

Matemaattista kielitiedettä ja kieliteknologiaa

Általános nyelvészeti tanulmányok II. A matematikai nyelvészet és a gépi fordítás kérdései. Szerkesztő KALMÁR LÁSZLÓ és TELEGDÍ ZSIGMOND. Akadémiai Kiadó. Budapest 1964. 322 s.

Maaliskuussa 1962 järjesti Unkarin tiedeakatemia Budapestissa konferenssin, jossa joukko matemaatikkoja ja kielentutkijoita käsitteli matemaattisen kielitieteen ja koneella kääntämisen ongelmia. *Általános nyelvészeti tanulmányok II:n* artikkelit perustuvat konferenssissa pidettyihin esitelmiin ja siellä käytettyihin puheenvuoroihin.

Kirja jakautuu kahteen osastoon. Alkupuolella käsitellään matemaattisen lingvistiikan ongelmia, loppupuoli on omistettu koneella kääntämiselle ja tätä alaa sivuaville kysymyksille.

Toimittajain alkusanojen ja Gyula Némethin avausesitelmän jälkeen ensimmäisenä varsinaisena artikkelina on László Kalmárin laaja, 64-sivuinen tutkielma »Matemaattiset ja kielelliset struktuurit». Kirjoittaja, joka tarkastelee asioita lähinnä matemaatikon näkökulmasta, on saanut esityksensä hyvin mielenkiintoiseksi ja nimenomaan lingvistille antoisaksi. Kielentutkijoita tuskin voidaan yleensä syyttää oman alansa ja naapuritieteiden välisten raja-alueiden kartoittamisen laiminlyönnistä. Näkemys on vanhastaan pyritty avarta-

maan kirjallisuuden- ja kansanrunouden tutkimuksen, kansatieteen, psykologian, sosiologian ja muidenkin lähinnä humanististen tieteiden suuntiin. Mutta lingvistiikan ja matematiikan yhteisalueeseen on kohdistettu huomiota vasta aivan viime vuosina, mikä onkin luonnollista, kun naapuruudesta ei ole oltu edes selvillä aikaisemmin. Kalmár toteaa, että matemaattinen kielentutkimus jakautuu ainakin kolmeen haaraan, jotka eroavat toisistaan mm. sen mukaan, kuinka laaja ja syvä on matematiikan osuus. Vanhin näistä haaroista on epäilemättä tilastollinen kielentutkimus. Yksinkertaista tilastolaskentaa on käytetty kielitieteen apuna ainakin 1700-luvulta lähtien, mutta tilastomatematiikan täysitehoinen hyödyntäminen kuuluu vielä tulevaisuuteen. Toinen, vasta aivan viime vuosina versonut haara on elektronisten tietokoneiden soveltaminen kielitieteeseen. Sovelluksista eniten huomiota on saanut osakseen koneella kääntäminen. Nämä matemaattisen kielentutkimuksen kaksi lohkoa eivät nähdäkseni vielä selvästi sijaitse kielitieteen ja matematiikan yhteisalueella. Niissä matema-

tiikka on alistettu tekniseksi apuneuvoksi, jonka on määrä vain edistää ja helpottaa puhtaasti kielitieteellisten ongelmien ratkaisua. Kokonaan toisella tasolla on kolmas lohko, jonka alueella vasta voi sanoa kielitieteen ja matematiikan kohtaavan toisensa tieteinä. Tätä periaatteellisesti kiintoisinta, mutta samalla vaateliainta matemaattisen kielentutkimuksen alaa Kalmár varsinaisesti käsittelee. On kysymys lähinnä kielen struktuurin tarkastelusta matemaattisten teoriain valossa. Vaateliasta tällainen tarkastelu on sikäli, että siihen tarvitaan molempien alojen perusteellista hallintaa. Kalmár huomauttaa, ettei matematiikka tutki pelkästään kvantiteetteja, vaan että monet matematiikan haarat, kuten abstrakti algebra, topologia, joukko-oppi ja matemaattinen logiikka käsittelevät myös mielivaltaisista elementeistä koostuvien järjestelmien rakennetta, jolloin myös kvalitatiivisiin näkökohtiin kiinnitetään huomiota. Juuri näitä aloja voidaan soveltaa kielentutkimukseen ehkä hyvinkin menestyksellisesti. Vaikeutena on lingvistin kannalta se, etteivät mainitut matematiikan haarat suinkaan kuulu matemaattiseen yleisivistykseen. Matemaatikostakin yritys saattaa näyttää ylivoimaiselta. Kalmár esittää useitakin mielenkiintoisia huomioita erinäisten kielellisten ja matemaattisten struktuurien analogisuudesta, mutta joutuu parissakin yhteydessä toteamaan, että elävä kieli on järjestelmänä monin verroin monimutkaisempi kuin mikään matematiikan koskaan tutkima järjestelmä.

Kokoelman muut artikkelit eivät ole niin laajoja eivätkä niin yleisluonteisia kuin Kalmárin esitys, joka sopii hyvin periaatteellisesti valaisevaksi johdannoksi koko kirjalle. Muissa kirjoituksissa tarkastellaan pikemminkin erikoiskysymyksiä. Ferenc Papp käsittelee kielen järjestelmää, kommunikaatioprosessia ja eräitä niiden matemaattisia malleja. Ferenc Kiefer, joka tutkielmasarjan ensimmäisessä osassa kirjoitti joukko-opin

soveltamisesta kielitieteeseen O. S. Kulagin mukaan, jatkaa nyt samasta aiheesta. Kulagin malli esitellään lukijalle melko yksityiskohtaisesti parillakymmenellä sivulla. Lisäksi Kiefer tarkastelee lyhyesti matemaattis-loogisen mallin tarjoamia mahdollisuuksia ja koskettelee siinä yhteydessä mm. J. Lambekin syntaktisten tyyppien järjestelmää.

Tilastollista tutkimusta edustaa kaksi artikkelia. Bálint Dömölki, Iván Fónagy ja Tamás Szende ovat suorittaneet äännetilastollisia tutkimuksia unkarin yleiskielestä ja laatineet niistä yhteisen raportin. Työ on suoritettu tietokoneen avulla, ja aineistona on käytetty Lőrinc Szabón ja Gyula Juhászín tuotantoa. Artikkeleissa selostetaan työn eri vaiheita, kuten esivalmisteluja ja teknistä toteutusta, sekä esitellään tutkimustuloksia. László Kozma puolestaan esittelee lyhyehkössä kirjoituksessaan varsin näppäräntuntuisen kielitilastollisen automaatin, koneen, joka seitsemässä tunnissa kykenee käsittelemään tilastollisesti kahden arkin laajuisen tekstin noin kuudeltakymmeneltä eri näkökannalta. Kozma arvioi saman työn vievän ihmisvoimin suoritettuna satakertaisen ajan.

Koneella kääntämisen ongelmat tuntuvat kovin askarruttavan monien unkarilaisten tutkijain mieltä. György Hell kirjoittaa venäjänkielisen tekstin koneelliseen unkarintamiseen liittyvistä periaatteellisista ja käytännöllisistä kysymyksistä, Sándor Kónyi taas koskettelee eräitä unkarista venäjään tapahtuvan koneellisen kääntämisen ongelmia. Ennen kuin koneella kääntämistä käytännössä voidaan ajatellakaan, tarvitaan suunnaton määrä valmistavaa tutkimustyötä. Päämääränä on löytää järjestelmä, joka tekisi mahdolliseksi kuvata kieltä mekanistisesti, koneen vaatimalla tavalla. Tämä pyrkimys on viime aikoina selvästi vaikuttanut erällä taholla koko lingvistiseen ajatteluun. Sielläkin, missä päämääränä ei ole kieliaineksen koneellinen käsittely, pidetään

usein mekanistista kielenkuvausta metodisena ihanteena. Siitä kuinka pitkälle tässä pyrkimyksessä saatetaan mennä, sopii näytteeksi Ferenc Hadasin toimittamastaan venäläis-unkarilaisesta kielio pillisestä sanakirjasta kirjoittama lyhyt esittely. Sanakirjassa on lueteltu aakkosjärjestyksessä kaikki venäjän kielio pilliset suffiksit ja suffiksiryhmät sekä annettu niiden unkarilaiset vastineet. Tekijä uskoo tällaisesta sanakirjasta olevan hyötyä venäjän kielen opetuksessa ja kääntämisessä. Lisäksi hän arvelee sen auttavan henkilöitä, jotka haluavat seurata venäjänkielistä ammattikirjallisuutta, mutta eivät tyydyttävästi tunne kieltä. Voin kuvitella suffiksisanakirjasta olevan monenlaista hyötyä kielentutkimukselle, mutta sen merkitystä Hadasin mainitsemilla käytännön aloilla rohkenen epäillä. Tekijä ei yritäkään salata sitä, että hän sanakirjansa käyttömahdollisuuksia kaavaillessaan on pitänyt esikuvanaan käännöskonetta. On kuitenkin kyberneettisesti nurinkurista kuvitella, että koneella kääntämisen äärimmäisen monimutkaiset menetelmät sopivat myös ihmisivoille.

György Szépe käsittelee artikkelissaan eräitä koneella kääntämisen kysymyksiä nimenomaan Unkarin kannalta. Ensin hän luo katsauksen alan ulkomaisiin koulukuntiin ja unkarilaisiin yrityksiin, siirtyy sen jälkeen koneella kääntämisen periaatteellisiin kysymyksiin, käsittelee koneella kääntämisen ja kielitieteen suhdetta ja lopuksi pohtii, mistä kielistä ja minkä alan erikoiskirjallisuutta olisi taroituksenmukaisinta kääntää koneella unkariksi. Nykyään käännetään eniten venäjistä. Tähän siis kipeimmin tarvittaisiin koneapua. Seuraavat lähdekielet ovat tärkeysjärjestyksessä saksa, englanti, tšekki, ranska ja puola. Koska ainakin aluksi on pakko rajoittaa koneellistamaan vain joidenkin erikoisalojen kielen kääntäminen, on aiheellista tutkia, millä aloilla tarve on suurin. Szépe luetlee eräitä keskeisimpiä aloja, jotka näyttävät kuuluvan lähinnä teollisuuden, tek-

niikan, kansantalouden ja maatalouden piiriin.

Gyözö Sipőczy on ottanut tarkastelakseen unkarin — ja miksei suomenkin — kannalta tärkeän ongelmaryhmää, nimittäin prepositioilmausten ohjelmoinnin. On kysymys mm. siitä, kuinka kone voi erottaa agentin ja paikanadverbiaalin, jos on käännettävä esimerkiksi englanninkieliset lauseet *he was kissed by the girl* ja *he was kissed by the bridge*, tai verbimääräyksen ja substantiivinmääräyksen, jos on käännettävä venäjänkieliset lauseet *они установили печь на твердом топливе* ('he asensivat uunin uudelle tehtaalle' ja 'he asensivat kiinteää polttoainetta käyttävän uunin'). Sipőczy päättyy ratkaisuluonnokseen, jota tuskin voi sanoa yksinkertaiseksi, mutta joka vaikuttaa koneen suorituskyvyn kannalta taloudelliselta ja käyttökelpoiselta, mikäli pysyttellään jonkin verraten vähäisiä ilmaisukeinoja käyttävän erikoisalan kielessä.

Koneella kääntämiseen vain välillisesti liittyy Alfréd Rényin artikkeli »Informaatioteoria ja kielitiede». Kirjoittaja katsoo, että sitä mukaa kuin informaatioteoriaa laajennetaan niin, että informaation määrän ohella voidaan ottaa huomioon myös informaation sisältö, se tulee yhä merkitsevämmäksi myös luonnollisten kielten tutkimukselle. Silti hän pitää puhtaasti kvantitatiivisia informaatioteoriaakin välttämättömänä esimerkiksi koneella kääntämisessä, sanakirjatyössä ja tyylintutkimuksessa. Kirjan viimeisenä artikkelina on Éva Ruzsiczkyn kirjoitus elektromekaanisten tietojenkäsittelykoneiden, lähinnä reikäkorttikoneiden soveltamisesta kielentutkimukseen.

Varsinaisten artikkelien lisäksi teos sisältää huomattavan määrän vähäisempiä kirjoituksia, jotka perustuvat konferenssissa esitelmien johdosta käytettyihin puheenvuoroihin.

Mikä on koneella kääntämiseen tähtäävän ja muun kieliteknologisen toimin-

nan suhde puhtaasti teoreettiseen kielentutkimukseen? Onko tässä puuhailussa lingvistillä mitään osaa? Kuten György Szépe sanoo, ei koneella kääntämisen ongelmia suinkaan ratkaista pelkällä insinööritaidolla, vaan yritykseen tarvitaan kielentutkijain, matemaatikojen ja insinöörien yhteistyötä. Toi-

saalta tällainen toiminta saattaa hyödyttää puhdasta lingvistiikkaa tarjoamalla ehkä uusia teknisiä apuneuvoja kielentutkijain käyttöön. Se voi vaikuttaa välillisesti myös yleiskielitieteelliseen ajatteluun esimerkiksi stimuloidessaan matemaattista kielentutkimusta.

MIKKO KORHONEN