

Leksikografiaa Leeuwardenissa

Euralex eli European Association for Lexicography kokoaa joka toinen vuosi eurooppalaisia ja muualtakin tulleita leksikografeja suureen kongressiin. Vuonna 2010, pahimman (tai parhaimman) hollantilaisen jalkapallohuuman aikaan 6.–10. heinäkuuta, kongressi pidettiin Alankomaissa, Friisinmaan pääkaupungissa Leeuwardenissa, ja sen järjestivät Fryske Akademy ja Leidenissa toimiva Istituut vor Nederlandse Lexicologie. Viisipäiväisessä tapahtumassa osanottajia oli noin 250 ja erityisteemana oli – friisin kielen puhuma-alueella kun oltiin – Euroopan vähemmistökielien (*European non-state or lesser used languages*) leksikografia. Kongressin ensimmäisenä päivänä kuningatar Beatrix piipahti vihkimässä käyttöön sähköisen version 25-osaisesta friisin kielen sanakirjasta.

Friisin kielen lisäksi Euroopan vähemmistökielistä kongressissa oli esitelmää mm. katalaanin, baskin, alatorbin, Skotlannissa puhuttavien vähemmistökielten gaelin ja skotin, friulin, Itävallan Burgenlandissa puhuttavan romanin ja viittomakielten sanakirjoista. Esitelmää oli myös joidenkin Euroopan ulkopuolisten kielten, mm. Etelä-Afrikassa puhuttavien afrikkalaisten kielten leksikografiasta.

Erittäin kiinnostavan plenaariesitelmän Euroopan vähemmistökielten leksikografian tilasta piti Anne Tjerk Popkema. Hän totesi ensinnäkin, että leksikografeilla on tärkeä osuus kielen elämissä. Kielen kodifikaatio ja kehitystyö ovat keskeisiä tehtäviä, kun vähemmistökielestä halutaan tehdä nykyaikainen käyttökieli, kun kieli halutaan ”emansi-

poida”. Nämä tehtävät kuuluvat leksikografian työhön. Popkema totesi, että leksikografilla on tärkeä tehtävä myös kieltä uhkaavassa kuolemassa. Kuoleva tai uhanalainen kieli on dokumentoitava sanakirjan muodossa. Kaikki kielet sisältävät vuosisatojen, kenties vuosituhansien ajalta kulttuurista ja ekologista tietoa, joka saatetaan kokonaan menettää, jos sitä ei dokumentoida kunnolla. Popkema siteerasi Unescon kieliasiantuntijoiden lausuntoa vuodelta 2003, jossa todetaan, että dokumentaatio on tärkeää, koska se rikastuttaa inhimillistä älyllistä omaisuutta ja tuo esiin sellaisen kulttuurisen perspektiivin, joka saattaa olla aivan uusi nykyiselle tietämyksellemme. Dokumentaatioprosessi usein myös auttaa elvyttämään kielellistä ja kulttuurista tietoa.

Pääosan Popkeman esitelmästä muodosti katsaus tuloksiin, jotka saatiin Euralexin järjestämästä kyselystä, jonka avulla haluttiin arvioida vähemmistökielten leksikografian tilaa. Kysely oli lähetetty kymmenille eurooppalaisille instituutiolle ja henkilöille, jotka ovat osallisina vähemmistökielten sanakirjojen teossa. Vähemmistökieliksi katsottavia kieliä on Euroopassa n. 60, ja niitä puhuu n. 55 miljoonaa ihmistä. Islantia lukuun ottamatta jokaisessa Euroopan maassa on ainakin yksi vähemmistökieli.

Kyselyn avulla hankittiin tietoa ensinnäkin kielten sosiolingvistiksesta asemasta (puhujien määrä, valtion antama ”tunnustus”, uhanalaisuus), toiseksi leksikografisesta infrastruktuurista (valtion antama tuki leksikografiseen työhön, tietotekniikka) ja kolmanneksi leksikogra-

fisten tuotosten ”moninaisuudesta”. Tätä moninaisuutta arvioitiin sen perusteella, onko kielestä olemassa yksikielisiä sanakirjoja, kaksikielisiä sanakirjoja (valtakielestä vähemmistökieleen, vähemmistökielestä valtakieleen), verkkosanakirjoja (sekä yksi- että kaksikielisiä), eriasteisia koulusanakirjoja, sanastoja ja metaleksikografista kirjallisuutta.

Kyselyn kolmen osa-alueen tuloksia vertailemalla Popkema oli tehnyt monenlaisia mielenkiintoisia havaintoja. Esimerkiksi Iberian niemimaan leksikografia vaikuttaa olevan hyvin kehittyntä: sekä leksikografisessa ”moninaisuudessa” että modernin teknologian käytössä iberialaiset kielet kuuluvat kyselyn kärkeen. Joidenkin kielten saamat alhaiset pisteet leksikografisessa moninaisuudessa voivat johtua voimakkaasta dialektaalista differentaatiosta, joka saattaa estää kirjoitetun standardikielen luomisen. Tällainen standardikieli on kuitenkin tärkeä tekijä kielen emansipaatiossa.

Sektioesitelmiä ja ohjelmistoesittelyitä kongressissa oli kaiken kaikkiaan noin 130. Vähemmistökielten lisäksi sektioiden aiheita olivat mm. tietokoneleksikografia ja -leksikologia (joka oli esitelmien määrällä mitattuna suosituin sektio), historiallinen ja tieteellinen (*scholarly*) leksikografia ja etymologia, kaksikielinen leksikografia, fraseologia ja kollokaatiot, leksikologiset aiheet, joilla on leksikografista relevanssia, sekä raportit leksikografista ja leksikologisista projekteista.

Useissa esitelmissä käsiteltiin korpuksen pohjalta sanakirjatyötä varten luotuja laajoja leksikaalisia tietokantoja. Tällainen tietokanta on mm. englantilainen DANTE (Database of Analyzed Texts of English), jota tekijät kuvaavat lingvistisin periaattein tehdyksi systemaattiseksi ja perusteelliseksi sanaston leksikogra-

fiseksi kuvaukseksi. Ranskan kielestä on tekeillä Definiens-niminen semanttinen tietokanta, jossa yhtenä tavoitteena on jakaa *Trésor de la Langue Francaise informatisé* -sanakirjan sisältämät määritelmät komponenteiksi, merkitä eksplisiittisesti määritelmässä esiintyvät yläkäsitteet (*genus*) ja erityspiirteet (*differentiae*).

Esillä oli myös kansallisia WordNet-hankkeita, jotka perustuvat relationaaliin näkemykseen sanaston rakenteesta ja jotka eksplisiittisesti pyrkivät tuomaan esille sanojen väliset suhteet, sekä FrameNet-hankkeita, joiden pohjana on Charles Fillmoren kehyssemantiikka. Tanskalaisessa WordNetissä tavanomaiseen wordnet-tietoon on lisätty tietoa sanojen konnotaatioista ja James Pustejovskyn generatiivisen leksikon teorian mukaisista *qualia*-rooleista. Alankomaissa kehitetään leksikaalis-semanttista Cornetto-tietokantaa, jossa yhdistyvät WordNet ja FrameNetin tyyppinen sanojen semanttisia kehyksiä koskeva tieto.

Kongressissa esiteltiin myös eri maissa tekeillä olevia sanakirjaportaaleja, joissa yhdistetään joukko sähköisiä sanakirjoja siten, että niistä kaikista voidaan etsiä tietoa yhteisillä hauilla. Tällainen portaali on mm. saksalaisen Institut für Deutsche Sprachen OWID (*Online-Wortschatz-Informationssystem Deutsch*). Alankomaissa Instituut voor Nederlandse lexicologie kokoaa parhaillaan GiGaNt-nimistä tietokantaa, joka tulee kattamaan 16 vuosisataa hollannin kieltä. Tähän tietokantaan yhdistetään sekä sanakirjoja (mm. kolme eri aikakausilta tehtyä vanhan kirjahollannin sanakirjaa) että korpuksia, joten esimerkiksi korpuksen käyttäjä voi välillä klikata etsimään tietyn sanan merkitystä sanakirjoista.

Alankomaat lienee yksi eurooppalainen edelläkävijämaa leksikografian alalla,

ja edellä mainittujen lisäksi kongressissa oli esillä paljon muitakin kiinnostavia hollantilaisia sanakirjoja, sanakirjahankkeita ja tutkimuksia. Yksi kiinnostava hanke on aloitusvaiheessa oleva hollannin murteiden sanakirja, johon kootaan suuri joukko maallikkojen tekemiä, aakkosellisesti järjestettyjä murre-sanakirjoja ja jonka tiedot tulevat täydentämään olemassa olevia *Wörter und sachen* -ideologian pohjalta onomasiologisin periaattein aikoinaan koottuja suuria Brabantin, Limburgin ja Flandersin alueen murre-sanakirjoja. Esillä oli myös tietävästi maailman suurin sanakirja, 43-osainen historiallinen *Woordenboek der Nederlandsche Taal*, jota tehtiin lähes 150 vuotta. Sanakirja on nyt päässyt siihen vaiheeseen, että se on ilmaisena sähköisenä versiona kaikkien kiinnostuneiden käytössä. Parhaillaan hakusanoihin yhdistetään ”ulkoista tietoa”: kielikarttoja, kuvia ja etymologista tietoa.

Hollantilaisen mammuttiteoksen lisäksi kongressissa esiteltiin muitakin kiinnostavia historiallisia sanakirjoja, joista mainittakoon tekeillä olevat skotlantilainen gaelin kielen *Faclair na Gàidhlig* ja saksalainen *Deutsches Rechtswörterbuch*, joka kuvaa saksalaisten lakitekstien koko sanaston 600-luvulta 1800-luvulle. Historiallisista sanakirjoista sekä erilaisista sanakirjoja yhdistävistä portaaleista ja tietokannoista esitelmissä yleisesti ennakoitiin tulevan tärkeitä väli-nejä mm. sanaston (historiallisessa) tutkimuksessa ja historian tutkimuksessa. Gaelin kielen sanakirjan tekijät totesivat, että korpuksinkin tarvitsevat rinnalleen sanakirjan, joka selittää, mitä ”aarteita” korpuksiin sisältyy.

Kongressissa oli runsaasti esitelmiä myös erilaisista leksikografian tutkimuksista. Monissa näistä todettiin sanakir-

jojen sisältävän erilaisten kielen ilmiöiden esittämisessä paljon epäjohdonmukaisuutta. Useissa esitelmissä päädyttiin siihen, että sanakirjahankkeen alkuvaiheessa on luotava ilmiön kuvausta varten malliartikkeli (*template entry*), jonka mukaan kaikki vastaavalla tavalla käyttäytävistä sanoista tehdyt sanakirja-artikkelit voidaan rakentaa.

Hollantilainen Josefien Sweep käsiteli esitelmässään verbien ”metonymisia objektin vaihdoksia”, joilla hän tarkoittaa vaihtelutyyppejä *to clear the plates / the table, to light wood / fire / the fireplace, to press oranges / juice*. Sweep oli tutkinut sitä, miten kaksi suurta hollannin sanakirjaa tulkitsevat näiden vaihteluiden vaikuttavan verbin merkitykseen. Kumpikin sanakirja (ainakin periaatteessa) tuo objektin muutoksen eksplisiittisesti ilmi (*objectsverwisseling*), mutta osassa tapauksista sanakirjat tulkitsivat verbin merkityksen tai alamerkituksen vaihtuneen, osassa taas eivät. Sweep tuli siihen tulokseen, että eri objektien avulla tilanteeseen otetaan erilainen perspektiivi mutta että verbin merkitys ei muutu.

Metonymia-aihe oli myös Sylwia Wojciechowskalla, joka tarkasteli sitä, miten viisi yksikielistä englannin oppijansanakirjaa esittävät tiettyjen sanojen metonymiset merkitykset. Systemaattisen polysemian nimellä substantiivien metonymisia merkityksiä käsitteli posterissaan virolainen Margit Langemets. Langemetsin tutkimuksen tarkoituksena oli löytää periaatteet, joiden mukaan substantiivien metonymia voidaan kuvata Eesti keele instituutin uudessa sanakirjassa. Tätä varten Langemets on tarkastellut sitä, miten metonymia esitetään jo ilmestyneessä yksikielisessä viron sanakirjassa. Taustateorianä hänen(kin) tutkimuksessaan on Pustejovskyn genera-

tiivisen leksikon teoria. Mielenkiintoinen sattuma on se, että sekä Langemets että puolalainen Wojciechowska väittivät metonymia-aiheestaan vuonna 2009.

Piet Swanepoel puhui hauskaista aiheesta, MONSTER-kategorian määrittelemisestä semanttisten piirteiden, narratiivisten kategorioiden ja kehysten (Fillmoren *frame*, Lakoffin *idealized cognitive model*) avulla. Swanepoel tarkasteli sitä, miten eri englanninkielisissä sanakirjoissa määritellään toisaalta kokokategoriaan viittaava sana *monster* ja toisaalta sellaiset kategorian jäsenet – tai Swanepoelin mielestä kategoriaan kuuluvat sanat – kuin *centaur*, *cyclops*, *vampire* jne. Swanepoelin lähtökohta oli Fillmoren ajatus, että sanojen merkitysten ymmärtämiseksi tarvitaan tietoa niiden kehystistä, toisin sanoen niiden taustana olevista ensyklopedisista tietorakenteista tai käsitteistyksistä. MONSTER-kategoriaan kuuluvien sanojen määritelmiin pitäisi Swanepoelin näkemyksen mukaan sisältyä tieto yhteisestä kehyksestä, mutta hänen sanakirja-aineistossaan tätä kehystä ei ollut.

Esimerkkeinä muista esillä olleista leksikografian tutkimuksista mainittakoon Barbara Bertin tutkimus kuudesta tunneadjektiivista ja niiden kollokaateista kolmessa englanti–italia-sanakirjassa, Stephen Coffeyn tutkimus offensiivisista sanoista viidessä yksikielisessä

opiskelusanakirjassa, Chris Mulhaalin tutkimus idiomien hakusanoituksesta viidessä italia–englanti-sanakirjassa, Lukasz Pakulan tutkimus värien määrittelemisestä englantilaisessa, japanilaisessa ja puolalaisessa leksikografiassa sekä Marijke Mooijaartin tutkimus hollannin *ne*-kieltopartikkelista kirjoitetuista sanakirja-artikkeleista hollannin historiallisissa sanakirjoissa. Mooijaart oli tullut siihen tulokseen, että *ne*-partikkelin kaltaisten funktiosanojen kuvaus sanakirjassa tarvitsee pohjakseen kielitieteellisiä tutkimuksia hollannin kielen varhaisempien vaiheiden kieltorakenteista. Funktiosanojen kuvauksesta puhui myös Erik Hoekstra. Hän piti sanakirjojen ongelmana sitä, että *ne* yrittävät funktiosanoillekin antaa parafraasin, vaikka näistä sanoista pikemminkin pitäisi esittää syntaktista tietoa.

Sympaattisen Leeuwardenin ilmastoidussa World Trade Centerissä ei päiväpäivältä voimistunut kuumuus haitannut. Rakennuksen ylimmässä kerroksessa, hollantilainen lakeus ikkunoiden takana, syötiin juhlaillallista, jonka yhtenä ohjelmanumerona kolme muusikkoa esitti Leonard Cohenin musiikkia – tietysti friisin kielellä. Suomalainen leksikografi tunsikin kuuluvansa eurooppalaiseen perheeseen.

NINA KAMPPI
etunimi.sukunimi@kotus.fi