

Hyvönen - Karanta - Syrjänen, eds.,
Teknodyn ensyklopedia, Gaudramus,
Helsinki 1993, 47-52.

1.4

Lingvistiikka

Fred Karlsson

YLEINEN LINGVISTIikka JA FILOLOGIA

Lingvistiikka [linguistics] tarkoittaa kieli-tiedettä tai kielentutkimusta (vrt. latinan *lingua* 'kieli'). Lingvistiikan harjoittaja on *lingvisti* [linguist] eli kielentutkija. Lingvistiikka voi kohdistua kieleen tai kieliin yleisesti, jolloin kysymyksessä on yleinen lingvistiikka eli *yleinen kielitiede* [general linguistics]. Tämä on kielitieteen teoria- ja menetelmäoppia, joka pyrkii vastaamaan seuraavanlaisiin kysymyksiin: Mikä on *luonnollinen kieli* [natural language]? Mitkä ovat luonnollisten kielten erityispiirteet verrattuna esimerkiksi muiden eläinlajien viestintäjärjestelmiin tai ihmisen keksimiin toissijaisiin viestintäjärjestelmiin? Mikä on luonnollisen kielen merkitys ihmiselle, hänen toiminnalleen ja kulttuurilleen? Mitkä ovat eri kielten rakenteen yhteiset ja erilaiset piirteet? Mitkä kielet ovat sukua keskenään? Mitkä formaaliset ominaisuudet luonnollisten kielten kielioppeilla on? Miten kielet muuttuvat aikojen kuluessa? Miten ja missä rajoissa luonnolliset kielet voivat vaihdella esimerkiksi iän, sukupuolen, koptaikkakunnan, sosiaaliluokan ja muiden kielenulkoisten parametrien mukaan? Mi-

ten ihminen omaksuu *ensikielensä* [first language, L1] eli *äidinkiellensä* [mother tongue] lapsena? Miten hän oppi myöhemmällä iällä vieraita kieliä? Mikä on kielen ja kulttuurin suhde? Vaikuttaako kieli (erityisesti ensikielen myötä omaksuttu käsitejärjestelmä) tapaan, jolla ihminen hahmottaa ja kokee ympäröivän maailman?

Lingvistiikasta puhutaan myös yhden tietyn kielen tutkimuksen kohdalla. Yksittäisistä kielistä käytetään lingvistiikassa termiä *erityiskieli* [particular language]. Näitä ovat suomi, ruotsi, kiina, hottentotti, suahili, venäjä jne. Kielentutkijoiden suuri enemmistö on erikoistunut tietyn kielen tutkimiseen, joskus toisilleen läheistä sukua olevien kielten tutkimiseen.

Lingvistiikka-termiin liittyy usein ajatus tietynlaisten strukturaalisten, formaalisten tai ainakin yleisyyttä tavoittelevien menetelmien käytöstä vastakohtana perinteiselle *filologialle* [philology]. Filologialla tarkoitetaan ahtaassa merkityksessä vanhojen tekstien ainutkertaisten piirteiden tutkimusta, erityisesti niiden kulttuuritaustaa vasten peilaten. Toisaalta filologialla voidaan tarkoittaa suurin piirtein samaa kuin lingvistiikkakin.

KIELEN RAKENNE JA KÄYTTÖ

Kieli on ihmisten välisen viestinnän keskeisin ja kehittynein väline. Kielen kaksi perusluottuvuutta ovat sen *rakenne* [structure] ja *käyttö* [use]. Sekä rakenteen että käytön kuvausta varten on lingvistiikassa kehitetty monenlaisia teorioita. Rakenne ja käyttö eivät ole toistensa vastakohtia tai vaihtoehtoja, vaan rakenne on olemassa, jotta kieltä voitaisiin käyttää tehokkaasti: rakenne palvelee käyttöä.

Kielen rakenteen kolme perusyksikköä ovat äänneet, sanat ja lauseet. Äännerakenteen kuvausta sanotaan *fonologiaksi* [phonology], sanojen rakenteen kuvausta *morfologiaksi* eli muoto-opiksi [morphology] ja lauserakenteen kuvausta *syntaksiksi* eli lauseopiksi [syntax]. Fonologia kuvaa puhetta sen äännerakenteen näkökulmasta, toisaalta puhe sekä sen tuottaminen ja havaitseminen ovat myös konkreettisia ilmiöitä, joita *fonetiikka* [phonetics] tutkii. Morfologiaa ja syntaksia (sekä joskus fonologiaakin) yhdessä kutsutaan *kieliopiksi* [grammar].

Kielen rakenteita käytetään viestittäessä, ja niiden avulla viestitään *merkityksiä* [meaning]. Merkityksetkin ovat osittain rakenteellisia, mutta ei samalla tavoin ulkoisesti näkyvästi kuin äänneet, sanat ja lauseet. Merkityksen tutkimusta sanotaan *semantiikaksi* [semantics]. Semantiikkaan läheisesti liittyvä ala on *pragmatiikka* [pragmatics], jolla tarkoitetaan sen tutkimista, miten merkitysten tulkinta riippuu puhetilanteesta ja sen osanottajista.

Kun kuvataan tietyn ajankohdan kielijärjestelmää, puhutaan *synkronisesta kielitieteestä* [synchronic linguistics]. Yleensä tällainen tutkimus kohdistuu nykykieleen, mutta se voi kohdistua tiettyyn varhempaankin kielimuotoon.

Ihmiset käyttävät kieltä moninaisesti elämänsä vaihtelevissa tilanteissa. Jokainen ihminen on monessa suhteessa vähän erilai-

nen kuin muut. Tämänkaltaisten tosiseikkojen valossa on luonnollista, että yhden ja saman kielen käytössä on paljon *vaihtelua* [variation]. Kielenkäytön vaihtelua aiheuttavia tekijöitä ovat muun muassa ikä, kotipaikkakunta, sosiaaliryhmä, sukupuoli ja puhetilanne.

Nuoret puhuvat eri tavalla kuin vanhemmat ihmiset, mikä on tärkeimpiä tekijöitä *kielenmuutoksissa* [language change]. Erityisesti lapsuuden aikainen kasvuympäristö predestinoi ihmisen puheenpartta tai murretta [dialect] sen seudun mukaiseksi. Toki ihminen myöhemmällä iällä saattaa muuttaa puhettomuksiaan, erityisesti "aksenttiaan" eli ääntämystään (esimerkiksi helsinkiläistyneet savolaiset tai raumalaiset). Kielenmuutoksia tutkittaessa verrataan usein kahta tai useampia kielimuotoja – tällöin puhutaan *diakronisesta kielitutkimuksesta* [diachronic linguistics].

Yhteiskuntajärjestelmästä riippuen kielialueella voi olla sosiaalisin seikkoihin perustuvaa vaihtelua. Englannissa niin sanotut *sosiaaliset murteet* [social dialects] ovat paljon selvemmin havaittavia kuin Suomessa, mikä korreloi sen kanssa, että sosiaaliset erot ovat Suomessa pienempiä.

Naisten ja miesten kielenkäyttöökään ei ole aivan samanlaista, vaan muun muassa sananvalinnoissa on eroja. Vihdoin jokainen puhuu ja kirjoittaa vähän eri tavoin tilanteesta riippuen: julkisuudessa puhuttaessa äännetään huolellisemmin kuin kotona arkitilanteissa, ystävälle kirjoitetaan toisella tavalla kuin yliopistollista opinnäytettä laadittaessa.

Puhtaan foneettista eli ääntämysvaihtelua aiheuttavat myös monet yksilölliset anatomiset tai aivan satunnaiset erot. Jos ihmisellä on pitkät ja/tai paksut äänihuulet, hän puhuu keskimäärin matalammalla äänellä kuin henkilö, jolla on lyhyemmät ja ohuemat äänihuulet. Jos kitapurje vähän vuotaa, puheeseen tulee lievä nenäsointi. Tavalli-

nen puhenopeus voi eri henkilöillä vaihdella paljonkin.

LUONNOLLISEN KIELEN KESKEISET OMINAISUUDET

Lingvistiikka on paljastanut 1900-luvulla joukon keskeisiä piirteitä, jotka ovat yhteisiä kaikille luonnollisille kielille.

1) "Luonnollinen" tarkoittaa, että on ihmisiä, jotka käyttävät (tai ovat joskus käyttäneet) tiettyä kieltä arkielämänsä keskeisimpänä viestintävälineenä. Tämän vaatimuksen täyttävät esimerkiksi sekä suomi että latina – niiden erona on se, että latinaa ei enää käytetä arkitilanteiden peruskielenä.

2) Sanat ja päätteet ovat *symboleja* [symbols], *mielivaltaisia* [arbitrary] äännettyjä merkkejä. Mielivaltainen tarkoittaa tässä sitä, että ei ole mitään pakottavaa tai luonnollista syytä, miksi esimerkiksi lehmää ilmaiseva sana on suomeksi juuri 'lehmä' eikä vaikkapa 'kehmä'. Normaalitapauksessa symbolit ovat *vokaalisia* [vocal] eli äännettyjä. Kuurojen viittomakieli on erikoistapaus, jossa symbolit (eli yksittäiset viittomat) ovat visuaalisia eli perustuvat näköaistiin.

3) Kielellisten symbolien rakenne on kaksijakoinen, *duaalinen* [dual], eli (ensimmäinen jako) symboleihin kuuluu merkitys ja muoto, mutta lisäksi (toinen jako) muoto koostuu pienemmistä yksiköistä, *foneemeista* [phoneme], kuten /a/, /i/, /p/, /s/.

4) Kielen järjestelmä on nimenomaan sanaston tasolla avoin. Pienellä määrällä foneemeja, tavallisesti muutamilla kymmenillä, voidaan luoda suuria sanastoja, joissa on satoja tuhansia tai jopa miljoonia sanoja ja joihin voidaan aina tarpeen vaatiessa muodostaa, keksiä tai lainata uusia sanoja.

5) Kielessä on sosiaalisin normeihin pohjautuvia *sääntöjä* [rules], jotka kielen-

puhujat tiedostamattaan tuntevat, vaikka he eivät yleensä osaakaan itse formuloida niitä. Säännöt määrittävät kielen sanat ja rakenteet ja niitä on kielen kaikissa osajärjestelmissä. Koira on suomeksi 'koira' (yksi pieni, yhtä tiettyä sanaa koskeva sääntö). Syntaktinen suomen kieltä koskeva sääntö on se, jonka mukaan vartalossaan konsonanttiin ja yhteen vokaaliin päättyvät verbit muodostavat yksikön 3. persoonan muodon vokaalinpidennyksellä ('hän tule+e', 'Leena mene+e'). Kielen kieliopin keskeisen osan muodostavat morfologiset ja syntaktiset säännöt.

6) Kielen käyttö on luovaa. Käyttämällä perussymboleja, sanoja, ja kielen sääntöjä voidaan käytännössä joustavasti muodostaa mitä tahansa sanomia.

7) Kielellä muodostettavat sanomat eivät ole oleellisesti tilannesidonnaisia (vaikka niissä voikin esiintyä yksittäisiä sanoja kuten pronomineja, esimerkiksi 'tuo', jotka viittaavat kielenulkoiseen tilanteeseen mukana olevaan olioon).

KIELITYPOLOGIA

Arviot maailman kielten lukumäärästä vaihtelevat yleensä 3000:n ja 7000:n välillä, usein mainitaan summittainen luku 5000. Täsmällistä lukumäärää ei tietenkään voi ilmoittaa. Kielen määritelmiä on useita ja kriteerit ovat usein ristiriitaisia. Klassinen kriteeri on ymmärrettävyys: jos kaksi kielimuotoa on keskenään ymmärrettäviä, ne ovat saman kielen muunnoksia. Tämän mukaan ruotsia ja norjaa voitaisiin pitää samana kielenä, mitä kuitenkin ei käytännössä poliittisista syistä tehdä (niitä puhutaan eri valtioissa).

Kielitypologia selvittää, minkälaisia lingvistisiä rakenteita maailman kielissä esiintyy. Kielet jaetaan tyyppeihin esimerkiksi sen mukaan, paljonko niiden sanojen vartaloissa on taivutuksen yhteydessä ta-

pahtuvia äännevaihteluita sekä päätteiden ja vartalon yhteensulautumisia (niin sanottu fuusion indeksi), paljonko päätteitä sanoissa voi olla peräkkäin (syntecttisyysindeksi), mikä on lauseiden perustava sanajärjestys (esiintyykö esimerkiksi subjekti normaalisti ennen verbiä vai sen jäljessä), esiintyvätkö substantiivin määritteet normaalisti ennen substantiivia vai sen jäljessä, jne.

Kielitypologian kriteerit ja menetelmät ovat keskeisiä myös silloin, kun selvitetään kielten sukulaisuussuhteita. Kielten typologiaa ja keskinäistä sukulaisuutta on tutkittu satoja vuosia. Arviot maailman kielikuntien lukumäärästä vaihtelevat silti vielä melkoisesti, 20:stä 200:aan.

LINGVISTIikka JA MUUT TIEETEET

Kielellä on ulottuvuuksia moneen suuntaan. Kieli voidaan ensiksikin ymmärtää omaehtoiseksi eli autonomiseksi järjestelmäksi, *systemiksi* [system]. Tämä on useimpien lingvistien mielestä kielentutkimuksen keskeinen tutkimuskohde. Kielijärjestelmän tutkimuksesta käytetään joskus termiä *autonominen kielentutkimus* [autonomous linguistics]. Autonomisuus viittaa siihen, että kielen järjestelmänluonnetta tutkittaessa ei välttämättä tarvitse ottaa kantaa kielenulkoisiin seikkoihin, esimerkiksi siihen, minkälaisia todelliset puheen ja sen ymmärtämisen prosessit ovat.

Käytännössä kieltä tarkastellaan osajärjestelmien kokonaisuutena, jossa keskeisimmät osajärjestelmät ovat edellä puheena olleet fonologia, morfologia, syntaksi ja semantiikka. "Osajärjestelmällä" pitää olla sille ominaisia yksiköitä tai peruskäsitteitä, jotka erottavat sen muista osajärjestelmistä.

Mutta kielisysteemi on myös edustuneena eli representoituneena yksittäisten ihmisten aivoissa, sekä neuraalisesti että yleisemmin mentaalisesti, enemmän tai vähem-

män selvinä todellisuudessa tapahtuvina, *aikapaikkaisina* [spatio-temporal] prosesseina. *Psykolingvistiikka* [psycholinguistics] selvittelee kielen mentaalirepresentatioita ja siihen liittyviä kielenkäytön prosesseja empiirisen (usein kognitiivisen) psykologian ja lingvistiikan menetelmin. *Neuro-lingvistiikassa* [neurolinguistics] tutkitaan kielen ilmiöitä ennen kaikkea aivotointojen ja niiden häiriöiden osalta.

Kielisysteemi on myös sosiaalinen toiseikka, jonka tietyn *kieliyhteisön* [language community] jäsenet jakavat siten kuin jakavat muutkin sosiaaliset norminsa. Kielen sosiaalisen ulottuvuuden ilmiöitä selvittelee *sosiolingvistiikka* [sociolinguistics], menetelmin ja taustamuuttujin, jotka saavat paljon vaikutteita sosiologiasta.

Antropolingvistiikan [anthropological linguistics] nimikkeen alla tutkitaan kieltä sekä kielen vaihtelua ja käyttöä erityisesti suhteessa ihmisen (muuhun) kulttuuriin ja uskomuksiin.

Tietokonelingvistiikka [computational linguistics] selvittää luonnollisen kielen rakennetta ja käyttöä laskennallisten mallien avulla. Näiden avulla mallinnetaan *luonnollisen kielen prosessointia* [natural language processing, NLP], esimerkiksi kirjoitettujen lauseiden automaattista *lauseenjäsennystä* [parsing]. Tätä laajempi, keskeinen NLP:n ongelma on automaattinen kääntäminen eli *konekäännös* [machine translation, MT].

Kielifilosofia [language philosophy, linguistic philosophy] on filosofiasta ja logiikasta paljon vaikutteita saanut suuntaus, jonka keskeisiä ongelmia ovat olleet muun muassa merkityksen luonne ja formalisointi, kielen suhde ulkomaailmaan sekä ne rajoitukset, jotka mahdollisesti sanelevat, mitä kielellä yleensä voi ilmaista.

Soveltavan kielitieteen [applied linguistics] nimikkeen alla kulkevat erityisesti sellaiset kielentutkimuksen sovellukset, jotka liittyvät kielen opetukseen ja oppimiseen,

niin ensikielen kuin vieraidenkin kielten osalta. Soveltavaksi kielitieteeksi luetaan usein myös kääntämisen tutkimus sekä kielienhuolto ja *kielisuunnittelu* [language planning].

LINGVISTIikka JA TEKOÄLY

Kielen käyttöä pidetään yleensä ihmisen älykkään käyttäytymisen yhtenä tyypillisimpänä piirteenä. Näin muodoin on selvää, että luonnollinen kieli ja sen käyttö sekä lingvistiikka laajemminkin ovat myös tekoälyn ongelmille ja tutkimukselle relevantteja.

Älyn käsitettä ei kielen kohdalla voi ymmärtää suppeasti, esimerkiksi ensisijaisesti ongelmanratkontakyknä. Kieli on nähtävä osana ihmisen muita *kognitiivisia toimintoja* [cognitive skills, cognitive abilities], joita ovat muiden muassa havaitseminen, muistaminen, oppiminen, päättely sekä tietäminen. Kaikissa näissä luonnollisella kielellä on jonkinlainen osa.

Tekoälyn oppikirjoissa on pitkään yhdessä tai useammassa luvussa käsitelty luonnollisen kielen prosessointia eli tietokoneohjelmia, jotka pystyvät "älykkäällä" tavalla käsittelemään luonnollista kieltä. Tällaisen käsittelyn vaiheita on useita:

1) Sananmuotojen *morfologinen analyysi* [morphological analysis] eli niiden palauttaminen perusmuotoon ynnä tietoihin taivutuspiirteistä, esimerkiksi sananmuoto "menisi" on "mennä"-verbin konditionaalin preesensin 3. persoonan muoto ("hän men+isi") taikka sen kieltomuoto ("en men+isi").

2) Morfologisten moniselitteisyyksien ratkominen eli *disambiguointi* [disambiguation], esimerkiksi lauseessa "Kalle men+isi sinne" sananmuodon 'menisi' oikea tulkinta on konditionaalin preesensin 3. persoona.

3) Morfologisesti analysoitujen lauseiden syntaktinen analyysi eli jäsenitys, esimerkiksi äskeisen lauseen täydellinen morfologinen ja syntaktinen rakenne on (1 = sananmuoto, 2 = perusmuoto, 3 = morfologiset piirteet, 4 = syntaktinen tehtävä):

1	2	3	4
Kalle	Kalle	substantiivi erisnimi yksikkö nominatiivi	subjekti
menisi	mennä	verbi konditionaali yksikkö 3. persoona	finiitti- predi- kaatti
sinne	sinne	adverbi	adverbi- aali

4) Lauseiden *kirjaimellisen merkityksen* [literal meaning] tulkinta, esimerkiksi verbi 'menisi' edellä tarkoittaa tietynlaista, ei kovin nopeaa konkreettista liikettä, jonka myötä elävä olio siirtyy, mahdollisesti paikasta toiseen. Olio on lauseen subjektin *tarkoite* [referent]. "Mennä"-verbillä on muitakin merkityksiä, jotka eivät tule tämän lauseen tulkinnassa kysymykseen mielekkäinä vaihtoehtoina, esimerkiksi abstraktinen merkitys "sujua" kuten lauseessa "Koe meni häneltä kehnosti", tai merkitys 'kuolla' kuten lauseessa "Meiltä meni hyökkäyksessä puolet miehistä".

5) Lauseen *kontekstimerkityksen* [contextual meaning] tulkinta, esimerkiksi sana 'Kalle' tarkoittaa puheena olevassa *diskursissa* [discourse] eli keskustelussa tiettyä, mahdollisesti aiemmin puheena ollutta oliota, joka mahdollisesti voidaan tarkasti identifioida. Samaten adverbille 'sinne' voidaan kenties antaa tarkka paikallinen tulkinta – johonkin paikkaan, joka selviää kontekstin avulla.

Tämänkaltaisessa luonnollisen kielen käytön *mallinnuksessa* [simulation] on

yleensä käsitelty vain melko pieniä sanastoja, joissa on ollut muutamia kymmeniä tai satoja eri sanoja. Luonnollisen kielen tietokone mallinnuksessa on runsaasti ratkaisemattomia ongelmia. Yleisiäkin sanoja on joka kielessä ainakin kymmeniä tuhansia, ja näistä monet ovat monimerkityksisiä eli polyseemisiä. Nimenomaan merkityksen kuvaus on vielä vailla yleisesti hyväksyttyä teoriaa. Monimerkityksiset sanat vaikeuttavat huomattavasti sanojen semantiikan kuvausta. Useimpien sanojen merkitys, mikä merkitys "sellaisenaan" sitten onkin, on *sumenta* [fuzzy] rajoiltaan.

Merkityksen tilannetulkinta edellyttää yleensä *arkitietämystä* [common sense knowledge] maailmasta, sen esineistä ja niiden suhteista jne. On epätodennäköistä, että (ainakaan lähiaikoina) pystyttäisiin esittämään eksplisiittistä teoriaa arkitietämyksestä, joka olisi niin täsmällinen, että se olisi tietokoneella implementoitavissa. Ja vihdoin merkitystulkinta – kuten yleensäkin ihmisen havaitseminen – toimii suurpiirteisesti, osittaisen ja puutteellisenkin informaation nojalla. Ihminen pystyy päättämään oikein vaillaisten tietojen pohjalta havaitsemiaan kokonaisuuksia ja analogioita, ja tällaisten "älykkäiden" kognitiivisten toimintojen mallintaminen on hankalaa.

Teoreettisista ongelmista huolimatta luonnollisen kielen tietokonesovelluksia on runsaasti muiden muassa automaattisen oikoluvun ja tavutuksen, informaation tallennuksen ja haun, koneellisen kääntämisen sekä puheen tunnistuksen ja tuottamisen alueilla.

KIRJALLISUUTTA

- Comrie, B. (1989). *Language Universals and Linguistic Typology*. Cambridge University Press: Cambridge.
- Crystal, David (1987). *The Cambridge Encyclopedia of Language*. Cambridge University Press: Cambridge.
- (1991). *A Dictionary of Linguistics and Phonetics*. Basil Blackwell: London.
- Hakulinen, A. ja Karlsson, F. (1988). *Nykysuomen lauseoppi*. Suomalaisen Kirjallisuuden Seura: Helsinki.
- Hautamäki, Antti (toim.) (1988). *Kognitiiviede*. Gaudeamus: Helsinki.
- Karlsson, Fred (1993). *Yleinen kielitiede*. Yliopistopaino: Helsinki.
- Lass, Roger (1984). *Phonology*. Cambridge University Press: Cambridge.
- Levinson, Stephen C. (1983). *Pragmatics*. Cambridge University Press: Cambridge.
- Lyons, John (1977). *Semantics. I, II*. Cambridge University Press: Cambridge.
- (1984). *Language and Linguistics. An Introduction*. Cambridge University Press: Cambridge.

Harnd

len