

Monimielisyydet liikaa tietokoneille

Tutkimusryhmän merkitys kasvaa myös humanistisissa tieteissä. Yksinäisten susien kouluttamisen rinnalle tarvitaan myös perehdyttämistä vuorovaikutukseen, uskoo professori Fred Karlsson.

Tietokoneingvistiikka on meillä vain noin kymmenen vuotta vanha ala, jossa selvitetään tietokonetta apuna käyttäen kielten yleisiä rakenteita. Ala kuului aluksi eräänlaisena aputieteenä yleiseen kielitieteeseen, mutta on nyt saavuttanut itsenäisen statuksen. Helsingin yliopistossa sitä on voinut lukea pääaineena kolmisen vuotta. Tutkimusyksikkö on jo niittänyt kansainvälistä mainetta.

— Perinteisesti humanistisissa tieteissä tutkija on nähty individualistina, joka hyvin yksilöllisesti asettaa kysymykset itselleen ja pyrkii niihin sitten vastaamaan. Asetelma on muuttumassa. Vaikka todellisia individualistejakin tarvitaan, tällä hetkellä näyttää siltä, että humanistisessa tutkijankoulutuksessa kannattaisi satsata entistä enemmän ryhmätyöskentelytaitoihin. Entistä useammin menestyvän tutkimuksen perusyksikkönä humanistisissa tieteissäkin alkaa olla ryhmä. Tämä on totta myös meidän alallamme, kertoo professori Fred Karlsson. Hän on ollut tietokoneingvistiikan pioneereja Suomessa.

Ryhmän merkityksen korostamisella Karlsson haluaa myös tuoda esiin sen piirteen, mikä tiedotusvälineitä usein leimaa: samat tutkijat nostetaan yhä uudestaan esiin samojen alojen symboleiksi. Karlsson haluaakin karistella harteiltaan "Suomen ainoan tietokoneingvistin" manttelia. Hänen mielestään esimerkiksi laitoksella työskentelevät Kimmo Koskenniemi, Lauri Carlson ja Atro Voutilainen ansaitisivat tulla mainituiksi aivan yhtä perustellusti kuin hän itse.

Ei ongelmia tiedekuntarajoista

Tietokoneingvistiikka on alana kulkenut Suomessa onnellisten tähtien alla. Suomen Akatemia on tukenut sitä pitkäjänteisesti vuodesta 1981. Myös Helsingin yliopisto on tullut vastaan. Tutkimusyksikkö on ollut suoraan konsistorin alainen vuodesta 1984. Yksikössä on myös täytettävänä alan ensimmäinen varsinainen professuuri Suomessa. Lisäksi Tekesin kautta on päästy mukaan kansainvälisiin projekteihin.

Tietokoneingvistiikan kehittymistä eivät myöskään ole haitanneet tiedekuntarajat.

— Itse asiassa olemme olleet erittäin onnekkaita. Joskus tunnen jopa omantunnon tuskia siitä, miten sujuvasti kaikki on mennyt. Esimerkiksi kognitio-

tiede, joka on eräällä tavalla verrattavissa tietokoneingvistiikkaan, ei ole Suomessa saavuttanut läheskään sellaista asemaa, joka sille kuuluisi. Muun muassa ihmisen mielen tietoprosesseja mallittava ala on tavallaan jäänyt väliinpuotoajaksi, jota nyt ylläpitävät erityisesti psykologia ja filosofia. Toisaalta resurssiltaan pienessä maassa ei mitenkään voida rahoittaa kaikkia kehittyviä aloja yhtä paljon, Karlsson sanoo.

Hän mainitsee, ettei oman alan erkaantuminen kuitenkaan merkitse suhteiden katkaisemista lähitieteisiin.

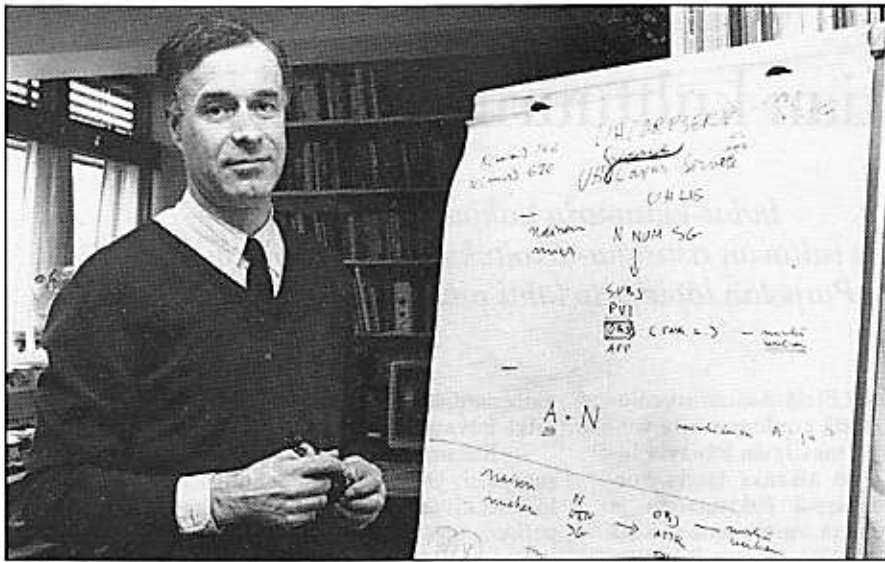
— Lähtökohtana tietokoneingvistiikassa on kieli, joka taas liittyy kaikkeen inhimilliseen toimintaan. Tämän takia yhteistyötä kielitieteessä on tehty ennenkin eri tieteiden välillä. Eräänä osoituksena monitieteellisyydestä on se, että pari vuotta sitten laudaturseminaarissa oli sekä teologian että lääketieteen tohtori suorittamassa jatko-opintoja. Tätä pidän hyvänä saavutuksena. Meillä on yleensäkin paljon sivuaineopiskelijoita muista tiedekunnista ja Teknillisestä korkeakoulusta, Karlsson kertoo.

Tekstipankin ohjelmistotoimittaja

Tietokoneingvistiikka kehittyi tarpeesta analysoida suuria määriä tekstejä. Sen kehitykseen vaikuttivat mm. ohjelmointikielten teoriat ja tietenkin tietokoneiden kapasiteetin kehittyminen. Yhtenä tietokoneingvistiikan tarkoituksena on löytää yleiset rakenteet yksittäisten kielten "takaa". Yleisemmän teorian pohjalta voidaan sitten rakentaa ohjelmia, jotka analysoivat eri kielten rakenteita.

Parhaisiin tuloksiin on päästy automaattisessa lausenjäsennyksessä. Karlsson näyttää, kuinka laitoksella rakennettu ohjelma jäsentää nopeasti esimerkiksi englanninkielisen auton käyttöoppaan tekstin. Jokaisen sanan jälkeen kone ilmoittaa perusmuodon sekä yksilöi taivutusmuodot.

Kymmenen vuoden tutkimuksen jälkeen ollaan päästy siihen, että useiden kielten jäsentäminen voidaan suorittaa automaattisesti tietokoneilla. Tästä on hyötyä kielitieteilijöille myös aineiston käsittelyssä. Tietokone voi poimia esimerkiksi suuresta määrästä tekstiä halutut taivutusmuodot, joita tutkija voi sitten ryhtyä käsittelemään. Tietokoneingvistiikan laitos järjestääkin kursseja mm. muiden yliopistojen



Aito Rusanen

— Huumoria ymmärtävän tietokoneohjelman rakentaminen tuskin onnistuu, epäilee professori Fred Karlsson.

kielitieteilijöille. Niillä kerrotaan uusista ohjelmista ja niiden sovellusmahdollisuuksista.

— Meidän laitoksellamme on rakennettu hyvin toimivat ohjelmat suomeksi, ruotsille, englannille ja venäjälle. Kaiken kaikkiaan kehittämiämme teorioita on sovellettu noin parinkymmenen kielen jäsenyksen, Karlsson kertoo.

Kehittyneintä tietokoneolingvistiikan tutkimus on USA:ssa. Myös Japani on panostanut paljon alan tutkimukseen. Euroopassa johtavia maita ovat Englanti ja Saksa. Helsingin yliopiston tietokoneolingvistiikan tutkimusyksikkö toimii läheisessä yhteistyössä parinkymmenen kansainvälisen yliopiston kanssa. Helsinkiläiset ovat kuuluisia muun muassa Kimmo Koskenniemen kehittämästä muoto-opin teoriasta. Yksikön ohjelmilla pystytään nopeasti analysoimaan suuria määriä juoksevaa tekstiä. Yhtenä osoituksena arvostuksesta on osallistuminen Bank of English -hankkeeseen.

— Birminghamin yliopistossa kehitetään tekstipankkia, johon "säilötään" viitisen miljoonaa englanninkielistä sanaa eri taivutusmuotoineen. Sanat ajetaan läpi meidän ohjelmillamme, jotka suorittavat jäsenyksen automaattisesti. Ruotsissa on käynnissä samanlainen projekti, ja myös he käyttävät hyväkseen meidän ohjelmiamme, Karlsson kertoo.

Tietokone ei erota ironiaa

— Kun nyt hallitsemme kielen muodon, voidaan seuraavaksi lähteä tutkimaan merkityksiä, Karlsson sanoo.

Hänen mukaansa tämä tulevaisuuden urakka on niin vaikea, ettei se täydellisesti onnistu ikinä. Luonnollisessa kielessä on piirteitä, joita ei pysty esittämään kattavan teorian avulla ja joista ei siis pysty rakentamaan tietokoneohjelmaa. Jo pelkästään tavallista sanakirjaa katsomalla huomaa, kuinka paljon merkityksiä eri sanoilla on. Lisähankaluu- tena ovat murteet, slangit ja erilaiset virheet, joita

ihmilliset puhujat tekevät. Myös monien sanojen merkitykset ovat hyvin liukenevia. Lauseita käytetään lisäksi usein sellaisina metaforina, joita on hankala purkaa osiin.

— Jos ajatellaan esimerkiksi lausetta "Aatteet taistelevat", tulkintatapoja on monia. Sanan "taistella" merkitys on alunperin ollut hyvin konkreettinen. Mitä sitten tarkoitetaan aatteiden taistelulla? Taistelevatko ideologiat todella keskenään yrittäen poissulkea toisensa, vai niiden kannattajat? Ovatko kannattajat käsirsyissä, vai käyvätkö vain sanallista väittelyä? Kattavan teorian luominen tämänkaltaisista kysymyksistä on melko mahdotonta, Karlsson pohtii.

Kieli on myös läheisessä suhteessa arkitietämykseen. Olisi vaikea luoda teoria, joka sisältäisi jonkinlaisen mallin niistä asioista, joita voidaan ajatella tavallisen ihmisen tietävän. Esimerkiksi sana "auto" on hyvin hankala määritellä sanallisesti. Käsitksemme siitä perustuu konkreettisiin esineisiin, joita näemme päivittäin. Toisissa kulttuureissa taas saattaa olla niille tuttuja ilmiöitä, jotka eivät avaudu muille.

Ongelmat eivät Karlssonin mukaan kuitenkaan merkitse sitä, että tietokoneolingvistien pitäisi kokonaan antautua merkityksiin liittyvien kysymysten äärellä.

— Joitakin osa-alueita merkitysoipista voidaan varmaan kuvata mielenkiintoisesti, hän uskoo.

Mutta on kuitenkin olemassa rajoja, joita ei hänen mielestään voida ylittää. Mahdotonta olisi hänen mielestään esimerkiksi se, että tietokone joskus osaisi tulkita ironiaa: sitä että lauseen merkitys tietyssä kontekstissa olisi täsmälleen päinvastainen. Eikä esimerkiksi poliitikkojen tarvitse pelätä sitä, että jonakin päivänä tietokone pystyisi suoraan puheesta kertomaan, mitä puhuja todella tarkoittaa sanoillaan.

Jari Siikari