

Informaatiotutkimuksen päivät 2010  
21. - 22. lokakuuta, Tampere

ABSTRAKTI

*Susanna Nykyri*

## ***Laadullinen bibliometriikka sanastotyön apuvälineenä : yhteissana-analyysi monikielisen tesaauruksen laadinnassa***

*Susanna Nykyri, Informationsförvaltning, Åbo Akademi, susanna.nykyri@abo.fi.*

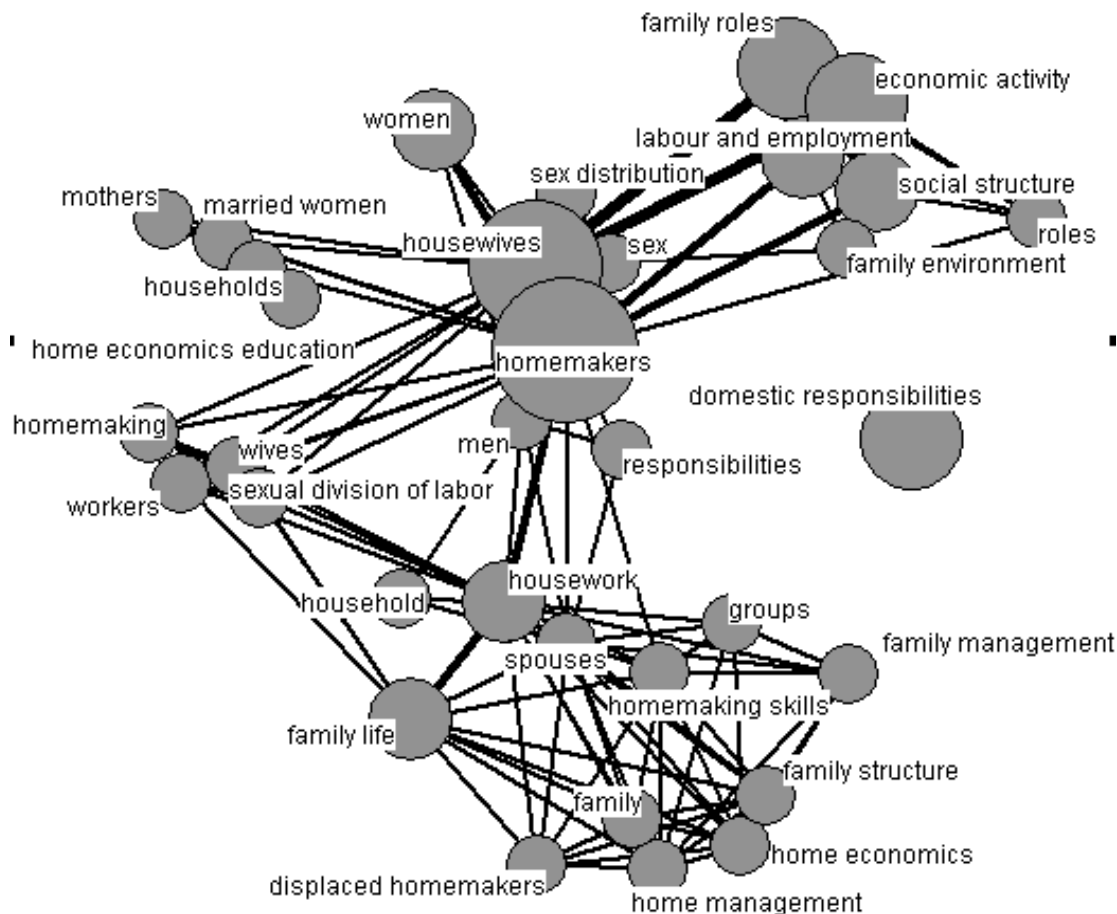
Tiedontallennus ja -haku yli kansallisrajojen lisääntyy kaiken aikaa. Kansainvälisten tietovarantojen luomisen ja käytön onnistuminen edellyttää kuitenkin yhteisiä työkaluja ja yhteisymmärrystä käsitteiden käytöstä. Tutkimuskohteena (Nykyri 2010) oli monikielinen ja -kulttuurinen yhteiskuntatieteellinen digitaalinen tiedonhaun tesaurus, jonka käyttöympäristönä on Internet. Siitä voidaan erottaa kahdeksan elimellisesti toisiinsa liittyvää aspektia - kieli, terminologia, kulttuuri, yhteiskuntatiede, tesaurus, digitaalisuus, tiedontallennus ja tiedonhaku. Tutkimuksessa keskityttiin selvittämään yhteiskuntatieteellisen tesaauruksen asiasanojen käännettävyyttä (kulttuuri- ja kielisidonnaisuutta) englannista suomeen ja ekvivalenssin käsitettä monikielisessä tesaauruksessa.

Tesaurusta tarkasteltiin kulttuurisena tuotteena. Jokainen tiedonhaku tietokannassa käsittää ainakin viisi erilaista kieltä: kirjoittajan, sisällönkuvailijan, synteettisen rakenteen, tiedonhakijan ja tiedonstrategian (Buckland 1999). Nämä kaikki edustavat toisistaan poikkeavaa diskurssia.

Kääntämisen merkitys on korostunut valtavasti esimerkiksi Internetin yleistymisen ja EU:n myötä, jotka osaltaan luovat tarvetta kehittää toimivampia tiedonhallinnan apuvälineitä. Ei ole kuitenkaan aina selvää, että mitä käännoksellä tavoitellaan – onko tavoite kotouttaa vastine kohdekulttuuriin vai heijastella lähtökulttuuria vai kenties pyrkiä kansainväliseen, mahdollisimman neutraaliin sisältöön (Koskinen 2000, Lindfors 2001, Ruokonen 2004). Yksi tapa tarkastella käännettävyyttä ja vertailla käännosvastineita on bibliometrinen yhteissana-analyysi (*co-word analysis*, ks. Persson 1991), joka oli yksi tutkimuksessa käytetyistä menetelmistä ja ko. menetelmän uusi aluesovellus.

Callonin, Courtialin ja Lavillen (1991) mukaan yhteissana-analyysin metodologinen perusta on ajatuksessa, että sanojen yhteisesiintyminen ja niistä muodostunut verkko kuvaavat dokumenttien sisältöä. Yhteissana-analyysi on hyvin samankaltainen kuin yhteisviittausanalyysi (*co-citation analysis*). Yhteissana-analyysissa tarkastellaan dokumentissa esiintyvien sanojen esiintymistä yhdessä, kun taas yhteisviittausanalyysissa tutkitaan yhteisiä lähdeviitteitä. (Horton et al. 1998) Yhteissana-analyysilla pyrittiin jäljittämään siis pikemminkin termien välisiä suhteita kuin dokumenttien liittymistä toisiinsa. Välineenä käytettiin Bibexceliä (Bibexcel 2001).

Tutkimuksen pääkysymys käsitteli ekvivalenssin käsitettä monikielisen ja –kulttuurisen luomisessa. Yksi metodisista alakysymyksistä koski yhteissana-analyysin käyttöä käännettävyyden tarkastelussa – mikä on sanojen pragmaattinen merkitys indeksoinnissa ja eroavatko merkitykset ja diskurssit eri tietokannoissa. Lisäksi menetelmää käytettiin haastatteluaineiston sana-assosiaatioiden analyysissä kuvaamaan termien käsitteellistä kontekstia ja sitä kautta selventämään sen semanttista ja pragmaattista merkitystä.



Kuva 1: ”Homemakers” tesaurusissa (esimerkki menetelmästä ja aineistosta)

Bibliometrinen yhteissana-analyysiä käytettiin siis uudella tavalla laadullisen aineiston kuvaamisessa ja tulkinna. Tämän tyyppisessä aineistossa bibliometriset kartat tarjoavat sekä kompaktin että monipuolisen kuvan tutkittavien termien leksikaalisesta verkostosta ja sanastollisista yhteisesiintymistä. Aineistoon ja menetelmään liittyvät ongelmat olivat varsin perinteisiä – esim. aineiston harmonisointi vs. autenttiset sanamuodot ja tietokantojen laatu ja puutteet (vrt. von Ungern-Sternberg 1994) - ja ne vaativat jatkuvaa tasapainottelua. Jotta on mahdollista antaa yleiskuva tai kuvittaa pääteemat, on erilaisten kielellisten ilmausten harmonisointi joskus välttämätöntä, mutta samalla menetetään paljon sävyjä. Aineiston käsittelyvaiheessa tehtävä harmonisointi sisältää myös merkittävän väärinkäsityksen riskin, kun käsitteiden sisältöjä ei ole siinä vaiheessa vielä tarkemmin analysoitu, mutta niitä harmonisoimalla puututaan niiden sisältöön. Menetelmä oli vaativampi haastatteluaineiston analyysissä kuin tietokantojen asiasanoituksessa. Ensinnäkin, sana-assosiaatiot olivat usein yksittäisen sanan tai sanaliiton asemesta pitkiä pohdiskeluja. Toiseksi, suomenkielinen aineisto vaati kääntämistä raportin julkaisukielelle eli englanniksi, ja aineistossa esiintyi paljon monitulkintaisia sanoja (esim. *poika* – *boy* vs. *son*, *kasvatus* – *education* vs. *pareting*). Erilaisten indeksointikäytäntöjen vuoksi tietokantadiskurssien analyysi oli tehtävä

enemmänkin termi- kuin asiasanatasolla, esim. asiasanoitus “*kotityöt/homework – miehet/men, koti-isät/stay-at-home fathers*” käsiteltiin analyysissä kolmena eri terminä: *homework, men, stay-at-home fathers*. Kun aineistoa muokattiin sanakarttana esitettävään muotoon, joitain sivumerkityksiä ja sävyjä siis menetettiin.

Tulosten perusteella yhteissana-analyysi on käyttökelpoinen menetelmä tutkittaessa termien käännettävyyttä. Menetelmän selkeä ansio on, että se tekee kontekstuaalisen ekvivalenssin tarkasteltavaksi ja näkyväksi. Se siis antaa käsityksen käännöksen dynaamisesta ekvivalenssista ja edelleen asiasanakäännöksen tulevasta käytettävyydestä. Se ei luonnollisestikaan tarjoa vastausta kysymykseen miten kääntää, mutta sitä voidaan käyttää eri vaihtoehtojen vertailussa. Sanan semanttista ja pragmaattista merkitystä selvittäessä menetelmästä on apua, kun vertaillaan erilaisia käyttöympäristöjä (esim. tietokannoissa olevia termiin liittyvien dokumenttien asiasanoituksia). Menetelmän käyttökelpoisuus etenkin ekvivalenssin tarkastelussa on kuitenkin sidoksissa muihin käytettyihin menetelmiin, tässä tapauksessa sana-assosiaatioihin (ks. kuva 1) ja semanttiseen komponenttianalyysiin (ks. kuva 2) sekä teoreettisen viitekehysten keskeisiin käsitteisiin. Menetelmää ei ole aiemmin käytetty vastaavaan tarkoitukseen kuin tässä tutkimuksessa eli kyseessä on uusi mielenkiintoinen sovellusalue.

HOMEMAKER characteristics:	Merriam-Webster	OED Online(a)	OED Online(b)	Cambridge International Dictionary of English	Dictionary.com	Word Net	NHD
woman	+	++				++	
mother	+						
wife	+	++				++	
housewife	+	++				++	
family member	+	++	+	++		++	++
manages a household	++	+	++	++	++	++	++
non-paid	+	++		+		++	++
takes care of children	+			+			++

Kuva 2: ”Homemakers” sanakirjoissa ja semanttinen komponenttianalyysi (esimerkki menetelmästä ja aineistosta)

## Lähteet

*Bibexcel - a toolbox for bibliometricians*. Version 2001-05-21. Developed by Olle Persson, Inforsk, Umeå University, Sweden. Uploaded from the World Wide Web <URL: <http://www.umu.se/inforsk/Bibexcel/index.html>>

Buckland, Michael (1999). Vocabulary As A Central Concept In Library And Information Science. Preprint of paper published as "Vocabulary as a Central Concept in Library and Information Science" I: Digital Libraries: Interdisciplinary Concepts, Challenges, and Opportunities. Proceedings of the Third International Conference on Conceptions of Library

and Information Science (CoLIS3, Dubrovnik, Croatia, 23-26 May 1999. Ed. by T. Arpanac et al. Zagreb: Lokve, s. 3-12

Callon, M., Courtial, J.P., Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. - *Scientometrics* 22. (1): 155-205

Horton, Thomas & Coulter, Neal & Grant, Emanuel (1998). Applying Content Analysis to Humanities Computing Research Literature. Joint International Conference ALLC/ACH'98, Association for Literary and Linguistic Computing, Association for Computers and the Humanities, 5 – 10 July 1998, Lajos Kossuth University Debrecen, Hungary. URL: <<http://lingua.arts.klte.hu/allcach98/abst/abs20.htm>> (23.4.2001)

Koskinen, Kaisa (2000). Beyond Ambivalence. Postmodernity and the Ethics of Translation [Ambivalenssin tuolle puolen. Postmoderni aika ja kääntäjän etiikka]. *Acta Electronica Universitatis Tamperensis*; 65, Tampereen yliopisto 2000. URL: <<http://acta.uta.fi/pdf/951-44-4941-X.pdf>>] (7.12.2005)

Lindfors, Anne-Marie (2001). Respect or ridicule: Translation strategies and the images of a foreign culture. Helsinki English Studies. - The Electronic Journal of the Department of English at the University of Helsinki. Volume 1, 2001. URL: <[http://www.eng.helsinki.fi/hes/Translation/respect\\_or\\_ridicule1.htm](http://www.eng.helsinki.fi/hes/Translation/respect_or_ridicule1.htm)> (7.12.2005)

Nykyri, Susanna (2010). Equivalence and Translation Strategies in Multilingual Thesaurus Construction. Doctoral thesis. Åbo: Åbo Akademi University Press.

Persson, Olle (1991). *Forskning i bibliometrisk belysning*. Umeå: INUM.

Ruokonen, Minna 2004. ”Schleiermacher, Berman ja Venuti: kolme käännösteoreettista näkökulmaa vieraannuttamiseen.” Teoksessa: Tommola Jorma (ed.) 2004: *Kieli, teksti ja kääntäminen. Language, text and translation*, pp. 63-80. Turku: Turun yliopisto.

von Ungern-Sternberg, Sara 1994. *Verktyg för planering av tvärvetenskaplig informationsförsörjning. En tillämpning på ämnesområdet bioteknik i Finland*. (A tool for planning information provision in interdisciplinary fields, applied to the field biotechnology in Finland). Åbo Akademi Förlag – Åbo Akademi University Press, Åbo. 318 pp. [Includes summary in English.]