitenkin vasta 6:nneksi yleisin lähisynonyymi akateemisissa teksteissä ja harvinaisin kaunokirjallisuudessa. Sanaselityksistä ei siis voitane suoraan päätellä lähisynonyymien yleisyyttä teksteissä. Tutkimuskorpuksen mukaan tämän synonyymipesyeen jäsenet ovat lähes kokonaisuudessaan selvemmin akateemiselle tekstilajille ominaisia: *olennainen*, *huomattava*, *oleellinen* ja *merkityksellinen* ovat lähes yksinomaan siinä esiintyviä adjektiiveja. Myös *keskeinen*, *merkittävä* ja *arvokas* sekä frekventein *tärkeä* ovat yleisempiä akateemisissa teksteissä. *Tähdellinen* ja *kallisarvoinen* ovat sen sijaan kummassakin tekstilajissa hyvin harvinaisia. Tekstilajia on siis pidettävä — etenkin frekventeimpiä adjektiiveja tarkasteltaessa — ainakin osittain lähisynonyymeja erottavana tekijänä (lähisynonyymisten sanojen ja tekstilajin suhteesta ks. Biber ym. 1998: 24, 43–53). Seuraavassa keskityn kahteen taajimmin esiintyneeseen lähisynonyymiin: *tärkeään* ja *keskeiseen*. Olen valinnut tutkittavaksi tekstilajiksi pelkästään akateemisen kirjallisuuden, koska kaunokirjallisuudessa vain *tärkeän* frekvenssi olisi kyllin suuri analysoitavaksi.

»TÄRKEÄN» JA »KESKEISEN» KOLLOKAATIT

Tässä tutkimuksessa tarkastelen lähisynonyymien *tärkeä* ja *keskeinen* kollokaatteja tilastollisen merkitsevyyden avulla. Kollokaation tilastollista merkitsevyyttä laskettaessa kyseeseen tulevat yleensä joko t- tai z-testi tai yhteisinformaatio eli Mutual Information (MI) (Barnbrook 1996: 94–100). Käytän analyysissäni MI-testiä⁶, koska se soveltuu analyysiin, jossa havaintoyksikköjä tutkitaan pienehkön korpuksen avulla. Pienen korpuksen ongelmana ovat vähäisen lukumäärän saavat kollokaatiosuhteet. Tällaiset vähäfrekvenssiset kollokaatiot voivat pienestä esiintymätiheydestään huolimatta olla tilastollisesti merkitseviä, jos niiden jäsenet esiintyvät todennäköisemmin yhdessä kuin erikseen. Edellä mainituista testeistä erityisesti t-testi ilmoittaa merkitsevimmiksi sellaiset kollokaatiot, joiden jäsenten kokonaisfrekvenssit ovat korpuksessa suuria; samalla se ei pidä merkitsevinä sellaisia kollokaatiota, joiden jäsenten kokonaisfrekvenssit ovat pieniä. Myös z-testi osoittaa taipumusta nostaa merkitsevimmiksi suuriin frekvensseihin perustuvat tapaukset. Sen sijaan MI-testi ottaa huomioon myös korpuksessa harvoin esiintyvät sanat ja niiden muodostamat kollokaatiot eikä suosi pelkästään frekventtejä tapauksia. Koska (pieniä) korpusta analysoitaessa on varmistuttava siitä, että myös harvoin esiintyvien sanojen muodostamat kollokaatiot tulevat otetuksi yhtäläisesti huomioon, olen valinnut viimeksi mainitun testin tässä käytettäväksi. Havainnot testien erilaisuudesta perustuvat omiin testeihini sekä Barnbrookin (mp.) tuloksiin.

Analyyssissa on aluksi määritelty havaintoalue eli kollokaatioväli (engl. *span*). Havaintoalueeksi olen valinnut alueen, johon kuuluvat hakusanan molemminpuoliset vieruskollokaatit eli sanat, jotka esiintyvät hakusanasta joko seuraavana vasemmalla (taulukoissa ja myöhempänä tekstissä 1L) tai seuraavana oikealla (1R)⁷. Havaintoalueessa esiintyvät sanat on seuraavaksi lemmatisoitu, toisin sanoen kaikki sanan taivutusparadigmaan kuu-

⁶ Mutual Information -testin kaava on $(MI) = \log_2 O / E$, jossa O on havaintojen frekvenssi ja E on odotettu frekvenssi (Barnbrook 1996: 98–99).

luvut muodot on ryhmitelty yhdeksi lemmaksi eli taivutusmuodot sisältäväksi sananmuotojen ryhmäksi. Lemmatisointi on tehty manuaalisesti konkordanssilistoista. Tilastollista merkitsevyyttä laskettaessa on verrattu hakusanan ja lemman muodostamien kollokaatioiden todellisten havaintojen määrää odotusarvoon eli todennäköisyyteen, jolla sanat esiintyvät toistensa kollokaatteina korpuksessa. Laskennassa käytetään hyväksi tietoa kollokaattina olevan lemman kokonaismäärästä korpuksessa, joka saadaan selville Word-List-ohjelman laskemasta frekvenssisanalista (laskennasta yksityiskohtaisemmin mm. Barnbrook 1996). Tarkasteluun olen ottanut mukaan sanat, jotka esiintyvät hakusanan yhteydessä vähintään neljä kertaa ja vähintään kahdessa eri tekstissä. Näin olen pyrkinyt vähentämään yksittäisten tapausten määrää ja eliminoimaan yhden kielenkäyttäjän idiolektin vaikutuksen tutkimustuloksiin.

Taulukossa 2 esitetään lähisynonyymien 1L-kollokaatit MI-arvon mukaiseen järjestykseen järjestettyinä. Olen lihavoanut taulukossa ne merkitsevät kollokaatit, joita ei toisella hakusanalla esiinny.

Taulukko 2. *Tärkeän ja keskeisen* 1L-kollokaatit merkitsevyyden (MI) mukaisessa järjestyksessä. Mukana kollokaatit, jotka esiintyvät korpuksessa vähintään neljä kertaa ja kahdessa eri tekstissä. f = kollokaatin frekvenssi 1L-asemassa. Viiva (–) osoittaa rajakohtaa, jonka alla olevat kollokaatit eivät ole tilastollisesti merkitseviä (raja-arvo on MI-testissä 1,58).

kollokaatti	tärkeä		keskeinen		
	f	MI-arvo	kollokaatti	f	MI-arvo
vähemmän	10	6,46	yksi	37	5,30
erityisen	6	5,62	eräs	4	4,20
yhtä	13	5,58	kaksi	9	4,04
paljon	4	5,39	kannalta	6	3,93
kannalta	15	5,15	tulla	13	3,28
eräs	4	4,10	politiikka	4	3,10
hyvin	4	4,03	teoria	5	2,90
edelleen	4	3,86	tutkimus	7	2,84
tulla	15	3,38	moni	5	2,71
syy	4	2,68	toinen	7	2,27
moni	5	2,60	olla	89	1,68
pitää	8	2,55	–		
yhteiskunta	4	2,42	joka	9	0,25
yksi	5	2,30	se	10	-0,09
olla	138	2,20			
kuitenkin	7	2,12			
jokin	4	1,66			
–					
myös	9	1,24			
hän	5	0,81			
tämä	7	0,56			
joka	7	-0,22			
se	9	-0,35			
ja	8	-1,53			

⁷ Sinclair (1991: 170; ks. myös Barnbrook 1996: 91) on määritellyt sopivaksi kollokaatiöväliksi viisi sanaa hakusanan kummallakin puolella (5L–5R). Kollokaatiotutkimuksessa on kuitenkin saatu merkittäviä tuloksia (esim. Biber ym. 1998; Bäcklund 1973; Kjellmer 1987) tutkimalla pelkästään vieruskollokaatteja, ja tässä esityksessä tydyn tarkastelemaan pelkästään niitä.

Tärkeän omista 1L-kollokaateista erottuu selkeästi yksi ryhmä: kvantiteettia ilmoittavat astemääritteet (ad-adjektiivit), joita ovat *vähemmän*, *erityisen*, *yhtä*, *paljon*⁸ ja *hyvin* (astemääritteistä ja ad-adjektiiveista ks. Hakulinen ja Karlsson 1988: 83–84; Vilkuna 1996: 44, 207–209). Näistä neljä ensimmäistä ovat *tärkeän* merkitsevimpiä 1L-kollokaatteja. Tähän ryhmään kuuluvia kollokaatteja ei *keskeisen* tilastollisesti merkitsevistä 1L-kollokaateista löydy ainuttakaan. (Esimerkit 1–3.)

1. Tämän seurauksena kristinuskon rooli muuttui *vähemmän tärkeäksi*, vaikka se edelleen toimi sekä normatiivisena rauhoittajana – –
2. Samastumisen ulottuvuus ei ylipäänsä ole *yhtä tärkeä* kuin Suomessa.
3. Ajallisen jatkuvuuden (samuuden) ohella myös erottautuminen on *hyvin tärkeää*.

Tärkeällä on edellisten lisäksi kaksi muuta adverbia (*edelleen* ja *kuitenkin*), kaksi substantiivia (*syy* ja *yhteiskunta*), yksi verbi (*pitää*) ja yksi pronomini (*jokin*) omina kollokaatteinaan. (Tässä tarkastelussa tarkoitan kursivilla ilmaistulla kollokaatilla ja hakusannalla lemmää, en pelkästään yhtä sananmuotoa.) *Keskeisen* omat 1L-kollokaatit ovat substantiivit *politiikka*, *teoria* ja *tutkimus* sekä lukusanakvanttorit *kaksi* ja *toinen*. Lisäksi on huomattava, että sen tilastollisesti merkitsevin kollokaatti on *yksi*, jolla on paljon suurempi frekvenssi *keskeisen* kollokaattina (37) kuin *tärkeän* (5) ja jonka MI-arvo on myös selvästi suurempi (5,30 vs. 2,30). Esimerkkejä *keskeisen* 1L-kollokaateista:

4. Näiden *kahden keskeisen* jännitteen strukturoima feminististen tietoppien kolmijako on Sandra Hardingin käsittein – –
5. Oppikirjat tarjoavat tähän monimutkaiseen tarinaan *yhden keskeisen* kertomuksen.
6. *Tutkimuksen keskeinen* idea maallikoiden ja asiantuntijoiden riskin määritelmien eroista jätetään mainitsematta – –

Astemääritteiden puuttumisen lisäksi *keskeisen* erottaa *tärkeästä* juuri sen 1L-asemassa olevat lukusanakvanttorit; lisäksi sillä on kollokaatteinaan pronominaaliset kvanttorit *eräs* ja *moni*, jotka myös ilmoittavat (epämääräistä) lukumäärää. *Keskeisellä* on merkitsevien kollokaattien joukossa kvanttoireita kaiken kaikkiaan selvästi enemmän kuin *tärkeällä*: *keskeisellä* niitä on 62; *tärkeällä* puolestaan vain 14. Kolmas erottava tekijä on substantiivien määrä 1L-kollokaattien joukossa: *keskeisellä* niitä on kolme yhdentoista merkitsevän kollokaatin joukossa; *tärkeällä* kaksi seitsemäntoista joukossa.

Hakusanoilla on myös yhdistäviä piirteitä. Edellä mainittujen *yksi*, *eräs* ja *moni* -sanojen lisäksi yhteisiä ovat *kannalta*, *tulla* ja *olla*. *Olla*-verbi on näistä mielenkiintoinen. Se on molemmilla adjektiiveilla frekventein 1L-kollokaatti, mutta ei suuren esiintymätiheydensä vuoksi saa kuitenkaan suurta MI-arvoa. Vaikka *olla* on näiden adjektiivien suhteen tilastollisesti heikosti merkitsevä, on se kuitenkin tärkeä kielellinen ilmiö: onhan

⁸ Luen tässä astemääritteiksi myös toisissa koteksteissa paljouskvanttorina toimivat *paljon*, *hiukan* ja *vähän* (*hiukan* ja *vähän* eivät näy kollokaatioanalyysissä esiintymien vähäisen määrän vuoksi mutta ovat mukana kolligaaatteja laskettaessa). Niillä on samanlainen syntaktinen asema kuin astemääritteillä, sillä ne voivat määrittää adjektiivia (*paljon tärkeämpi*), kvanttoria (*hiukan enemmän*) ja adverbja (*vähän kauemmaksi*). Hakulisen ja Karlssonin (1988) ad-adjektiiveiksi lukemiin määritteisiin verrattuna näiden erona on se, että nämä määrittävät adjektiivien, adverbien ja kvanttoireiden komparoituja muotoja, kun esimerkiksi *hyvin* ja *varsin* määrittävät positiivimuotoja.

näiden adjektiivien 1L + *tärkeälkeskeinen* -syntagmoista tavallisin juuri sellainen, jossa *olla* on 1L-asemassa (esimerkit 7 ja 8).

7. Mielikuvat ovat *tärkeitä*, mutta eivät merkitse koko asiaa.
8. Kuten myöhemmin käy ilmi, varsinaisen metaforan käytön ymmärtämiseksi Väyrysen kampanja on *keskeinen* osa laajempaa kontekstia.

Olla on myös tärkeä oikeanpuoleinen vieruskollokaatti, vaikka se ei kummallakaan hakusanalla yllä tilastollista merkitsevyyttä ilmoittavan rajan yläpuolelle (ks. taulukko 3).

Seuraavassa taulukossa on esitetty *tärkeää* ja *keskeistä* välittömästi seuraavat kollokaatit. Aineiston perusteella molemmilla adjektiiveilla on 1R-asemassa enemmän tilastollisesti merkitseviä kollokaatteja kuin 1L-asemassa.

Taulukko 3. *Tärkeän* ja *keskeisen* 1R-kollokaatit merkitsevyyden (MI) mukaisessa järjestyksessä. Mukana kollokaatit, jotka esiintyvät korpuksessa vähintään neljä kertaa ja kahdessa eri tekstissä. f = kollokaatin frekvenssi 1R-asemassa. Viiva (–) osoittaa rajakohtaa, jonka alla olevat kollokaatit eivät ole tilastollisesti merkitseviä (raja-arvo on MI-testissä 1,58).

tärkeä			keskeinen		
kollokaatti	f	MI	kollokaatti	f	MI
tekijä	10	5,03	teesi	4	6,51
seikka	4	4,87	elementti	10	6,41
rooli	11	4,76	teema	11	6,00
piirre	7	4,72	idea	4	5,50
tavalla	11	4,45	osuus	4	5,50
yhteiskunnallinen	7	4,04	piirre	9	5,19
osa	16	3,91	asema	17	5,13
erityisesti	5	3,79	sija	14	4,83
suhde	5	3,68	tehtävä	9	4,82
tehtävä	4	3,54	käsite	21	4,74
kysymys	12	3,23	kysymys	30	4,66
kuin	24	3,09	rooli	9	4,58
ero	7	3,03	tavoite	5	4,58
asema	4	2,93	nousta	6	4,42
koska	5	2,72	tekijä	6	4,40
merkitys	7	2,71	lähtökohta	5	4,25
syy	4	2,68	tavalla	6	3,68
asia	9	2,41	ongelma	10	3,66
poliittinen	4	2,40	paikka	4	3,42
sosiaalinen	4	2,36	kohde	6	3,40
tieto	4	2,30	syy	6	3,37
käsite	4	2,24	ajatus	5	3,25
–			osa	9	3,19
mutta	4	1,06	merkitys	7	2,82
että	10	0,56	–		
olla	36	0,26	kuin	4	0,61
ei	4	-0,93	myös	4	0,18
ja	12	-0,95	olla	13	-1,10
se	4	-1,52	ja	6	-1,84

Kummallakin adjektiivilla noin puolet tilastollisesti merkitsevistä 1R-kollokaateista on omia ja noin puolet yhteisiä kollokaatteja. *Tärkeän* omia kollokaatteja ovat substantiivit *seikka*, *suhde*, *ero*, *asia* ja *tieto*; adjektiivit *yhteiskunnallinen*, *poliittinen* ja *sosiaalinen*; konjunktiot *kuin* ja *koska* sekä adverbi *erityisesti* (esimerkit 9–11). Osa *kuin*-konjunktion lukumäärästä selittyy sillä, että *kuin* esiintyy syntagmassa, jossa hakusanan vasemmalla puolella on ekvatiivinen astemääräite, kuten esimerkissä 11.

9. Semioottisina olioina tekstejä yhdistää muuan tuiki *tärkeä seikka* — nimitäin se, että ne kaikki ovat representaatioita.
10. -- eri puolilla Lappia havaitut yksittäiset metsävauriot kasvoivat *tärkeäksi yhteiskunnalliseksi* ilmiöksi nimenomaan siksi, että --
11. Jos joku näistä seikoista olisi kulttuurin kannalta yhtä *tärkeä kuin* meille sattuu olemaan homous, juoppous tai aggressiivisuus --

Keskeisen omat 1R-kollokaatit ovat lähes yksinomaan substantiiveja: *teesi*, *elementti*, *teema*, *idea*, *osuus*, *sija*, *tavoite*, *lähtökohta*, *ongelma*, *paikka*, *kohde* ja *ajatus* (esimerkit 12–14). Joukkoon mahtuu vain yksi muuta sanaluokkaa edustava sana: verbi *nousta* (esimerkki 15). *Tärkeällä* ei ole 1R-kollokaattiansa joukossa ainoatakaan tilastollisesti merkitsevää verbiä.

12. Kirjoittajien *keskeinen teesi* on, että sosiaaliset ja pragmaattiset tekijät vaikuttavat luokittelujärjestelmiin --
13. Myös Lehtosen romaanissa paloviinan salapoltto ja salakauppa olivat mitä *keskeisin teema*.
14. Yksi puheaktiteorian *keskeinen ongelma* on epäsuoruus, jossa Searlen (1975, 60) mukaan puhuja esittää --
15. -- esimerkiksi sellainen, jossa informaation erittelyn asemesta *keskeiseksi nousee* merkityksen tulkinta.

Adjektiivien edellä käsitellyt omat kollokaatit noudattavat johdonmukaisesti kaikkien merkitsevien kollokaattien jakaumaa: *tärkeällä* on kollokaatteinaan 15 substantiivia, kolme adjektiivia, kaksi adverbia ja kaksi konjunktiota; *keskeisellä* 22 substantiivia, yksi adverbi ja yksi verbi. *Keskeisen* 1R-kollokaatit muodostuvat siis lähes yksinomaan substantiiveista, mutta *tärkeällä* kollokaattien sanaluokissa on enemmän hajontaa.

Adjektiiveilla on omien kollokaattiansa lisäksi 11 yhteistä 1R-kollokaattia, mikä merkitsee, että ne esiintyvät myös samanlaisissa syntagmaattisissa suhteissa. Näistä 10 on substantiiveja ja yksi adverbi. Vaikka molempien 1R-asetusta löytyvät substantiivit *asema*, *kysymys* ja *käsite*, ovat ne kuitenkin tavallisempia *keskeisen* kollokaatteina: *keskeisen* kollokaattina *asema* on 17 kertaa, *tärkeällä* vain neljä kertaa (MI-arvot ovat 5,13 ja 2,93); *kysymys* on peräti 30 kertaa *keskeisen* kollokaattina ja 12 kertaa *tärkeällä* (MI-arvot 4,66 ja 3,23). *Käsite* saa *tärkeän* kollokaateista pienimmän merkitsevyyssarvon (2,24 ja neljä esiintymää), kun taas *keskeisen* kollokaattina se on verrattain merkitsevä (MI-arvo 4,74, esiintymiä 21). Yhteisistäkin kollokaateista löytyy siis eroja, ja niiden frekvenssierot ja etenkin merkitsevyyssarvojen erot voidaan tulkita lähisyronyymeja erotteleviksi tekijöiksi.

Kaiken kaikkiaan näiden kahden lähisyronyymien välillä näyttää siis aineiston perusteella olevan kollokationaalista eroa: *tärkeällä* ja *keskeisellä* on yhteisten kollokaattien lisäksi runsaasti sellaisia kollokaatteja, joita toisella ei ole. Kollokaatioanalyysi antaa vielä

mahdollisuuden tarkastella tilastollisesti merkitsevien kollokaattien kokonaismäärää suhteessa kaikkiin kollokaatteihin ja kollokaatiosuhteiden kiinteyttä. Taulukossa 4 esitetään sekä vasemman- että oikeanpuoleisten vieruskollokaattien yhteenlasketut lukumäärät ja suhteelliset osuudet taulukkojen 2 ja 3 perusteella. *Tärkeällä* on yhteensä 17 erilaista tilastollisesti merkitsevää 1L-kollokaattia, jotka esiintyvät aineistossa vähintään neljä kertaa ja kahdessa eri tekstissä; niiden yhteenlaskettu määrä on 250. Nämä kollokaatit muodostavat yhteensä 41,5 % kaikista *tärkeän* saamista kollokaateista, joita on yhteensä 602 eli siis saman verran kuin itse hakusanan esiintymiäkin. *Keskeisellä* vastaavat luvut ovat selvästi pienemmät. *Tärkeän* merkitsevimmän 1L-kollokaatin (*vähemmän*) MI-arvo on myös suurempi kuin *keskeisen* merkitsevimmän 1L-kollokaatin (*yksi*) — itse asiassa kaikkiaan neljä *tärkeän* kollokaattia ylittää *keskeisen* tilastollisesti merkitsevimmän kollokaatin MI-arvon, joka on 5.30 (ks. taulukkoa 2). Korpuksen perusteella *tärkeällä* on suhteellisesti enemmän merkitseviä 1L-kollokaatteja ja sen kollokaatiosuhteet 1L-asemassa ovat *keskeistä* vahvemmat, mikä näkyy suurempina merkitsevyysarvoina. 1R-kollokaatteja tarkasteltaessa tilanne on sen sijaan päinvastainen, kuten taulukosta 4 voi havaita: nyt *keskeisen* tilastollisesti merkitsevät kollokaatit muodostavat selvästi suurimman osuuden kaikista hakusanan kollokaateista (38,8 %), ja sen erilaisten kollokaattien lukumäärä on myös hieman suurempi (24). Lisäksi merkitsevimmän 1R-kollokaatin MI-arvo on *keskeisellä* suurempi, ja peräti seitsemän kollokaatin merkitsevyysarvo ylittää *tärkeän* merkitsevimmän kollokaatin arvon 5,03 (ks. taulukkoa 3). Korpuksen mukaan oikeanpuoleiset kollokaatit ovat siis *keskeisellä* merkitsevempiä, ja niitä on koko kollokaattimäärästä enemmän. *Keskeinen* kollokoi siten 1R-asemassa olevien kollokaattiensä kanssa vahvemmin kuin *tärkeä*.

Taulukko 4. Yhteenvedo *tärkeän* ja *keskeisen* kollokaatiosuhteista taulukkojen 2 ja 3 mukaan. f = erilaisten tilastollisesti merkitsevien kollokaattien määrä. F = tilastollisesti merkitsevien kollokaattien yhteenlaskettu määrä, % = tilastollisesti merkitsevien kollokaattien prosenttiosuus kaikista hakusanan saamista kollokaateista. Suurin MI-arvo ilmoittaa tilastollisesti merkitsevimmän kollokaatin MI-lukeman.

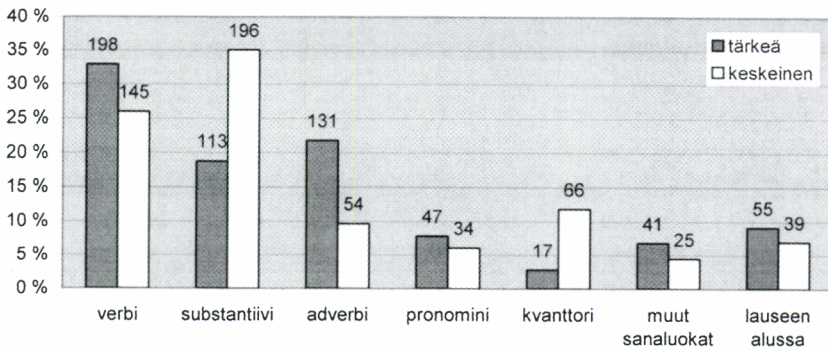
	1L-kollokaatit			1R-kollokaatit		
	f	F / %	suurin MI-arvo	f	F / %	suurin MI-arvo
tärkeä	17	250 / 41,5	6,46	22	168 / 27,9	5,03
keskeinen	11	186 / 33,3	5,30	24	217 / 38,8	6,51

Kollokaatioanalyysissä lähisyronyymeja yhdistäviksi tekijöiksi osoittautuivat erityisesti 1L-position yleisin kollokaatti *olla* ja 1R-position yhteiset substantiivit. Erottaviksi tekijöiksi ilmenivät 1L-position astemääritteet ja kvanttorit ja *keskeisen* taipumus saada 1R-positiiossa kollokaateikseen lähes pelkästään substantiiveja — *tärkeän* kollokaattikirjo oli selvästi laajempi. Lähisyronyymit eroavat toisistaan myös tilastollisesti merkitsevien kollokaattien määrän ja suhteellisen osuuden sekä kollokaatioiden kiinteyden suhteen.

▷

Edellä olevasta kollokaatioanalyysistä voi alustavasti päätellä, millaisissa syntagmaattisissa suhteissa *tärkeä* ja *keskeinen* ovat kollokaattiensa kanssa: vasemmalla puolella esiintyy adverbija, kvanttoireita tai verbejä, oikealla usein substantiiveja. Havainto on looginen suhteessa adjektiivien yleiseen käyttötapaan: ne voivat esiintyä attribuutiasemassa, jolloin substantiivit asemassa 1R ovat helposti selitettävissä; predikaattiasemainen adjektiivien käyttö taas selittää osaltaan verbien määrän 1L-positiossa. Adverbeihin lukeutuvat astemääritteet voivat esiintyä adjektiivin määritteenä, olipa adjektiivi attribuuttina (*He ostivat hyvin kauniin talon*) tai predikaatiivina (*Tyttö on hyvin kaunis*). Kvanttorit esiintyvät yleensä attribuutiasemaisen adjektiivin kanssa (*He omistavat kaksi pientä taloa*).

Olen seuraavassa koonnut molempien adjektiivien kaikkien kollokaattien sanaluokat kahteen kaavioon. Kuviossa 1 esitetään sanaluokat, jotka edeltävät *tärkeää* ja *keskeistä* 1L-positiossa. Kolligaatioanalyysissa lauseenalkuiset ja -loppuiset tapaukset on laskettu omaksi ryhmäkseen sanaluokkien rinnalle. Siten edellä olleen esimerkin 3 kolligaatioanalyysi tuottaa tuloksen, jossa 1L on adverbi (*hyvin*) ja 1R on laskettu lauseenloppuiseksi tapaukseksi. Tavoitteena on tällä tavalla tarkastella, eroavatko adjektiivit toisistaan lauseasemaltaan. Suurimmat kolligaattien väliset erot 1L-asemassa ovat substantiivien, adverbien, kvanttorien ja verbien määrissä: substantiivien ero on 16, adverbien 12, kvanttoireiden 9 ja verbien 7 prosenttiyksikköä.



Kuvio 1. *Tärkeän* ja *keskeisen* sanaluokkakolligaatit 1L-asemassa. Prosenttiluku ilmaisee sanaluokan osuuden kaikista kollokaattien sanaluokista, luku pylvään päällä kertoo sanaluokkaesiintymien todellisen lukumäärän.

Keskeinen esiintyy useimmiten syntagmoissa, joissa 1L-asemassa on substantiivi:

16. Tervejärkisyys, naiivi realismi, kuten älymystö arvioisi, oli Suomen *pe-lastumisen keskeisiä* kulmakiviä.
17. *Poliitikkojen keskeinen* vaatimus oli Suomen osallistuminen Kuolan saastepäästöjen vähentämiseen.

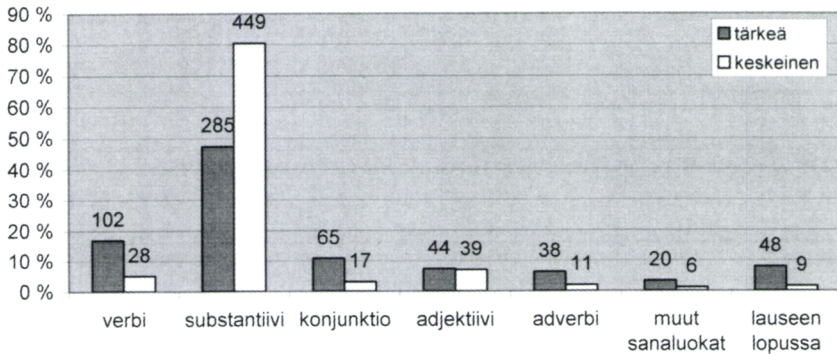
Myös *tärkeää* käytetään samanlaisessa syntagmaattisessa suhteessa mutta huomattavasti harvemmin. Edellä oli kollokaatioiden käsittelyn yhteydessä havaittavissa, että *keskeisen* tilastollisesti merkitsevistä kollokaateista moni oli kvanttori, mutta *tärkeällä* kvanttoireita ei 1L-positiossa juuri ollut — tämä näkyy kolligaatioanalyysissä eroina kvantto-

rien suhteellisissa osuuksissa. Samassa positiossa on nähtävissä myös adverbien suhteellisten osuuksien ero. Ero johtuu astemääritteiden määrästä: *tärkeällä* on vasemmanpuoleisina vieruskolligaatteinaan astemääritteitä 69 kappaletta, kun *keskeisellä* on niitä vain 18 (ks. myös lukua Hakusanojen *tärkeä* ja *keskeinen* morfologia ja syntaktinen funktio, s. 185–188). *Tärkeä*-adjektiivilla on myös enemmän verbejä tässä positiossa, ja sen yleisin sanaluokkakolligaatti onkin verbi (33 %), kun se *keskeisellä* on substantiivi. Pronominien ja muiden sanaluokkien (adjektiivit, konjunktiot, post- ja prepositiot) eroissa, samoin kuin lauseenalkuisten esiintymien määrässä, ei ole kovin suurta suhteellista eroa.

Tarkasteltaessa positiota 1R (kuvio 2) yksi erottava tekijä nousee selvästi esiin: *keskeisellä* on tässä positiossa huomattavasti enemmän substantiiveja kuin *tärkeällä* — ero on peräti 33 prosenttiyksikköä. Substantiivi onkin *keskeisen* tavallisin 1R-sanaluokkakolligaatti (80 %); muiden osuudet jäävät hyvin pieniksi. Substantiivi on myös *tärkeän* tavallisin kolligaatti, mutta sen osuus on selvästi pienempi (47 %). Pieniä eroja on myös verbien ja konjunktioiden määrissä: *tärkeällä* näitä on 1R-positiossa enemmän kuin *keskeisellä*. Konjunktioiden määrä selittyy osittain *kuin*-konstruktiolla, jossa *tärkeä* on komparoitui (esimerkki 18) tai sillä on positiivimuotoisena määrite *yhtä* (19; vrt. aiemmin *kuin* kollokaattina). Myös lauseenloppuinen positio on *tärkeälle* tavallisempi kuin *keskeiselle*.

18. Se, että oltiin naisia, katsottiin *tärkeämmäksi kuin* naisten väliset erot.

19. Siellä kuninkaanvaltaa rajoitettiin jo 1600-luvulla, eikä hovin merkitys voinut siten myöskään tulla yhtä *tärkeäksi kuin* Ranskassa.



Kuvio 2. *Tärkeän* ja *keskeisen* sanaluokkakolligaatit 1R-asemassa. Prosenttiluku ilmaisee sanaluokan osuuden kaikista kollokaattien sanaluokista, luku pylvään päällä kertoo sanaluokkaesiintymien todellisen lukumäärän.

Kolligaattianalyysi osoittaa selvästi, että tutkittavia lähisynonyymeja käytetään erilaisissa syntagmaissa. Erottavia tekijöitä löytyy sekä 1L- että 1R-asemien sanaluokkakolligaattien jakaumista. Seuraavassa luvussa tarkastelen, osoittautuuko kollokaattien morfologia merkittäväksi tekijäksi näiden lähisynonyymien analyysissa.

KOLLOKAATTIEN MORFOLOGINEN JAKAUMA

Kollokaattien morfologiasta otan esiin substantiivien sijataivutuksen. Substantiivien määrä on sekä positiossa 1L että 1R niin suuri, että analyysi on mahdollista. Myös verbien mää-

▷

rä, etenkin positiossa 1L, sallii analyysin. Verbien morfologiassa ei kuitenkaan osoittaudu olevan eroja: verbien kollokointi ei poikkea pääluokan, moduksen tai tempuksen suhteen merkittävästi toisistaan. Tässä korpuksessa kollokoivien verbien morfologia ei siis toimi näitä lähisynonyymeja erottelevana tekijänä. Substantiivien sijataivutuksessa sen sijaan on eroja, joskin niitä esiintyy pelkästään vasemmanpuoleisissa vieruskollokaateissa; oikeanpuoleisien kollokaattien taipuminen on kummallakin adjektiivilla samantapaista (sijojen osuuk-sien erot jäävät alle neljän prosenttiyksikön). 1L-positiossa olevien substantiivien sijataivutus on erilaista genetiivissä, partitiivissa ja elatiivissa, kuten seuraavasta taulukosta voi havaita. Muiden sijojen kohdalla ei ole merkitsevää eroa — sijojen osuuk-sien erot jäävät alle viiden prosenttiyksikön.

Taulukko 5. *Tärkeän ja keskeisen* 1L-substantiivikollokaattien sijataivutuksen määrälliset ja suhteelliset erot genetiivissä, partitiivissa ja elatiivissa. Prosenttiluku ilmoittaa sijamuodon osuuden kaikista sijamuodoista.

	tärkeä		keskeinen	
	f	%	f	%
genetiivi	52	46,0	137	69,9
partitiivi	16	14,2	4	2,0
elatiivi	16	14,2	9	4,6

Keskeinen esiintyy usein rakenteessa substantiivi(gen.) + *keskeinen*, jonka osuus on liki 70 %. Tämä rakenne on *keskeisellä* selvästi yleisempi kuin *tärkeällä*: ero on lähes 24 prosenttiyksikköä. Kyseisellä rakenteella ilmaistaan virkkeissä, minkä asian (teema, konkreettinen) keskeinen ilmiö (reema, usein abstrakti) on kyseessä; tätä havainnollistavat esimerkit 20–22. Myös *tärkeä* esiintyy samanlaisissa rakenteissa (esimerkit 23 ja 24), mutta niiden suhteellinen osuus on selvästi pienempi.

20. Yksi *tämän oppikirjan keskeinen* tavoite onkin osoittaa se, mitä toimintojen koordinoimisella tarkoitetaan.
21. *Poliitikkojen keskeinen* vaatimus oli Suomen osallistuminen Kuolan saastepäästöjen vähentämiseen.
22. Yleisön merkityksen korostaminen on *Perelmanin retoriikkanäkemyksen keskeinen* teema.
23. *Oluen tärkein* valintakriteeri on tällöin jotain muuta kuin alkoholiprosentti.
24. – – että sosiaalidemokraatit nähtiin *kommunistien tärkeimpänä* vastustajana, päävihollisena – –

Tärkeä-adjektiivin substantiivikollokaattien partitiivin yleisyys selittyy rakenteilla, joissa ilmoitetaan, mitä pidetään tärkeänä (25) tai mitä tärkeämpi jokin toinen asia on (26); tällaiset rakenteet ovat tässä korpuksessa *keskeisellä* harvinaisia. Harvinaisia ovat myös rakenteet, joissa *keskeistä* edeltää elatiivimuotoinen substantiivi. *Tärkeällä* näitä rakenteita esiintyy suhteellisesti useammin (ks. taulukkoa 5). Elatiivia käytetään useasti rakenteissa, joissa ilmaistaan kenen mielestä (27) tai mistä syystä (28) jokin asia on tärkeä. Ela-

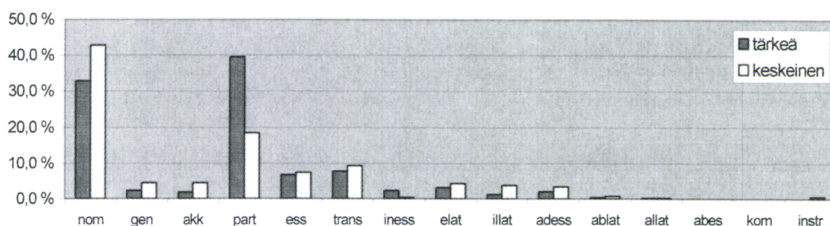
tiivi esiintyy myös superlatiivimuotoisen hakusanan kanssa, kun ilmoitetaan, mistä joukosta jokin on tärkein (29).

25. – – Juliet Mitchell pitikin *freudilaista psykoanalyysia tärkeänä* ja hyödyllisenä feministiselle ajattelulle.
26. *Todennäköisyyksien arviointia tärkeämmäksi* tulee päätösten seurausten vakavuuden – –
27. Suhtautuminen kieltolakiin oli vuosisadan alkupuolella *mielestäni keskeinen* alkoholipoliittisen mielipideilmaston mittari.
28. Kun suomi nyt toimi Karjalassa kansallisena kielenä — mikä oli *poliittisista syistä tärkeää* — — —
29. Turvallisuus on *elämän arvoista tärkein*, ja sen piiriin kuuluvat ennustettavuus – –

Tärkeän ja *keskeisen* vieruskollokaattien morfologiaa tarkasteltaessa erottelevaksi tekijäksi nousee 1L-position substantiivien sijataivutuksen erilaisuus. Synonyymisyyttä korostavat sen sijaan oikeanpuoleisten substantiivikollokaattien ja kaikkien verbikollokaattien morfologian samanlaisuus, mitä ei kuitenkaan tässä yhteydessä ole tarkasteltu lähemmin.

HAKUSANOJEN »TÄRKEÄ» JA »KESKEINEN» MORFOLOGIA JA SYNTAKTINEN FUNKTIO

Tässä jaksossa siirrän huomion itse hakusanoihin. Tarkastelen, esiintyykö hakusanojen morfologiassa ja syntaktisessa funktiossa sellaista vaihtelua, jota voidaan pitää näitä adjektiveja erottelevana tekijänä. Edellä tuli ilmi, että substantiivien tiettyjen sijojen määrälliset osuudet ovat erilaisia 1L-positiossa. Onko sitten mahdollista, että itse hakusanojen sijat käyttäytyvät eri tavalla? Korpuksen perusteella eroja ilmenee kahden sijan määrissä: erilaisia ovat nominatiivin ja partitiivin osuudet. Muiden sijojen erot ovat marginaalisia (sijojen osuuksien erot jäävät alle kolmen prosenttiyksikön). Tilannetta havainnollistaa kuvio 3.



Kuvio 3. Hakusanojen *tärkeä* ja *keskeinen* sijataivutusmuotojen suhteelliset osuudet.

Kuviosta käy ilmi, että *keskeinen* on yleisimmin nominatiivissa (lähes 43 % kaikista sijoista, esiintymiä yhteensä 239), mutta *tärkeä* partitiivissa (lähes 40 %, esiintymiä 238). *Keskeisen* toiseksi yleisin sijamuoto on partitiivi, mutta sen osuus on selvästi, 21 prosenttiyksikköä, pienempi kuin *tärkeällä*. *Tärkeän* toiseksi tavallisimpana sijana esiintyy taas nominatiivi, ja sen osuus on 10 prosenttiyksikköä pienempi kuin *keskeisen* vastaava osuus.

▷

Näiden sijojen osalta erot ovat selvät, ja sijataivutusta voidaan näiltä osin pitää — ei siis pelkästään kollokaatteja, vaan myös itse hakusanoja tarkasteltaessa — lähisynonyymeja erottavana tekijänä. Mikä sitten selittää kyseisen vaihtelun hakusanojen sijaesiintymistä? Tämä selviää tarkasteltaessa hakusanojen syntaktisia funktioita, jotka näkyvät taulukossa 6.

Taulukko 6. Hakusanojen *tärkeä* ja *keskeinen* syntaktiset funktiot.

	tärkeä		keskeinen	
	f	%	f	%
adjektiiviattribuuttina	312	51,8	477	85,3
luokituspredikatiivina	217	36,0	54	9,7
predikatiiviadverbiaalina	58	9,6	27	4,8
muuna jäsenenä	15	2,6	1	0,2

Tärkeällä ja *keskeisellä* on selvästi erilaiset syntaktiset tehtävät lauseessa: kun *keskeistä* käytetään suurimmaksi osaksi attribuuttina NP:ssa (noin 85 %), *tärkeä* toimii sekä attribuuttina (noin 52 %) että luokituspredikatiivina (36 %)⁹. Nimenomaan se, että *tärkeä* toimii usein luokituspredikatiivina, lisää partitiivin osuutta sen sijataivutuksessa, sillä partitiivi on adjektiivisen luokituspredikatiivin tavallisin sija (Hakulinen ja Karlsson 1988: 194–196). Useimmat adjektiivit esiintyvät partitiivissa silloin kun lauseessa on lause- tai infinitiivisubjekti tai kun subjektin tarkoite on aine-, kollektiivi-, abstrakti tai monikollinen sana (mp.). *Tärkeän* sisältävissä lauseissa on usein juuri infinitiivisubjekti (esimerkit 30 ja 31) tai lausesubjekti (32): lausesubjekteja on yhteensä 30 ja infinitiivisubjekteja 58 kappaletta. Vastaavat määrät ovat *keskeisellä* 2 ja 1. Tämä näkyi osittain jo edellä kolligaatti-analyysin yhteydessä, jolloin havaittiin, että verbien ja konjunktioiden määrä 1R-asemassa on *keskeisellä* pienempi — infinitiivisubjekti eli verbi voi seurata adjektiivia heti positiossa 1R ja lausesubjektin aloittaa useimmiten konjunktio *että*. Infinitiivisubjekti ei ole kuitenkaan aina tässä positiossa (esimerkki 31). Tällöin vieruskolligatteihin rajoittuva analyysi ei riitä antamaan oikeaa tulosta, vaan kotekstia on tarkasteltava laajemmin.

30. Neljänneksi on ollut *tärkeää pyrkiä* luomaan teoriaa, joka pyrkii ottamaan huomioon – –
31. – – mutta analyysin kannalta ensiarvoisen *tärkeää* on kuitenkin *muistaa* myös toinen seikka – –
32. Kun tietotekniikkaa käytetään yhä useammin tekstintulkinnan apuvälineenä, on *tärkeää, että ohjelmien käytön ongelmista ja hyödyistä keskustellaan*.

Tärkeän käyttö luokituspredikatiivina selittää siis partitiivien osuuden sen sijataivutuksessa, mutta mistä johtuu nominatiivien suurempi määrä *keskeisen* sijataivutuksessa?

⁹ Tuomikosken (1973) mukaan vajaat 69 % adjektiivien käytöstä muodostuu attribuutiivisesta käytöstä. Tähän verrattuna *tärkeä* esiintyy adjektiiveja harvemmin kongruoivana attribuuttina. Predikatiivien osuus on Tuomikosken tutkimuksessa vain vajaat 8 %, *tärkeällä* tämä funktio on siis huomattavasti tavallisempi kuin adjektiiveilla keskimäärin. *Keskeisen* attribuutiivisen käytön osuus on taas selvästi suurempi kuin adjektiivien keskimäärin, mutta predikatiivisen käytön osuus noudattaa yleistä linjaa.

Selitys löytyy siitä, että *keskeinen* on *tärkeää* yleisimmin adjektiiviattribuuttina NP:ssa. NP:iden tavallisimpina funktioina on toimia lauseen subjektina, objektina ja predikaatiivina. Etenkin subjektin ja myös identiteettipredikaatiivin tavallisin sija on juuri nominatiivi (Hakulinen ja Karlsson 1988: 182, 194). Koska *keskeinen* on yleinen nimenomaan attribuuttina ja koska attribuutti kongruoi (tässä nominatiivimuotoisen) pääsanansa kanssa, tämä syntaktinen funktio lisää nominatiivien osuutta *tärkeään* verrattuna. Seuraavissa esimerkeissä *keskeisen* sisältävä NP on subjektina (33) ja identiteettipredikaatiivina (34):

33. *Yksi keskeinen kiista* yhteiskunnallisten ongelmien tutkimuksessa ja konstruktivismiin piirissä koskee asiantilojen ja ongelmanmäärittelyjen välistä suhdetta.
34. *Kirjoittajien keskeinen teesi* on, että sosiaaliset ja pragmaattiset tekijät vaikuttavat luokittelujärjestelmiin – –

Kjellmerin (1990: 169–170) mukaan adjektiivit, jotka ovat syntaktiselta funktioltaan yleensä joko attribuutteja tai predikaatiiveja, voivat olla selkeästi kollokoivia; molemmissa tehtävissä esiintyvät adjektiivit kollokoivat sen sijaan heikommin. Syynä tähän on se, että vakiintuneessa rakenteessa esiintyvän lekseemin kontekstuaalinen ennustettavuus on helpompaa ja sen potentiaalisten kollokaattien määrä on rajatumpi kuin sellaisen lekseemin, joka voi olla usean rakenteen jäsen. Yleisesti ottaen Kjellmer pitää adjektiiveja kuitenkin heikosti kollokoivina: niiden kollokaatiovalinta ei ole rajoitettu (mp.) Miten Kjellmerin havainto toteutuu tässä käsiteltyjen adjektiivien kohdalla? Koska *keskeinen* on pääasiallisesti attribuuttiasemassa eli yhdessä syntaktisessa funktiossa lauseessa, sen pitäisi olla vahvemmin kollokoiva kuin *tärkeän*, jolla on selkeästi kaksi eri funktiota — attribuuttivinen ja predikaatiivinen. *Keskeisellä* on kuitenkin vähemmän merkitseviä kollokaatteja kuin lähisynonymillaan. Lisäksi adjektiivien kollokaatiosuhteiden määrä ja vahvuus (merkitsevyys) vaihtelevat kollokaattien position mukaan. (Ks. aiempaa kollokaatioanalyysia ja taulukkoa 4.) Edellä todettu *keskeisen* vahvempi kollokaatiosuhde 1R-positiossa olevien lekseemien kanssa voi olla kuitenkin selitettävissä sillä, että *keskeisen* käyttö on keskittynyt erityisesti yhteen syntaktiseen funktioon. Näiden adjektiivien kollokointia ja syntaktisen funktion suhdetta tarkasteltaessa Kjellmerin havaitsema kaava ei toteudu yksiselitteisesti, vaan kollokaatiosuhteiden määrä ja vahvuus ovat erilaisia riippuen myös kollokaattien positiosta, ei pelkästään syntaktisten funktioiden kirjosta.

Aivan lopuksi tarkastelen vielä näiden kahden lähisynonymien vertailtavuutta ja vertailuasteiden morfologiaa. Molemmat adjektiivit ilmaisevat 'tärkeyttä', joka ominaisuutena on luontaisesti vertailtavissa. Ne luetaan siksi relatiivisten (ks. Hakanen 1973: 21) adjektiivien ryhmään, jolle on ominaista sekä kvantitatiivisen vaihtelun että kvalitatiivisten komponenttien olemassaolo (Orpana 1988: 121). Vertailun sallivilla adjektiiveilla asterot voidaan ilmaista joko vertailumuodoilla tai etumääritteiden eli jo aiemmin käsitelmiäni astemääritteiden avulla (Hakulinen ja Karlsson 1988: 79; Quirk ym. 1985: 435). Edellä kollokaatioanalyysissa huomattiin, että *tärkeän* 1L-positiossa juuri astemääritteet olivat tilastollisesti merkitsevimpiä kollokaatteja ja että *keskeisellä* näitä ei ollut ainuttakaan. Kaikkiaan astemääritteiden määrä on kaikkien 1L-kollokaattien joukossa suurempi juuri *tärkeä*-adjektiivilla: sillä on etumääritteitä 69 kappaletta (11,5 % kaikista kollokaateista), kun *keskeinen* on saanut astemääritteen ainoastaan 18 kertaa, mikä on vain 3,2 % sen kaikista 1L-kollokaateista. Onko sitten adjektiivien herkkyydellä saada kvan-

tifioiva etumääräte yhteyttä adjektiivien vertailuasteiden määrään? Aiempaa tutkimusta tällaisesta vertailusta ei liene saatavilla. Taulukossa 7 on esitetty näiden kahden adjektiivien vertailuasteiden määrät ja niiden suhteelliset osuudet.

Taulukko 7. Hakusanojen *tärkeä* ja *keskeinen* vertailuasteet ja vertailuasteiden suhteelliset osuudet.

	tärkeä		keskeinen	
	f	%	f	%
positiivi	425	70,6	507	90,7
komparatiivi	57	9,5	11	2,0
superlatiivi	120	19,9	41	7,3

Kuten yllä olevasta taulukosta käy ilmi, *tärkeän* vertailuasteista 29,4 % muodostuu komparatiiveista ja superlatiiveista; vastaava yhteenlaskettu suhdeluku on *keskeisellä* 9,3 prosenttia. *Keskeistä* käytetään siten lauseissa selvästi useammin komparoirottomana. Komparaatiomuotojen ja etumääritteiden lisäksi *tärkeällä* on vielä erityisiä yhdyssanamuotoja, jollaisen kaltaiset ovat *keskeiselle* vieraita: *elintärkeä* esiintyy materiaalissa kuusi kertaa ja *hengentärkeä* kerran. Kyseisissä tapauksissa yhdyssanan alkuosalla on samanlainen kvantifioiva merkitys kuin astemääritteilläkin: *elintärkeän* ja *hengentärkeän* merkitys on kutakuinkin sama kuin *ehdottoman* tai *äärimmäisen tärkeän* (adjektiivien vahvikemorfeista ks. Tuomi 1991: 7–9). Astemääritteiden ja komparaatiomuotojen välillä näyttäisi olevan siten yhteys: selkeästi enemmän kvantifioivia määritteitä saava *tärkeä* esiintyy *keskeistä* useammin komparatiivissa ja superlatiivissa. Yleistä sääntöä adjektiivien morfologisin keinoin tehdystä vertailusta ja kvantifioivien etumääritteiden käytöstä ei voi yhden tutkimuksen perusteella tehdä, mutta havainto kohdistuu mielenkiintoiseen ilmiöön, joka voi varsin mahdollisesti koskea adjektiiveja yleisemminkin.

Kotekstianalyysin lisäksi myös itse hakusanojen tarkastelu osoittautui hyödylliseksi lähisyronymien tutkimisessa: Ensinnäkin eroja löytyi hakusanojen sijaituvutuksesta, mikä selittyy hakusanojen syntaktisten funktioiden erilaisuudella. Toiseksi adjektiivit käyttäytyivät eri tavalla komparaatiossa, ja komparaatiolla näyttäisi olevan yhteys etumääritteillä tehtyyn kvantifiointiin.

LOPUKSI

Tässä artikkelissa olen tarkastellut 'tärkeyttä' ilmaisevien adjektiivien lähisyronymipeseyttä ja erityisesti kahta lähisyronymia, *tärkeää* ja *keskeistä*. Olen pyrkinyt esittelemään korpuslingvistiikan suomin keinoin niitä eroja ja yhtäläisyyksiä, joita adjektiiveilla on ja jotka siten erottavat ja yhdistävät tutkimiani lähisyronymieja. Ensinnäkin erottavaksi tekijäksi osoittautui tekstilaji. Akateemisissa teksteissä ja kaunokirjallisuudessa — tutkimissani tekstilajeissa — adjektiivien käyttö erosi toisistaan: tutkitut adjektiivit olivat lähes yksinomaan akateemisille teksteille ominaisia, ja ainoastaan *tärkeä* oli yleinen kaunokirjallisuudessa, vaikka sekään ei yltänyt lähelle akateemisissa teksteissä il-

mennyttä frekvenssiä. Toiseksi: merkittäviksi tekijöiksi osoittautuivat myös monet kontekstuaaliset valintasäännöt. Lähisynonyymien kontekstissa eli lähimmässä tekstiyhteydessä oli runsaasti erottavia mutta tietenkin myös paljon yhdistäviä tekijöitä; tämä on edellytys, jotta sanoja voi pitää lähisynonyymeina.

Analyyseissä kiinnitin huomioni erityisesti *tärkeän* ja *keskeisen* leksikaaliseen kontekstiin eli kollokaatteihin sekä kieliopilliseen ympäristöön, tässä sanaluokkakollokaatteihin. Kollokaatioanalyyseissä merkittäviksi eroiksi nousivat erityisesti astemääritteiden ja kvanttoreiden käyttö IL-positiossa ja erilaiset kollokaatit IR-positiossa. Myös kollokaatiosuhteiden kiinteys (merkitsevyys) ja hakusanojen erilaisten kollokaattien määrä osoittautuivat näillä lähisynonyymeilla erilaisiksi — kuitenkin siten, että erottavana tekijä toimi myös kollokaatin positio. Lähisynonyymeja erottavia tekijöitä löytyi myös sanaluokkakolligaahteista ja substantiivikollokaattien sijaituvuudesta. Verbikollokaattien morfologia ei osoittautunut erottavaksi tekijäksi. Lopuksi tarkastelu kohdistui itse hakusanoihin sekä niiden morfologiaan ja syntaktiseen funktioon. Myös näiden piirteiden analyysi toi esiin sellaisia ominaisuuksia (sijaituvuus, komparaatio ja ero attributiivisen ja predikatiivisen käytön välillä), jotka osoittavat *tärkeän* ja *keskeisen* käytön erilaisiksi. Tavoitteenani oli myös tarkastella, pystyvätkö analyysin yhtäällä antamat tulokset selittämään tai tukemaan toisaalla saatuja tuloksia. Tämä osoittautui mahdolliseksi etenkin silloin, kun tavoitteena oli selvittää hakusanojen sijavaihtelua; se selittyi pitkälti hakusanojen syntaktisen funktion kautta. Kollokaatio- ja kolligatioanalyyseissä olivat toisiaan tukevia, ja jälkimmäisen tuloksia selitti osittain hakusanojen syntaktinen analyysi.

Saamiini tuloksiin saattaa vaikuttaa korpuksen koko, joka ei esimerkiksi suomen kielen tekstipankkiin verrattuna ole kovin suuri. Erityisesti kollokaatit voivat olla herkkiä pienessä korpuksessa korostuvalle idiolektien vaikutukselle. Yhden tekstintuottajan vaikutusta pyrin kuitenkin välttämään rakentamalla tutkimuskorpuksen siten, ettei mikään teksti yli- tai aliedustu siinä merkittävästi. Kolligaahteihin, syntaktiseen asemaan ja morfologiaan kohdistuneiden analyysien tulokset ovat sen sijaan otaksuttavasti jähmeämpiä muuttamaan, vaikka korpuksen kokoa kasvatettaisiinkin. Sen vuoksi pidän korpuksen kokoa tarpeeksi isona, kun tarkasteltavana ovat näinkin frekventit adjektiivit.

Analyyseissäni tulokset ovat rohkaisevia: Ensinnäkin ne tukevat jo aiemmin muista kielistä tehtyjä havaintoja synonyymipesyeiden jäsenten erilaisuudesta. Ennen kaikkea ne kuitenkin tuovat lisävaloa adjektiivien käyttöön — siihen, millaiset piirteet voivat lähisynonyymisten adjektiivien käytössä olla samanlaisia ja millaiset puolestaan erilaisia. Tutkimus herättää myös uusia kysymyksiä: miten adjektiivien syntaktiset funktiot vaikuttavat kollokaatiosuhteisiin ja millaisessa suhteessa toisaalta eri kvantifioinnin keinot ovat toisiinsa? Analyyseissä esiin tulleita semantiikan, syntaksin ja morfologian sekä osin myös tekstilajien tutkimuksen yhdistäviä adjektiivien »käyttöohjeita» on mahdollista selvittää vain suurten tekstimassojen ja niiden analyysiin soveltuvien ohjelmien avulla. Analyysiini ja yleensäkin korpuslähtöisen tutkimuksen tavoitteena onkin luoda kuvaus kielen todellisuudesta, joka ei aina tule ilmi intuition avulla tai yksittäisiä tapauksia tarkastelemalla.

LÄHTEET

- BAKER, MONA 1995: Corpora in translation studies: An overview and some suggestions for future research. – *Target* 7:2 s. 223–243. Amsterdam: Benjamins.
- BARNBROOK, GEOFF 1996: *Language and computers. A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- BEAUGRANDE, ROBERT DE 1999: Reconnecting real language with real texts: Text linguistics and corpus linguistics. – *International Journal of Corpus Linguistics*, Vol. 4 (2) s. 243–259. Amsterdam: Benjamins.
- BIBER, DOUGLAS – CONRAD, SUSAN – REPPEN, RANDI 1998: *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- BÄCKLUND, ULF 1973: *The collocation of adverbs of degree in English*. Acta Universitatis Upsaliensis 13. Uppsala.
- CARTER, RONALD 1987: *Vocabulary. Applied linguistic perspectives*. London: Allen & Unwin.
- CD-Perussanikirja 1997. Kotimaisten kielten tutkimuskeskuksen julkaisuja 94. Helsinki: Edita.
- CRUSE, D. A. 1986: *Lexical semantics*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- FIRTH, J. R. 1957: *Papers in linguistics*. London: Oxford University Press.
- 1968: *Selected papers of J. R. Firth 1952–1959* (toim. F. R. Palmer). London: Longman.
- FORSGREN, TUULI 1997: What is a collocation? – Mariann Skog-Södersved & Per Råberg (toim.), *Ethische Konzepte und mentale Kulturen 1. Auf der Suche nach ethnischen Konzepten in Wortschatz und Phraseologie* s. 59–64. Umeå studies in the humanities 139. Umeå.
- HAKANEN, AIMO 1973: *Adjektiivien vastakohtasuhteet suomen kielessä*. Suomalaisen Kirjallisuuden Seuran Toimituksia 311. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HAKULINEN, AULI – KARLSSON, FRED 1988: *Nyky-suomen lauseoppia*. Suomalaisen Kirjallisuuden Seuran Toimituksia 350. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HAKULINEN, AULI – OJANEN, JUSSI 1976: *Kielitieteen ja fonetiikan termistöä*. Suomalaisen Kirjallisuuden Seuran Toimituksia 324. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HOEY, MICHAEL 1991: *Patterns of lexis in text*. Oxford: Oxford University Press.
- ITKONEN, ERKKI 1966: *Kieli ja sen tutkimus*. Helsinki: WSOY.
- JONGE, BOB DE 1993: The existence of synonyms in a language: two forms but one, or rather two, meanings? – *Linguistics* 31 s. 521–538.
- KARLSSON, FRED 1999: *Yleinen kielitiede*. Helsinki: Yliopistopaino.
- KENNEDY, GRAEME 1991: *Between and through: the company they keep and the functions they serve*. – Karin Aijmer & Bengt Altenberg (toim.), *English corpus linguistics. Studies in honour of Jan Svartvik* s. 95–110. London: Longman.
- 1998: *An introduction to corpus linguistics*. New York: Longman.
- KJELLMER, GÖRAN 1987: Aspects of English collocations. – Willem Meijs (toim.), *Corpus linguistics and beyond*. Proceedings of the seventh international conference on English language. Research on computerized corpora s. 133–140. Amsterdam: Rodopi.

- 1990: Patterns of collocability. – Jan Aarts & Willem Meijs (toim.), *Theory and practice in corpus linguistics* s. 163–173. Amsterdam: Rodopi.
- LEECH, GEOFFREY 1990: *Semantics. The study of meaning*. Harmondsworth: Penguin Books.
- LEINO, ANTTI – LEINO, PIRKKO 1990: *Synonymisanasto*. Keuruu: Otava.
- LYONS, JOHN 1977: *Semantics. Vol. I–II*. Cambridge: Cambridge University Press.
- 1981: *Language, meaning and context*. London: Fontana.
- MAURANEN, ANNA 1998: Käännössiomi ja kääntämisen universaalit. Projektisuunnitelma. Kansainvälisen viestinnän laitos. Joensuun yliopisto.
- 2000: Strange strings in translated language: A study on corpora. – Maeve O’han (toim.), *Intercultural faultlines. Research models in translation studies 1. Textual and cognitive aspects* s. 119–141. Manchester ja Northampton: St. Jerome Publishing.
- ORPANA, TERTTU 1988: *Kuvaus vai kommentti. Tutkimus suomen kielen adjektiivadverbien semanttisesta tulkinnasta*. Opera Fennistica & Linguistica 2. Tampere.
- PALMER, F. R. 1976: *Semantics*. Cambridge: Cambridge University Press.
- PARTINGTON, ALAN 1998: *Patterns and meanings. Using corpora for English language research and teaching*. Amsterdam: John Benjamins.
- QUIRK, RANDOLPH – GREENBAUM, SIDNEY – LEECH, GEOFFREY – SVARTVIK, JAN 1985: *A comprehensive grammar of the English language*. London: Longman.
- RÉBÉK-NAGY, GÁBOR 1997: Modulation of authors’ claims in the introduction and discussion sections of medical research articles [verkkodokumentti]. Aston University, 1999, [viitattu 15.8.2000]. Saatavissa: <http://www.les.aston.ac.uk/lsu/diss>.
- SAUKKONEN, PAULI – HAIPUS, MARJATTA – NIEMIKORPI, ANTERO – SULKALA, HELENA 1979: *Suomen kielen taajuussanasto*. Helsinki: WSOY.
- SINCLAIR, JOHN 1991: *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- TOGNINI BONELLI, ELENA 1996: *Corpus. Theory and practice*. Birmingham: Tuscan Word Centre.
- TUOMI, TUOMO 1991: Töpötäysi ja patajuoppo. – *Kielikello* 4/1991 s. 7–9.
- TUOMIKOSKI, RISTO 1973: Adjektiivin ja adverbien erosta. – *Virittäjä* 77 s. 199–214.
- VEHMASKOSKI, MAILA 1995: Äijät ja ämmät, vaarit ja muorit. Isovanhempien nimitykset suomen murteissa. – Maija Länsimäki (toim.), *Sana kiertää. Kirjoituksia suomen kielestä* s. 70–75. Kotimaisten kielten tutkimuskeskuksen julkaisuja 80. Helsinki.
- VILKUNA, MARIA 1996: *Suomen lauseopin perusteet*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 90. Helsinki: Edita.
- WESTERLUND, HANNA 2000: Leksikaalisista kollokaatioista EU-asetusten käännöksissä ja Suomen lainsäädännön kielessä. – Päivi Jauhola, Outi Järvi & Detlef Wilske (toim.), *Erikoiskielet ja käännösteoria*. VAKKI-symposiumi XX, Vaasa 11.–13.2.2000. Vaasan yliopiston käännösteorian ja ammattikielten tutkijaryhmän julkaisu 27 s. 400–406, Vaasa.

TUTKIMUSMATERIAALI

Käännössiomen korpus. Käännössiomen sähköinen tutkimusaineisto. Käännössiomi ja kääntämisen universaalit -hankkeessa koostanut Joensuun yliopiston kansainvälisen viestinnän laitos 1997–. Saanti: kansainvälisen viestinnän laitoksen palvelin, Joensuun yliopisto.

▷

WHAT CAN CORPUS-LINGUISTIC ANALYSIS REVEAL ABOUT NEAR SYNONYMS?

Synonymy is a phenomenon of interest not only to linguists but to ordinary language users. Few 'instructions' are available, however, for using synonyms. Dictionary definitions, for example, do not explain the differences and similarities of synonyms but simply give indicative, largely stylistic guidance. The article examines how synonyms can be distinguished from each other using corpus linguistics. The aim is to identify the rules by which a word is selected from amongst a group of synonyms with paradigmatic semantic relationships to be used in a relationship which is syntagmatic.

The article is based on the theory (e.g. Cruse 1986; de Jonge 1993) that synonyms have both features in common and features that distinguish them from each other. For this reason, the term 'near synonym' (Lyons 1981) can better illustrate the relationship of the words to each other than the standard term 'synonym'. Near synonyms are distinguished from each other either by extralinguistic selection criteria, such as the circumstances of communication and the associated aims, or intralinguistics selection criteria, such as text type, style and linguistic phenomena. The article focuses on intralinguistic criteria and especially the rules evident in the cotext of a node word.

The writer examines the phenomenon using two adjectives commonly found in academic texts in particular: *tärkeä* and *keskeinen* (both meaning 'important', 'significant'). Corpus-linguistic methods are used in the analysis of these adjectives, as the research material is a collection of texts in electronic form. Synonymy is examined by studying the cotext of node words and their properties. Use of near synonyms is guided by collocational (lexicon-related) and colligational (grammatical-category related) selection arising in the cotext. In addition, morphological variation evident in the cotext has a partial function in distinguishing near synonyms from each other. Analysis of node words, on the other hand, reveals properties related to the syntactic use and to the morphology of the adjectives as distinguishing features. Common features of the adjectives are also clearly distinguishable in the analysis, which is, of course, necessary if the words are to be considered near synonyms. The results obtained in each part of the analysis explain and support the observations made in the other parts. Syntactic 'instructions', for example, prove to be features explaining morphological variation. ■

Kirjoittajan osoite (address):
Kansainvälisen viestinnän laitos
Joensuun yliopisto
PL 48
57101 Savonlinna
Sähköposti: *jarmo.jantunen@joensuu.fi*