



kaasti niin kielitieteessä kuin sen lähi-aloillakin, erityisesti psykologiassa ja tekoälyn tutkimuksessa.

## Tutkimuksen nykytila

Tällainen tutkimus on kaikkialla alkuvaiheessaan, jos tavoitteena on ihmisen arkisen viestintäkyvyn kuvaus. Parhaat ohjelmistot pystyvät "keskustelemaan" vain tarkoin rajatuista aihepiireistä, kuten tautien määrityksistä tai tieteellisten näytteiden analyysistä. Ohjelmistoissa on vain muutaman tuhannen sanan sanasto ja ne pystyvät käsittelemään vain suppeaa lauserakenteiden valikoimaa.

Ihminen taas hallitsee kymmeniä- tai satojatuhansia sanoja (joillakin niistä on monia merkityksiä) sekä suurta lauserakenteiden joukkoa. Lisäksi hänellä on joustava kyky tuottaa ja ymmärtää myös vaillinaisia lauseita, ja hän voi ymmärtää paitsi kirjaimellisesti myös assosioimalla ja vertailemalla sanoja ja merkityksiä.

## Koneellinen kääntäminen

Koneellinen kääntäminen on puhutanut vuosikymmeniä. Ensimmäiset yritykset epäonnistuivat. Viime vuosina ajatus on taas tullut esiin mm. japanilaisten suunnitteleman ns. viidennen sukupolven tietokoneen yhteydessä.

Sen ohjelmistokehittelyn keskeisimpiä tavoitteita on parantaa koneen kielenkääntökykyä erityisesti japanista useihin muihin kieliin. Ratkaisu ei kuitenkaan ole olennaisesti lähempänä kuin 1960-luvulla. Oireellista onkin, että uusimpien tietojen mukaan japanilaiset ovat reivaamassa hankkeensa kielenkääntöpuolta realistisemmaksi.

Korkealuokkainen, tekstilajiin sitoutumaton automaattinen kääntäminen kompastuu ennen kaikkea siihen, ettei sanavarastoa pystytä käsittelemään niin kuin pitäisi. Kysymys ei ole vain kahden kielen sadoistatuhansista sanoista ja niiden monista merkityksistä ja rajattomista yhdistelymahdollisuuksista. Luonnolliselle kielelle tyypillistä on myös se, että sanat, lausekkeet ja kokonaiset lauseet voivat saada idiomattisia erikoismerkityksiä, jotka eivät vastaa toisiaan siirryttäessä kielestä toiseen.

Sanakirjat tulevat siis säilyttämään merkityksensä.

## Käytännön tuloksia

Edellä on pantu jäitä hattuun. Ihmisen ja tietokoneen viestintä luonnollista kieltä käyttäen on ongelma, jonka ratkaiseminen kaipaa ennen kaikkea aikaa vievää kielitieteellistä perustutkimusta.

Tällaista työtä tehdäänkin niin meillä kuin muualla. Helsingin yliopiston yleisen kielitieteen laitoksella ovat tri Kimmo Koskenniemi ja allekirjoittanut

harjoittaneet tietokonekielintieteen teorian yhteistyötä kohta viiden vuoden ajan.

Mielenkiintoista on ollut nähdä, kuinka sanarakenteen sekä muoto- ja lauseopin teoriaan kohdistunut työ on sivuraiteella poikunut tuotantoelämässä toimivia sovelluksia. Tällaisia ovat esimerkiksi lähes virheetön suomen kielen tavutuslauseke ja suomenkielisessä informaation haussa toimiva sanamuotojen tunnistusohjelma.

Näistä sovelluksista on silti pitkä matka vapaan viestinnän mallintamiseen.

## Köyhtyykö kieli?

Usein pelätään, että tietokoneiden kasvava käyttö köyhdyttäisi jopa kielen rakennetta.

Vaaroja piilee lähinnä siinä, että pienet lapset voivat alkaa yksipuolisesti harrastaa tietokonepelejä ja -ohjelmointia 3-7-vuotiaina, jolloin he ovat vasta omaksumassa äidinkieltään. Silloin koneen viestintärajoitukset saattavat haitata normaalia äidinkielen omaksumista, erityisesti jos sana- ja käsitevaraston kasvattaminen kirjoja lukemalla samalla jää vähemmälle.

Äidinkielen omaksuminen on tärkeältä osaltaan laajojen käsiteverkostojen ja uusien yhteyksien oppimista monipuolisen aineiston ja vapaan mielikuvituksenkin varassa.

Viestintä tietokoneen kanssa on parhaimmillaankin melko esineellistä. Siihtä ei ole haittaa, jos on kyse tieteellisistä tai kaupallis-hallinnollisista sovelluksista. Mutta toisin on, kun tietokoneet tulevat koteihin ja kouluihin eli tavallisten ihmisten viestintäympäristöön.

On keskeinen kulttuuri-intressi, että tietokoneet pystyisivät mahdollisimman sujuvaan ja rajoittamattomaan viestintään. Muuten rengistä tulee herra. Käsitellään vain asioita, joihin kone pystyy. Pahimmillaan viestintä koneen kanssa latistaa.

Toinen näkökulma on kansallisen kulttuurin vaaliminen. Englanti on atk:n valtakiel. Silti haluamme varmaan käyttää kansalliskieliamme tulevaisuuden tietokoneviestinnässä. Nämä ovat arvoja, joista tuskin halutaan tinkiä, niin kansainvälistyvässä maailmassa kuin elämissä.

Ja lopuksi: Mitä luonnollisemmalla kielellä koneet ja ihmiset viestivät keskenään, sitä useammat ihmiset pystyvät käyttämään tulevaisuuden tietojärjestelmiä ja sitä vähemmän he tulevat passiivisesti manipuloiduiksi. Näin voidaan edistää tasa-arvoa ja humanistisia ihanteita.

Fred Karlsson



Kirjoittaja on Helsingin yliopiston yleisen kielitieteen professori.