

Jarmo Saarti

Kaunokirjallisuuden indeksointi ja kaunokirjallisuuden tesauruksen kehittäminen

Kaunokirjallisuuden indeksointi ja kaunokirjallisuuden tesauruksen kehittäminen [The indexing of fiction and the development of fiction thesauri]. Informaatiotutkimus 17 (4): 69-79, 1998.

The article discusses firstly theoretical questions about fiction indexing. Previous studies are divided into two categories: firstly general papers about fiction indexing and aspects to be indexed about fiction, secondly making book indexes for fictional works. After that the paper presents different attempts to create word-lists and thesauri for fiction, in which the greatest emphasis is in presenting the structure and desining of the Kaunokki - the Finnish fiction thesaurus which was published in 1996.

Address: Brahenkatu 3 B 13, FIN-70100 Kuopio, Finland. Email: jarmo.saarti@oulu.fi.

1. Johdanto

Kaunokirjallisuuden tallennus ja haku ovat viime vuosina nousseet erityisen kiinnostuksen kohteeksi informaatiotutkimuksen parissa. Tähän ovat vaikuttaneet useat seikat. Suurin niistä lienee, että kaunokirjallisuuden sisällönkuvauksen tarve ja toisaalta sen mahdollisuudet ovat kasvaneet tietoteknisten välineiden kehityksen myötä. Nämä välineet ovat mahdollistaneet fiktiiviselle aineistolle ominaisen moniulotteisen kuvailun toteuttamisen. Toisaalta tietotekniikan tulon myötä on havaittu, että suuri osa kirjastojen välittämästä aineistosta on jäänyt indeksoimatta vuosien saatossa ja että juuri tämä osa aineistoa on ollut pääasiassa fiktiivistä.

Lisäksi vähitellen tapahtuva siirtyminen aineiston digitaaliseen jakeluun korostaa fiktiivisen aineiston sisällönkuvailun tarvetta. Tämä sen vuoksi, että kuvailematonta aineistoa on vaikea löytää ja myös jaella digitaalisessa ympäristössä. Fiktiivisen aineiston osalla kuvailun tärkeyttä korostaa

sekin, että vapaatekstihaku on sen kohdalla tehottomaa verrattuna esimerkiksi luonnon-tieteisiin, joissa teksteille on ominaista yksiselitteisyys ja alaspesifisyys. Informaatiotutkimuksen kannalta hedelmällisin vaikutus on kuitenkin sillä, että kaunokirjallisuus ja sen välitys asettaa perinteiset tiedonvälitysmallit kyseenalaiseksi ja näin ollen testaa ja laajentaa informaatiotutkimuksen teoreettista välineistöä. (Vrt. esim. Beghtol 1994b ja 1997; Green 1997.)

Seuraavassa käsitellään kaunokirjallisuuden sisällönkuvailua ja erityisesti sen indeksointia. Pääpaino on kaunokirjallisuuden sisällönkuvailun aspektien tutkimuksen sekä Kaunokin, suomalaisen kaunokirjallisuuden tesauruksen kehittämistyön esittelyllä ja sillä, miten edellinen on vaikuttanut Kaunokin rakenteeseen ja kehitykseen. Kaunokirjallisuus on tässä rajattu suorasanaiseen proosaan, lyriikka ja draama on jätetty tietoisesti käsittelyn ulkopuolelle. Ei siksi, että ne olisivat vähäarvoisempia, vaan siksi, että niiden ominaislaatu asettaa sisällönkuvaukselle vielä suurempia haasteita kuin suorasanaisten proosan.

2. Kaunokirjallisuuden indeksointi

Kaunokirjallisuuden indeksointia koskeva tutkimus voidaan jakaa kahteen pääryhmään: ensinnäkin niihin, joissa käsitellään yleisellä tasolla kaunokirjallisuuden indeksointia ja sen perusteita¹, toiseksi niihin, joissa käsitellään teoskohtaisten indeksien laatimista. Teoskohtaista indekseistä on kirjoitettu eniten anglo-amerikkalaisissa maissa, joissa perinteisesti on korostettu teoskohtaisten indeksien laatimista.

Kaunokirjallisuuden indeksoinnin yleisperusteiden käsittelyssä ovat anglo-amerikkalaisten lisäksi olleet merkittäväällä sijalla pohjoismaat ja niistä etenkin Tanska, jossa Annelise Mark Pejtersenin työ on ollut uraa uurtavaa. Kirjoituksissa on lisäksi käsitelty mm. kaunokirjallisuuden kompleksisuuden hallintaa kuvailussa sekä aboutness-käsitettä kaunokirjallisuudessa (Andersson & Holst 1996, Beghtol 1992, Bell 1991, Pulli 1992, Ranta 1991). Kahteen viimeiseen ryhmään liittyy läheisesti Sanna Kivimäen tutkimus siitä, kuinka hyvin romaanin takakansi kuvailee teoksen sisältöä ja voiko sitä käyttää apuna esimerkiksi kirjastossa (Kivimäki 1992). Tähän ryhmään voidaan lukea osin myös kehittyneet, tekoälyä hyväksi käyttävät tiedonhakujärjestelmät (Rich 1979 ja 1986). Lisäksi on raportoitu yksittäisissä kirjastossa toteutettuja kokeiluja (esim. MacPherson 1987, jossa kerrotaan koulukirjaston lastenkirjallisuuden indeksoinnista).

2.1. Kaunokirjallisuuden indeksointi-perusteet

Kaunokirjallisuuden sisällönkuvauksen tutkimuksessa eräänä olennaisena kiinnostuksen kohteena on ollut löytää ne seikat, jotka kaunokirjallisesta teoksesta tulisi indeksoida ja/tai luokitella hakua varten. Perinteisesti kaunokirjallisuuden luokituksessa on korostettu teoksen muotoa ja sen julkaisuajankohtaa ja julkaisumaata, myöhemmin myös joitakin spesifejä sisällöllisiä seikkoja. Nämä ovat edelleenkin suurten

yleisluokitusten eräs pääluokitusperusteista (Iivonen 1985, 9-23; Juntunen & Saarti 1992, 74 ja 139). Muoto, aika ja maantieteellinen alue ovat siitä kiitollisia luokitusperusteita, että niiden määrittely on varsin helppoa ja yksiselitteistä. Sen sijaan varsinaisen sisällön kuvailu on jäänyt paitsioon kaunokirjallisuuden luokitteluissa.

Ranta erottaakin kaksi perustasoa kaunokirjallisuuden kuvailussa – denotatiivisen ja konnotatiivisen. Denotatiivisella tasolla esitetään teoksista faktuaaliset asiat kuten esimerkiksi henkilöt, miljö, juonen faktuaaliset seikat. Konnotatiivisella (eli imaginatiivisella) tasolla esitetään taas seikat, jotka liittyvät teoksen teemaan ja sen tulkintaan, sekä tunneilmaisuuksiin liittyvät seikat. (Ranta 1991, 20-23.) Tässä Ranta hyödyntää Shatfordin Panofskyn pohjalle rakentamaa jaottelua kuvien indeksointiin. Shatford jakaa merkityksen faktuaaliseen ja ilmaisulliseen (factual and expressional meaning). Merkitysten ero on siinä, että faktuaaliset merkitykset ovat objektiivisia, ilmaisulliset subjektiivisia. Edellinen kuvaa, mitä kuvassa on, jälkimmäinen, mistä kuva on. ("The former describes what the picture is Of, the latter, what it is About.") Faktuaalisten merkitysten indeksointi on paljon helpompaa kuin ilmaisullisten. Of-kategorioita Shatford erottaa lisäksi kaksi – spesifisen (specific of) ja geneerisen (generic of). (Shatford 1986, 42-50.)

Lisäksi kaunokirjallisuuden luokituksille on ollut ominaista teoreettinen lähestymistapa. Niinpä perinteiset denotatiiviset luokitukset, joiden lähtökohtana on ollut romantiikan kauden historiallis-kielitieteellinen ja kasvatuksellis-ideologinen perinne² eivät juurikaan ota huomioon varsinaisia käyttäjien tarpeita. Tästä johtuen tanskalainen A. M. Pejtersen teki kyselytutkimuksen tanskalaisissa yleisissä kirjastoissa. Sen tarkoituksena oli selvittää mitä asioita kirjaston käyttäjät haluaisivat teoksista kuvattavan. Hän havaitsi, että käyttäjien teoksia koskevat kysymykset voitiin jakaa neljään taulukossa 1. lueteltuun ryhmään (Pejtersen 1983, 234).

Pejtersenin kategoriat voidaan jakaa konnotatiivisiin (aihe ja miljö) sekä denotatiivisiin (tekijän intentio). Lisäksi hänellä on mukana sisällönkuvailussa perin-

teisesti bibliografisen kuvailun puolelle jätettyjä seikkoja ryhmässä lähestyttävyyttä (mm. teoksen fyysinen muoto, sen sijainti kirjastossa). Niinpä lukijoitten tarpeitten mukaan luodusta kaunokirjallisuuden kuvailujärjestelmästä tulee hyvin monipuolinen ja se sisältää sekä helposti kuvattavia seikkoja että vaikeita ja tämän hetken käytännölle vieraita seikkoja (mm. arvottaminen). Lisäksi Pejtersenin tuloksista huomaa, että kirjaston työntekijöille selvä jako teoksen bibliografiseen kuvailuun ja sisällön kuvailuun on käyttäjille toisarvoinen seikka – olennaista on, että he saavat tarpeelliset tiedot etsiessään itselleen lukemista. Rebecca Green esittääkin, että kaunokirjallisuutta kuvaavat indeksointitermit tulisi jakaa sisältöä kuvaaviin ja attribuutteja kuvaaviin (subject indexing vs. attribute indexing). Edelliset hän rajaa pelkästään topikaalisen indeksoinnin piiriin eli termeiksi, "that reflects what a document or a user need is about." Tällöin: "This leaves attribute indexing to reflect such other characteristics of documents and user needs as language, recency, author affiliation, intended audience, and so on." (Green 1997, 86.)

Ongelmallisin kohta Pejtersenin kuvailussa on tekijän intentio, joka kuvailijan kannalta on täysin tulkinnallinen seikka, etenkin kun

halutaan kuvailla tunnekokemusta, joka ei välttämättä liity teokseen sinänsä, vaan kokijaan eli tässä tapauksessa lukijaan.⁴ Ongelmalliseksi tämän käsitteen tekee se, että itse teoksesta on vaikeaa, jollei mahdotonta päätellä tekijän intentiota. Lisäksi, kuten Wellek ja Warren huomauttavat, tekijä saattaa tulkita oman intentionsa väärin: "It happens to all of us that we misinterpret or do not fully understand what we have written some time ago (Wellek 1980, 148)." Lisäksi tekijän intention selvittäminen vaatisi sen kysymistä tekijältä itseltään, joka on hyvin vaikeaa ja monessa tapauksessa jo mahdotonta.

Pejtersenin luokitusta kehittivät edelleen Rolf Andersson ja Erik Holst tutkimuksessaan, jossa he haastattelivat 100 käyttäjää kahdessa ruotsalaisessa kirjastossa. Sitten he analysoivat näin saadut romaanien juonen (plot) kuvailut ja vertasivat niitä lisäksi kirjastoissa tehtyihin indeksointeihin. (Andersson 1996, 88.) He täydensivät ja muokkasivat Pejtersenin esittämästä ryhmittelystä oman mallinsa, joka sisälsi taulukossa 2. (sivulla 72) esitetyt ryhmät.

Ruotsalaiset tutkijat ovat lisänneet Pejtersenin kategorioihin kaunokirjalliseen viestintään olennaisesti kuuluvia aspekteja. Näitä ovat mm. lainattu motiivi, tarkemmin ana-

Taulukko 1. A. M. Pejtersenin kaunokirjallisuuden kuvailuaspektit

1. Aihe Teoksen aiheisältö eli se, mitä tarina kertoo.	-toiminta (action and course of events) -psykologinen kehittäminen ja kuvaus (psychological development and description) -sosiaaliset suhteet (social relations) ³
2. Miljö Tekijän teokselle valitsema ympäristö – aika-kausi ja paikka.	-tapahtuma-aika (time) -paikka/ympäristö (place) – maantieteellinen, sosiaalinen, ammatillinen
3. Tekijän intentio (author's intentio) Teoksen teema – teoksen tekijän suhde aiheeseensa nähden. Niiden ajatusten ja tunteiden joukko, jonka tekijä haluaa välittää lukijoilleen.	-tunnekokemus (emotional experience) -tietosisältö (cognition and information)
4. Lähestyttävyyttä (accessibility) Kommunikaatiota helpottavat tai estävät tekijät, esim. kielen vaikeusaste, kompositio, tyografia, koko.	-luettavuus (readability) -fyysiset ominaisuudet, kirjallinen muoto (physical characteristics, literary form)

lysoitu teoksen tapahtumasisältö sekä lukijan kokemus. Ongelmallisen 'tekijän intention' he ovat puolestaan korvanneet 'sanoman'-käsitteellä.

Mielenkiintoista edellä käsitellyissä kaunokirjallisuuden sisällönkuvailun kategorialuetteloissa on se, että hyvin harvat ottavat mukaan esteettiseen objektiin olennaisesti liittyvän arvottavan elementin. Eräänä syytä tähän on tietysti se, että arvottaminen on yleensä hyvin subjektiivista ja sopii siten huonosti perinteisen sisällönkuvauksen arvo maailmaan, jossa pyritään objektiivisuuteen. Toisaalta arvottaminen kuuluu olennaisena osana esteettiseen kulttuuriin, joten sen huomiotta jättäminen teosten kuvailussa unohattaa siitä erään olennaisen aspektin pois. Lisäksi näyttäisi siltä, että lukijat haluavat jonkinlaista teosten arvottamista. Tämän voi havaita esimerkiksi kaunokirjallisuuden markkinoinnissa, kritiikissä ja erilaisissa palkinnoissa. Erään mahdollisuuden jonkinasteiseen objektiivisuuteen arvottamisessa tarjoaa Hilderleyn ja Raffertyn (1997) esittämä demokraattisen indeksoinnin malli. Siinä annetaan käyttäjille mahdollisuus osallistua indeksointiin. Näin saadut raakakuvailut suodatetaan sitten yleiseen käyttöön tarkoitetuksi tietokannaksi. Tämä näyttäisikin olevan eräs mahdollisuus kehittää fiktiivisten, henkilökohtaiseen tulkintaan perustuvien

aineistojen kuvailua.

Edellä esitetyistä kategorialutteloista voi huomata myös sen, että kuvailtavat asiat pyritään rajoittamaan mahdollisimman objektiivisiksi (konnotatiivisiksi). Tosin Pejtersenillä ja hänen pohjaltaan oman systeeminsä kehittäneillä Rolf Anderssonilla ja Erik Holstilla on mukana myös joitakin muuttuvia, lukijan kokemukseen perustuvia аспекteja. Molempien esittämistä kategorioista jää kuitenkin tyystin puuttumaan teoksen tulkintojen historia sekä teoksen sijoittuminen kirjallis-historialliseen jatkumoon, joka joidenkin teosten kohdalla olisi mielenkiintoinen ja lisävalaistusta antava aspekti. Myös tekijä ja hänen roolinsa jäävät näissä luetteloissa pitkälti toisarvoiseksi. Toisaalta tämä osoittaa sen, että kaunokirjallisuuden sisällönkuvailussa on tehtävä rajauksia niiden asioiden suhteen, joita teoksista kannattaa kuvailla. Epäilemättä tähän vaikuttaa myös se ympäristö, mihin kuvailu tehdään.⁵

2.2. Teoskohtaiset indeksit

Toinen merkittävä ryhmä kaunokirjallisuuden indeksointia koskeva tutkimuksia on spesifisti yksittäisen kaunokirjallisen teok-

Taulukko 2. Andersson & Holstin kaunokirjallisuuden kuvailuaspektit

1. Tapahtumat (Phenomena)	1A. Toimijat ja heidän luonteenpiirteensä (Agents and their characters) 1B. Tapahtumien kulku (Course of events) 1C. Tunteet (Feelings) 1D. Suhteet (Relationships) 1E. Esineet (Objects) 1F. Abstraktit tapahtumat (Abstract phenomena)
2. Miljö	2A. Aika (Time) 2B. Maantieteellinen paikka (Geographic space) 2C. Yleinen paikka (General space)
3. Tekijän intentio (Author's intention)	3A. Sanoma (Message) 3B. Lukijan kokemus (Reader's experience) 3C. Kirjallinen muoto (Literary form) 3D. Lainattu motiivi (Borrowed motif) 3E. Genre

sen indeksointia käsittelevät kirjoitukset. Ransley (1987a ja 1987b) indeksoi kaksiosaisessa artikkelissaan Nathaniel Hawthornen romaanin *Scarlet Letter* ja käsittelee samalla kaunokirjallisuuden ja erityisesti yksittäisen kaunokirjallisen teoksen indeksoinnin ongelmia. Hänen mukaansa kaunokirjallisuuden sisällön indeksoinnin voi jakaa karkeasti kahteen tyyppiin, denotatiiviseen ja konnotatiiviseen viittaamiseen (Ransley 1987a, 46-47). Denotatiivisella viittaamisella hän tarkoittaa viittauksia niihin kaunokirjallisen teoksen aspekteihin, jotka ovat samankaltaisia kuin faktuaalisissakin kirjallisuudessa (esim. paikat, ajat ymv.). Konnotatiivinen viittaaminen puolestaan tarkoittaa kaunokirjallisen teoksen teemojen ja symbolien kuvailua (Ransley 1987b, 108).

Denotatiivinen viittaminen on ollut Ransleyn mukaan tyyppistä myös kaunokirjallisuuden indeksoinnissa. Sen sijaan konnotatiivinen viittaminen on ollut harvinaisempaa, koska se on työläämpää ja perustuu osin subjektiivisiin seikkoihin. Niinpä konnotatiivinen indeksi onkin enemmän lukijan opas (reader's guide) -tyyppinen johdatus teokseen ja sen maailmaan kuin perinteinen asiahakemisto (Ransley 1987a, 49). Tällainen hakemisto lähenee hengeltään kirjallisuuskritiikkiä, "...novel index can be, in itself, a rudimentary form of criticism..." (Ransley 1987a, 53)". Hakemisto voi ottaa huomioon myös kirjallisuushistorian tutkimustulokset antamalla lukijalle esimerkkejä erilaisista lukutavoista, joilla etenkin kirjallisuuden klassisia teoksia on tulkittu (Ransley 1987b, 104).

Tällaisella denotatiivisella hakemistolla on vaikutuksia myös itse indeksoinnin tekniikkaan. Ensinnäkin Ransleyn mukaan juonen mukana kehittyviä aspekteja kuvailevien termien esitysmuodon on oltava kronologinen eikä aakkosellinen. Näitä elementtejä ovat mm. symbolit ja teoksen tematiikka. (Ransley 1987a, 50-51). Tämä sen vuoksi, että esimerkiksi symboliikka ja yksittäisten symbolien merkitykset kehittyvät teoksen kronologian myötä ja niiden esittäminen aakkosissa ei paljastaisi lukijalle tätä kehitystä.⁶ Niinpä:

A novel indexer will have to decide whether it

is the specific occurrences of the noun named in the heading that are to be collated or whether a broader classified heading will be used as an umbrella under which to shelter a number of thematically linked images (Ransley 1987a, 52).

Toiseksi tällaisessa indeksoinnissa sanat eivät riitä aina kuvailutermeiksi vaan mukaan on otettava kokonaisia lauseita (Ransley 1987b, 108). Pejtersen kehitti tälle rakenteelle syntaktiset säännöt omassa indeksointi-järjestelmässään ja kutsui näin aikaansaatuja asiasanaketjuja konkreettisiksi (concrete indexing strings) (Pejtersen 1996, 25). Ransley korostaa, että kuvattaessa etenkin symboliikkaa ja pyrittäessä aukaisemaan sitä lukijoille on mahdotonta, että kuvailutermit olisivat yksittäisiä sanoja. Esimerkkinä hän esittää kukkien käytön Hawthornen romaanissa:

Flower(s)

Black: prison as black flower of civilized society 76; of evil, will blossom as it may 92; to blossom darkly and bear black fruit betimes 235...

Pearl: a lovely and immortal flower 113; made beautiful by wearing 223; crowned and girdled with 226... (Ransley 1987a, 52).

Hazel K. Bell teki vastaavanlaisen kokeen indeksoimalla A. S. Byattin romaaneja. Hänen mielestään suurin ongelma kaunokirjallisen teoksen indeksoinnissa on sen sisällön rikkaus ja kompleksisuus – ydintermien ja käsitteiden löytäminen on hankalaa. Lisäksi teosten sisältämät symboliset ja muut tulkinnalliset seikat vaikeuttavat indeksointia. Bell korostaa myös jo Ransleyn tekemää havaintoa, että indeksointitermit eivät voi kaunokirjallisuuden kohdalla olla yksittäisiä sanoja. Ongelmaksi nousee myös indeksoijan tulkinnan syvyys – mitä indeksoija voi kertoa lukijalle ja mitä hänen tulee jättää lukijan oman löytämisen varaan⁷. Samalla indeksoijan tulee päättää kuinka suhtautua kerronnan rakenteessa mukana oleviin faktuaalisiin henkilöihin ja tapahtumiin – mikä on niiden funktio tekstissä ja mikä on niiden suhde oikeisiin, historiallisiin esikuviiin. Myös kerronnan rakenteen monikerroksisuus ja siinä käytetty näkökulmatekniikka aiheuttavat ongelmia indek-

soinnille. Esimerkiksi, jos teos käyttää usean kertojan tekniikkaa symbolien esittämisessä, tulee indeksoinnissa tämä tulla esille. (Bell 1991, 252-253.)

Bell kehittää Ransleyn esittämää kronologista termien esittämisjärjestystä korostaessaan, ettei historioissa ja elämäkertoissakin käytetty kronologinen indeksitermien esittämistapa aina riitä kaunokirjallisuudessa. Tämä sen vuoksi, että teoksessa voidaan käyttää kerrontatekniikkana mm. takautumia, jolloin teoksen tematiikan kannalta tärkeä, ajallisesti varhainen tapahtuma voi paljastua vasta teoksen lopussa⁸. Bellin mukaan tämä tapahtuman konteksti tulee näkyä indeksoinnissa (esim. "X muistaa tapahtuman y"). Lisäksi Bell korostaa sitä, että kaunokirjallisuuden indeksoinnissa merkittävän ongelman luo käytettävä termistö. Jos se on voimakkaan standardisoitua, se saattaa tyypistää indeksoinnin ilmaisuvoimaa. Lisäksi termien valinnassa pitää tehdä arvovalintoja. Nämä tulisi kuitenkin aina tehdä kunnioittaen teoksen tekijän omia tarkoituksia: "Does one list these as religion: telepathy: madness: spiritualism: occult? The author's intentions and attitudes must be divined and reflected in the vocabulary used." (Bell 1991, 254.)

Judith A. Ranta puolestaan käsittelee artikkelissaan (1991) kaunokirjallisuuden (imaginative literature) indeksointia tutkimustarkoitukseen akateemisessa kirjastoympäristössä. Hänkin käyttää esimerkkinään jo edellä mainittua N. Hawthornen teosta *Scarlet Letter* kehittäessään järjestelmän, jossa hän käyttää hyväkseen Kongressin kirjaston asiasanastoa (LCSH). Ranta korostaa aikaisempien tutkimusten pohjalta, että kaunokirjallisuuden sisällönkuvailujärjestelmän tulee olla: "multi-dimensional, giving access to abstract and concrete levels of meaning (Ranta 1991, 5)". Rantan mukaan kirjallisuudentutkijat käsittelevät nykyaikaisessa kirjallisuudentutkimuksessa kaunokirjallisuutta useiden eri arvohierarkioiden perusteella. Lisäksi kaunokirjallisuutta voidaan käyttää tutkimuskohteena myös ulkokirjallisten seikkojen vuoksi, esimerkiksi tutkittaessa eri kulttuureita. (Ranta 1991, 6.)

Ranta päätyy suosittelemaan, että:

...I agree with Ransley, Pejtersen, and Shatford that it is important to provide subject access to both the literal (or denotative) and figurative (or connotative) levels. Thus, headings should be assigned to cover historical content, themes, point of view, genre, and literary form. (Ranta 1991, 19.)

3. Kaunokirjallisuuden asiasanastojen kehitystyö

Kaunokirjallisuutta varten tehdyt asiasanastot ovat kehittyneet paljolti yksittäisten kirjastojen tarpeisiin ja/tai yksittäisten ihmisten aloitteesta. Ensimmäiset sanastot ovat olleet sanaluetteloa tai kaunokirjallisuutta varten täydennettyjä yleisiä asiasanastoja. Näiden sanastokokeilujen pohjalta on lähdetty kehittämään sanastoja, joilla mahdollistettaisiin kuvailujen jonkinasteinen yhtenevyys sekä keskitettyjen palvelujen tuottaminen. (Pulli 1992.) Yhteispohjoismaisen, paljolti BookHousen ideoille pohjaavan projektin tarkoituksena on ulottaa palvelut yhteensopiviksi eri Pohjoismaiden välillä (Pejtersen & al. 1996, 75).

Yhdysvalloissa sanastokehitys yleisellä tasolla alkoi, kun ALA:n Subject Analysis Committee julkaisi ohjeet Kongressin kirjastolle kaunokirjallisuuden indeksointia varten (Guidelines on subject access to individual works of fiction, drama etc, 1990). Ohjeissa komitea suosittelee, että kaunokirjallisuudesta tulisi kuvata seuraavat seikat: muoto/genre, henkilöt, kerronnan puitteet (setting) sekä aiheet/teemat (topics). Tämän suosituksen ja sen tueksi laaditun 23-sivuisen sanaluettelon pohjalta Yhdysvalloissa käynnistyi vuonna 1991 projekti, jossa kymmenen kirjastoa indeksoi kaunokirjallisuutta. (Pulli 1992, 12-13.) Sen lisäksi Steven Olderr on laatinut vuonna 1991 laajemman kaunokirjallisuuden asiasanaston täydentämään Kongressin kirjaston tesarusta (Young 1992, 89-94.)

Laajin Ruotsissa koottu sanasto on Eiler Jansonin ja Bo Södervallin toimittama

Tesaurus för indexering av skönlitteratur, joka ilmestyi vuonna 1987. Teos jakautuu systemaattiseen ja aakkoselliseen osaan, jossa termit on järjestetty tesaurusmaisesti. Systemaattisessa osassa termit on koottu kolmen pääfasetin alle, jotka ovat kehys (ram), henkilöt (person) sekä aihe (ämne). Nämä puolestaan jakautuvat alafasetteihin siten, että kehys jakautuu aikaan (tid) sekä tilaan (rum); henkilöt jakautuvat kehitykseen (utveckling), sosiaaliisiin suhteisiin (sociala relationer) sekä ammattiin/toimeen (yrke/ verksamhet); ja aihe jakautuu ideologiaan (ideologi), toimintaan (aktivitet), luontoon (natur) sekä ihmisruumiiseen (människokrop). Kuten toimittajat sanovat, fasettien väliset rajat ovat liukuvia ja joidenkin termien asettelu vain yhteen fasettiin on joskus väkivaltainen ratkaisu. (Jansson & Södervall 1987, 4-6.) Mainittakoon tässä vielä muista Pohjoismaista Tanskan ja Norjan BookHousen ideoiden perusteella laaditut sanastot, joiden pohjana oleva Pejtersenin kehittämä rakenne on esitelty jo aiemmin.

Eräs olennainen ongelma kaunokirjallisuuden indeksoinnin kehittämisenä on ollut, että olemassa olevat tietojärjestelmät on rakennettu voimassa olevien standardien, yleensä eri maiden MARC-standardien päälle. Tästä johtuen kaunokirjallisuuden indeksointi on jouduttu usein toteuttamaan olemassa olevien tietojärjestelmien ehdoilla, joihin indeksointia varten tarkoitettujen kenttien ovat yleensä kehitetty tietoaineiston asiansanoitusta varten. (FINMARC-formaattiin varattiin Kaunokirjallisuuden asiansanaston osakenttineen kuten Yleisen suomalaisen asiansanaston (YSA) asiansanoja varten varattua kenttää 652.)

Tämä ongelma on kaunokirjallisuuden osalla vielä merkitsevämpi kuin tietokirjallisuuden, koska kaunokirjallisuudessa indeksointi muodostuu fasetoiduksi ja sisältää loogisesti ja sisällöllisesti erilaisissa funktioissa olevia aspekteja. Tietojärjestelmissä tulisi olla mahdollisuus erottaa ja tarkentaa nämä hakuprosessin aikana. Yleensä ensimmäinen ongelma tällaisissa järjestelmissä on jo se, miten hakujoukosta voidaan erottaa mahdollisimman helposti tieto- ja kaunokirjallisuus haki-

jan sitä niin halutessa. Ongelman ratkaisuna on joko MARC-pohjaisten ratkaisujen sisällönkuvauskenttien ja hakujärjestelmien kehittäminen tai sitten kokonaan fiktiiivistä aineistoa varten tarkoitettujen tietojärjestelmien luominen. (Hyvänä esimerkkinä jälkimmäisistä on Pejtersenin kehittämä BookHouse -järjestelmä, ks. tästä esim. Pejtersen 1989, 1992, 1996).

4. Kaunokki – suomalaisen kaunokirjallisuuden tesaurus

4.1. Tausta

Suomalaisen kaunokirjallisuuden asiansanaston taustalla ovat useat yksittäiset kokeilut, joista merkittävimmät ovat Kirjavälityksen – Olli Tuuterin merkittävällä henkilökohtaisella panoksella – luoma KIRSTI tietopankki kirja-alan erityistarpeisiin sekä Erkki Salon luoma ROMA-indeksi, joka oli tarkoitettu yleisten kirjastojen käyttöön. Molemmat niistä käyttivät mahdollisuuksien mukaan YSA:n sanastoa, jonka molemmat totesivat puutteelliseksi ajatellen kaunokirjallisuuden indeksointia. (Pulli 1992, 2-4, ks. myös Kaunokki 1996, 7.) Näiden kokeilujen sekä BookHouse-konseptin mukaisten kokeilujen perusteella esille nousi keskitettyjen palveluiden tarve. Kaunokirjallisuuden indeksointi on työlästä ja aikaavievää – konkreettisesti tämän koki Erkki Salo, joka lopetti ajanpuutteen vuoksi oman toimen ohella toimitetun ROMAn ylläpidon. Lisäksi kaunokirjallisuuden indeksoinnin yleistymisen leviämisen esteenä oli, ettei sille ole olemassa sellaisia perinteitä ja käytänteitä kuin tietokirjallisuuden indeksoinnille – kaunokirjallisuuden indeksoijilta puuttui myös tärkeä väline, asiansanasto.

Sanastosta päätettiin tehdä vastaava YSA:n perheeseen kuuluva sanasto, kuten esimerkiksi Musiikin sanastosta. Sanaston toimitustyö alkoi syksyllä 1993. Vastaavan toimittajan lisäksi työtä tukemaan valittiin työryhmä, joka kävi läpi toimittajan tekemät useat

sanastoversiot. Lisäksi sanastoa testattiin käytännössä kirjastoissa, joissa oli jo aiemmin indeksoitu kaunokirjallisuutta. Teoksen kokeilupainos julkaistiin vuonna 1996. Työn rahoittajina ja organisoijina toimivat Helsingin yliopiston kirjasto sekä BTJ Kirjastopalvelu.

4.2. Rakenne

Ensimmäinen konkreettinen ongelma kaunokirjallisuuden asiasanaston luomisessa oli rakenne, jonka mukaan sanastoa ja sen sisältämää termistöä ruvettaisiin kokoamaan ja järjestämään. Sanastosta päätettiin tehdä tesaurusmuotoinen sekä järjestää tesaurus systemaattisesti niiden fasettien/ryhmien mukaan, joita työryhmä ja aiempi tutkimus olivat pitäneet olennaisina kaunokirjallisuuden kuvauksessa. Tämän lisäksi sanastoon liitettiin aakkosellinen osuus, jossa kaikki termit olivat samassa luettelossa.

Kaunokin alanmukainen osa on jaettu kuuheen ryhmään. Ne ovat:

- genrejä kuvaavat asiasanat ja niiden selitykset
- aiheita, motiiveja ja teemoja kuvaavat asiasanat
- toimijoita kuvaavat asiasanat
- miljöötä eli tapahtumapaikkaa kuvaavat asiasanat
- tapahtuma-aikaa kuvaavat asiasanat
- muita, lähinnä typografisia ja teknisiä seikkoja kuvaavat asiasanat.

Ryhmistä neljä – aiheet, toimijat, miljööt sekä tapahtuma-ajat – ovat esiintyneet lähes kaikissa aikaisemmissa kaunokirjallisuuden sisällönkuvaustutkimuksissa niinä fasetteina, jotka kertomakirjallisuudesta tulisi kuvata. Beghtol vetääkin omassa väitöskirjassaan sen johtopäätöksen, että: "Characters, Events, Spaces and Times may be taken as fundamental data categories for fiction." (Beghtol 1994a, 157).

Jos Beghtolin esittämää luetteloa vertaa Ranganathanin PMEST-fasetteihin – kuten

Shatford (1986, 49) tekee omassa kuvien sisällönkuvaukseen luodussa järjestelmässään – nähdään, että ne vastaavat Ranganathanin MEST (Matter, Energy, Space, Time) -fasetteja. Shatford tekee sen ratkaisun, että hän yhdistää Personality ja Matter -fasetit toimijoiksi ja viittaa Energy-fasetilla siihen, mitä nämä toimijat tekevät. Kaunokissa ratkaisu puolestaan tehtiin siten, että Personality-fasettia vastaa genre, joka määrää kaunokirjallisuudessa faktisestikin teoksen persoonallisuuden eli sen minkä tyyppinen teos on, teoksen tematiikan ja miljöön sekä paljolti myös kerronnan keinot ja välineet (kts. tästä tarkemmin Tenhunen 1989 ja Saarti 1996, 26-34). Matter-fasettia vastaa puolestaan Kaunokissa teemat ja Energy-fasettia toimijat. Indeksointia voisi kehittää edelleen ottamalla mukaan Ranganathanin Basic Subjectin (Ranganathan 1969, 200), jolla voitaisiin erotella lisäksi eri fiktiivisen esittämisen/julkaisemisen muodot toisistaan saman indeksointijärjestelmän sisällä.

Muut-ryhmään sijoitettiin asiasanoja, jotka kuvaavat pääasiassa tekstin sisällön ulkopuolisia seikkoja. Ne ovat kuitenkin käytännön kirjastotyössä osoittautuneet aspekteiksi, joita teoksista tulee kuvata. Tähän kuuluvat mm. jo edellä esitetyt Pejtersenin lähestyttävyys-ryhmän seikat, jotka koostuvat paljolti teoksen luettavuuteen liittyvistä asioista (Pejtersen 1983, 234). Lisäksi tähän ryhmään otettiin mm. teoksen julkaisuhistoriaan liittyviä seikkoja kuvaavia termejä.

Jo sanastoa kootessa huomattiin, että kaunokirjallisuuden sanastoissa käyttökonteksti ratkaisee hyvin paljon, minkä tyyppistä ja tasoista sanastoa tulisi valita. Konkreettisenä ongelmana oli lasten- ja nuortenkirjallisuus, joka päätettiin sisällyttää Kaunokkiin, vaikka sitä varten voisi tehdä oman sanastonsa. Toinen ongelma, joka tuli esiin työryhmässä, oli käyttöympäristö. Kaunokki päätettiin hyvin varhaisessa vaiheessa tehdä yleisten kirjastojen käyttöön. Tällöin esimerkiksi ne aspektit, joita kirjallisuudentutkimus pitää teoksissa olennaisina, päätettiin jättää sanaston ulkopuolelle, ellei niitä ollut jostakin erityisyydestä perusteltua ottaa mukaan. Toisaalta tämä ongelma ratkeaa työryhmän kirjallisuudentutkijajäsen Lasse Koskelan

laatiman kirjallisuudentutkimuksen asia-sanaston myötä. Kolmas ongelma oli sanaston termien luonne. Tässä työryhmä päätyi siihen, että termien tulee olla yleiskielisiä ja mahdollisimman yksiselitteisiä. Lisäksi genreistä annettiin selitykset, koska niiden kohdalla sanasto on kirjavaa ja vakiintumatonta. Sisällönkuvauksen tarkkuus voitiin ilmaisuvoiman vaateen, jota fiktiivisen aineiston kuvailulta myös kaivataan.

Osa aspekteista, jotka sisältyvät kirjallisuudentutkimuksen sanastoon voisivat olla relevantteja myös osalle yleisen kirjaston käyttäjäkuntaa. Tämä tulee esille Nielsenin esittämästä kriittisestä. Hänen mukaansa perinteistä, faktoihin sitoutunutta kaunokirjallisuuden indeksointia tulisi laajentaa myös temaattisiin seikkoihin paneutuvalla kuvailulla sekä kerronnan rakenteita ja tekniikkoja paljastavalla kuvailulla. Tämä sen vuoksi, että modernistisessa ja postmodernissa kerronnassa olennaista ei usein ole se, mitä kerrotaan vaan se, miten kerrotaan. Kolmas asia, joka Nielsenin mukaan voitaisiin ottaa mukaan kaunokirjallisuuden indeksointiin, on teoksen ja tekijän saamat kulttuurihistorialliset vaikutteet, joihin liittyviä asiansanoja Kaunokkiin otettiin jonkin verran mukaan. (Nielsen 1997, 171-180.)

5. Päätäntö

Kaunokin kokeiluversio on nyt ollut pari vuotta käytössä, ja sen myötä on jo saatu palautetta sanastosta sekä sen kehittämistarpeista. Lisäksi teoksesta on laadittu ruotsinkielinen versio Bella (1997), joka antaa mahdollisuuden kuvailujen ja sanastojen vertailuun kahden eri kielen ja kulttuurin välillä. BTJ Kirjastopalvelu on laajentanut palvelujaan kaunokirjallisuuden indeksointiin, ja näin saadaan vähitellen kokemusperäistä tietoa kaunokirjallisuuden keskittystä kuvailusta, sen ongelmista ja eduista.

Jatkotutkimuksen ja -selvittelyn aiheita on lukuisia muitakin. Eräs olennainen on kaunokirjallisuuden indeksoinnin riippuvuus indeksointiympäristöstä, joka jo vähäisen

kokemuksen perusteella näyttäisi olevan ainakin yhtä komplisoitua kuin tietokirjallisuuden indeksoinnin riippuvuus indeksointiympäristöstään. Tämä on kriittinen tekijä myös keskitettyjä palveluja ajatellen. Siihen liittyy myös ns. demokraattisen indeksoinnin mallin testaus kirjastoympäristössä.

Tärkein jatkotutkimuksen ja kehitystyön aihe on kaunokirjallisuuden tietojärjestelmien luominen. Eräänä mallina näissä voisivat olla kaupalliset ratkaisut eri fiktiivisen aineiston aloilla (esim. Cinemania) sekä kaunokirjallisuutta varten kehitetyt mallit, joista tunnetuin on BookHouse.

Hyväksytty julkaistavaksi 22.8.1998

Lähteet

- Andersson, Rolf & Holst, Erik (1996). Indexes and other depictions of fiction: a new model for analysis empirically tested. Swedish Library Research/Svensk biblioteksforskning (2-3):77-95.
- Beghtol, Clare (1992). Toward a theory of fiction analysis for information storage and retrieval. Teoksessa: Classification research for knowledge representation and organization. Editors N. J. Williamson & M. Hudson. Amsterdam: Elsevier.
- Beghtol, Clare (1994a). The classification of fiction: the development of a system based on theoretical principles. Metuchen: Scarecrow Press.
- Beghtol, Clare (1994b). Domain analysis, literary warrant, and consensus: the case of fiction studies. Journal of the American Society for Information Science 46 (1):30-44.
- Beghtol, Clare (1997). Stories: applications of narrative discourse analysis to issues in information storage and retrieval. Knowledge Organization 24 (2):64-71.
- Bell, Hazel K. (1991). Indexing fiction a story of complexity. The Indexer 17 (4):251-256.
- Eagleton, Terry (1991). Kirjallisuusteoria: joh-

- datas. Suomentaneet: Yrjö Hosiaisuus ym. Helsinki: Vastapaino.
- Green, Rebecca (1997). The role of relational structures in indexing for the humanities. *Information Services & Use* 17 (2-3):85-100.
- Hidderley, Rob & Rafferty, Pauline (1997). Democratic indexing: an approach to the retrieval of fiction. *Information Services & Use* 17 (2-3):101-109.
- Iivonen, Mirja (1985). Kaunokirjallisuuden kirjastoluokituksesta. Tampere: Tampereen yliopisto. (Tampereen yliopiston kirjastotieteen ja informatiikan laitoksen tutkimuksia, 20.)
- Jansson, Eiler & Södervall, Bo (1987). *Tesaurus för indexering av skönlitteratur*. Borås: Högskolan i Borås. (Högskolan i Borås. Institutionen Bibliotekshögskolan. Specialarbete, 1987:7.)
- Juntunen, Arja & Saarti, Jarmo (1992). Kaunokirjallisuuden sisällönkuvaus yleisissä kirjastoissa. Kirjastotieteen- ja informatiikan pro gradu -tutkielma. Oulu: Oulun yliopisto. Kaunokki: kaunokirjallisuuden asiansaasto (1996). Vastaava toimittaja Jarmo Saarti. Helsinki: BTJ Kirjastopalvelu.
- Kivimäki, Sanna (1992). Takakansi romaanin tuoteselosteena: kotimaisten romaanien takakansitekstit vuosina 1930-1989. *Kulttuuritutkimus* 9 (4):12-23, 48.
- Macpherson, Ruby (1987). Children's literature indexes at Moray House. *Library Review* 4:254-260.
- Nielsen, Hans Jørn (1997). The nature of fiction and its significance for classification and indexing. *Information Services & Use* 17 (2-3):171-182.
- Pejtersen, Annelise Mark (1989). The book house: modelling users' needs and search strategies as a basis for system design. Risø: Risø national laboratory.
- Pejtersen, Annelise Mark & Villumsen, Steffen & Weber, Steen (1992). Rapport om førsog med Boghuset i skolebibliotek. Risø: Forsknincenter Risø.
- Pejtersen, Annelise Mark & Albrechtsen, Hanne & Lundgren, Lena & Sandelin, Ringa & Valtonen, Riitta (1996). Subject access to Scandinavian fiction literature: index methods and OPAC development. Copenhagen: Nordic Council of Ministers. (Tema-Nord: Culture, 608).
- Pejtersen, Annelise Mark & Austin, Jutta (1983). Fiction retrieval: experimental design and evaluation of a search system based on users' value criteria: part 1. *Journal of Documentation* 39 (4):230-246.
- Pulli, Riitta (1992). Kaunokirjallisuuden keskitetty indeksointi Suomessa. Helsinki: Helsingin yliopiston kirjasto.
- Ranganathan, S. R. (1969). Colon classification library edition 7 (1971): a preview. *Science with a Slant to Documentation*. 6 (3) : 193-242.
- Ransley, Ambrose (1987a). Towards a fiction index: part one. *The Australian Library Journal* 36(1): 44-54.
- Ransley, Ambrose (1987b). Towards a fiction index: part two. *The Australian Library Journal* 36(2): 103-113.
- Ranta, Judith A. (1991). The new literary scholarship and a basis for increased subject catalog access to imaginative literature. *Cataloging & Classification Quarterly* 14 (1):3-27.
- Rich, Elaine (1979). User modeling via stereotypes. *Cognitive Science* (3):329-354.
- Rich, Elaine (1986). Users as individuals: individualizing user models. Teoksessa: R. Davies (editor). *Intelligent information systems: progress and prospects*. Chichester: Ellis Horwood.
- Saarti, Jarmo (1996). Romaanit hukassa?: kertomakirjallisuuden hyllyluokituksen vaikutus kirjastonkäyttöön ja kertomakirjallisuuden lainaukseen. Helsinki: BTJ Kirjastopalvelu.
- Shatford, Sara (1986). Analyzing the subject of a picture: a theoretical approach. *Cataloging & Classification Quarterly* 6 (3):39-62.
- Tenhunen, Iris (1989). Genre teoria kirjallisuudenhistorian näkökulmasta. Teoksessa: Kirjallisuushistoria tänään. Helsinki: SKS. (Kirjallisuudentutkijain Seuran vuosikirja, 43.)
- Vainionpää, Marja-Leena (1974). Kirjallisuustieteen perusteita. Porvoo: WSOY.
- Wellek, René & Warren, Austin (1980). *Theory of literature*. (Rep.) Harmondsworth: Penguin.
- Young, J. Bradford (1992). Olderr's fiction subject headings: a supplement and guide to the LC thesaurus. (Kirja-arvostelu.) .

Cataloging & Classification Quarterly 15 (1):89-94.

Viitteet

1. Osa kaunokirjallisuuden luokitustutkimuksesta voidaan lukea myös tähän ryhmään. Näin etenkin Annelise Mark Pejtersenin työ. Hänen luokitusysteeminsä on itseasiassa indeksoinnin ja ryhmittelyn hybridi.
2. Kts. esim. Eagleton 1991, 26-66. Vastaava kasvatuksellis-ideologinen perinne on liittynyt tietysti yleisen kirjaston tehtävään ja heijastunut myös siten kaunokirjallisen aineiston valintaan. Vrt. Saarti 1996, 22-24.
3. On mielenkiintoista, miten nämä kategoriat vastaavat lähes yksi yhteen tyypillisimpien romaanikategorioiden luetteloa runousopeissa. Esimerkiksi Vainionpää (1974, 102–103) jakaa romaanit seuraaviin kategorioihin: -toiminta- eli tapahtumaromaanit, juoniromaanit
-henkilöromaanit, luonneromaanit
-elämänpiiri- eli raumromaanit.
4. Tämän on Pejtersen huomannut myös itse ja myöhemmissä versioissa tätä kohtaa on selvennetty vähemmän tulkinnanvaraiseksi.
5. Hyvänä esimerkkinä on tekijästä kuvailtavat seikat. Kirjastossa, jossa kaunokirjallisuudesta etsivät tietoa esitelmää tekevät koululaiset, on käyttäjien kannalta hyvin relevanttia, että teosten tekijöistä löytyy myös tietoa. Se onko tämä tieto teosten indeksoinneissa vai jotenkin muuten saatavissa on tietysti asia, joka vaatii selvittämistä.
6. Teknisesti tällaisen pitkän kuvailun voikin käsittää yhden asiakokonaisuuden kuvailuksi, esimerkiksi viittaus synnin käsitteeseen romaanissa, jossa synti on ydinsymboli ja ydinteema. Vasta koko teoksen lukeminen paljastaa kokonaismerkityksen, jonka teema teoksessa saa.
7. Karkeana esimerkkinä voisi esittää salapolii-siromaanin henkilöhakemiston, joka paljastaa lukijalle murhaajan.
8. Venäläiset formalistit käyttivät tästä jaosta käsiteparia *sujet-fable*. Fable on temporaalis-kausallinen tarina, joka on kerronnan pohjana, ideaalisesti aina samana. *Sujet* on puolestaan tarina kerrottuna eli kerronnan narratiivinen rakenne. (Wellek 1980, 218.) Kaunokirjallisuudessa tätä käytetään hyväksi myös kerrontatekniikkana esimerkiksi siten, että sama tarina kerrotaan romaanin sisällä eri kertojanäkökulmista. Vrt. myös esimerkiksi Faustin tarina ja siitä kirjoitetut kirjat.