



MONTA TAPAA »AJATELLA» — TILASTOLLISTEN MENETELMIEN HYÖDYNTÄMINEN AINEISTOLÄHTÖISESSÄ SANASTONTUTKIMUKSESSA

ANTTI ARPPE

VÄITÖKSENALKAJAISIESITELMÄ
HELSINGIN YLIOPISTOSSA
19. JOULUKUUTA 2008

Väitöstilaisuuden avaavan lektion pitäminen tässä salissa osoittautui minulle merkityksellisemmäksi, kuin aluksi kuvittelinkaan. Alkuperäinen tarkoitukseni oli olla ensimmäinen Arppe, joka väittelee kaukaisen sukulaiseni mukaan nimetyssä yliopiston rakennuksessa — mutta tässä salissa pidin itse asiassa ensimmäisen Suomessa pitämäni kielitieteellisen esitelmän Kari Pitkäsen vuonna 1999 pitämällä kurssilla. Kurssin nimi oli *Kieli ja konteksti*, esitelmän päivämäärä 21.10.1999, ja aiheenani kirjaesitelmä pohjautuen Revere D. Perkinsin (1992) tutkielmaan kielen ja kulttuurin tilastollisesta riippuvuussuhteesta. Tämän

yli yhdeksän vuoden takaisen esitelmäni teemat ja tuolloin esittämäni peruskysymykset ovat edelleen ajankohtaisia ja relevantteja — niin kielitieteessä ylipäänsä kuin väitöskirjassani tekemälle tutkimukselle erityisesti.

Esimerkiksi pohdin tuolloin, mikä on kielen suhde muihin (inhimillisiin) ilmiöihin. Onko kieli riippuvainen tai vuorovaikutussuhteessa muihin ilmiöihin? Voiko kieltä tutkia itsenäisenä ilmiönä vai tarvitseeko kielen vuorovaikutussuhdetta muihin ilmiöihin ottaa huomioon? Näille rinnakkaisia kysymyksiä ovat miten kieli kehittyy ja erityisesti, mitkä kulttuuriset tekijät vaikuttavat kielen kehitykseen ja muutokseen. Tuolloin näkemäni perusvastakainasettelut siinä, miten kielen sekä kielen kehityksen ja muutoksen luonne käsitettiin,

▷

vallitsevat kielitieteessä yhä. Yhdet näkevät kielen itsenäisenä ja itseriittoisena, tai abso-
luuttisena ilmiönä, jolloin yksittäinen kieli
koetaan ylipäänsä täydellisenä ja aina yhtä
ilmaisuvoimaisena — verrattuina muihin
samanaikaisesti olemassa oleviin kieliin tai
aikaisempiin kielen tiloihin. Toiset näke-
vät kielen muiden inhimillisten tekijöiden
kanssa vuorovaikuttavana — kasvavana,
laajentuvana, kehittyvänä ja muuttuvana il-
miönä. Tuolloin koin kuuluvani jälkimmäi-
seen ryhmään, eikä mielipiteeni ole järin
muuttunut, mitä väitöskirjani tulokset ovat
vain vahvistaneet. Tuolloin esitin lisäksi,
että nämä näkemyserot heijastuvat myös
eriävinä käsityksinä kielen muutoksesta.
Yksien mielestä kieli on oma autonominen
ilmiönsä, joka kehittyy omien sisäisten me-
kanismiensa ja prosessiensa varassa. Toiset
taas näkevät, että kieli muuttuu sitä käyttä-
vien ihmisten kulttuurin komplisoituessa,
sosiologisista syistä. Tässäkin kuuluin ja
edelleen kuulun jälkimmäiseen ryhmään,
nyt vain perusteluni näkemykselleni ovat
pelkkää omaa järkeilyä painavimmat —
annan tästä esimerkin tai pari myöhemmin
tässä lektiössä.

Siirryn nyt viime vuosituhannen viimei-
sestä vuodesta tähän päivään ja tilanteeseen,
johon olen päätenyt saatuaani nyt tarkastet-
tavan väitöskirjani valmiiksi. Tavoitteeni
on valottaa paitsi väitöskirjani keskeisiä kä-
sitteitä, kysymyksenasetteluja ja tuloksia,
myös yleistä näkemystäni siitä miten yri-
tän ymmärtää kielen ja kielitieteen. Yleiset
tutkimustavoitteeni kielitieteessä ovat seu-
raavat kolme. Ensimmäiseksi minua askar-
ruttaa, mikä on luonnollisen kielen käytön
sekä sen taustalle oletetun kielellisen järjes-
telmän — ja sitä kuvaavan teorian — suhde.
Tässä erityisenä mielenkiintoni kohteena
ovat sanasto- ja rakennevarianttien valinta,
sekä tällaisia valintoja selittävät tekijät ja
mallit. Toiseksi haluan selvittää ja kokeilla,
miten tätä suhdetta voidaan mallintaa eri

tilastollisilla monimuuttujamenetelmillä
— mistä **nimenomaisesti** väitöskirjassa-
ni on kyse. Ja kolmanneksi, olen edelleen
kiinnostunut siitä, mikä on kielen ja sitä
käyttävän kulttuurin suhde.

Tutkimuskohteenani ovat siis väitöskir-
jassani samaa merkitsevät rakennevariantit.
Näitä ovat toisaalta kieliopilliset rakenne-
variantit, toisaalta synonyymiset sanat.
Esimerkki rakennevarianteista on suomen
kaksi vaihtoehtoista vertailurakennetta:
voimme joko sanoa, että *hän on Pekkaa
parempi* tai *hän on parempi kuin Pekka*,
ilman että ilmaisun merkitys mitenkään
oleellisesti muuttuu. Esimerkki synonyymi-
sista sanoista ovat tutkimani suomen ajat-
telemisverbit eli *ajatella*, *mieltiä*, *pohtia*,
harkita, *tuumia*, *tuumata*, vain yleisimmät
mainitakseni.

Aikaisemmissa tutkimuksissa on havait-
tu laaja kirjo eri tekijöitä, jotka vaikuttavat
siihen mikä rakennevariantti kulloinkin va-
litaan. Näitä ovat ainakin sanallinen ympä-
ristö, kieliopillinen ympäristö, tekstityyppi
tai tyyli, tai jopa sanansisäinen rakenne eli
taivutusmuodot, joissa sanoja käytetään.
Sanallisen ympäristön vaikutuksen näem-
me siinä, että esitämme suomessa mieluum-
min *keskeisen kysymyksen* kuin *tärkeän* sel-
laisen, kun taas havaitsemme enemmän
tärkeän seikan kuin *keskeisen* — ainakin
jos laajoja tekstiaineistoja on uskominen.
Kieliopillisen ympäristön voimme taas hu-
mata siinä, että ensimmäisessä persoonassa
yksilöllisenä tekijänä mieluiten *minä mie-*
tin, mutta kollektiivinen tekijä kuten *halli-*
tus pohtii. Tekstityypin tai tyylin luoma ero
puolestaan näkyy siinä, että sanomalehtien
politiikkasivuilla mieluiten *pohditaan* kun
taas urheilun puolella enemmän *mietitään*.
Ja lopuksi suomenkielinen esimerkki sa-
nansisäisen rakenteen eli taivutusmuotoon
perustuvasta erosta on, että kykenemistä
tarkoittavien *voida-* ja *pystyä-* verbien käy-
tössä menneessä aikamuodossa useimmin

pystyin eilen tekemään jotain, kun taas nykyhetken viitattaessa aivan hyvin *voin tänään tehdä jotain* — *voin tehdä eilen ei kuulosta yhtä hyvältä.*

Mainitsemani aiempi tutkimus on nostanut esille merkittävää ja hyödyllistä tietoa niistä pienistä mutta todellisista eroista, joita samaa tarkoittavien rakenteiden ja sanojen käytössä on, mutta näitä koskien on myös esitetty aiheellista kritiikkiä. Aiemmalle tutkimukselle on ollut tunnusomaista vaihtoehtojen välisen valinnan selittäminen vain yhden ympäristötekijän tai tekijäkategorian avulla sekä rajoittuminen pareittaisiin vaihteluihin. Toinen esitarkastajistani, Stefan Th. Gries (2003) on vakuuttavasti osoittanut, että yhden muuttujan selitysmallit ovat riittämättömiä tai jopa keskenään ristiriitaisia — Griesin mukaan kielenpuhujien ja kirjoittajien tekemät valinnat ovat huomattavasti paremmin selitettävissä monen muuttujan yhteisvaikutuksena — ja olen tästä aivan samaa mieltä. Niin ikään muun muassa Dagmar Divjak ja Stefan Th. Gries (2006) ovat muutama vuosi sitten nostaneet esille sen, että mahdollisia vaihtoehtoja on usein enemmän kuin kaksi.

Tästä on erinomaisena esimerkkinä suomen verrattain vapaa sanajärjestys. Voimme sanoa tekijästä, verbistä ja kohteesta koostuvan lauseen kuudella tapaa, esimerkiksi:

*Hän näki valon.
Hän valon näki.
Näki hän valon.
Näki valon hän.
Valon näki hän.
Valon hän näki.*

Näissä kussakin sanajärjestyksen muutos lähinnä muuttaa painotusta muttei olennaista merkityssisältöä.

Toinen kelpo esimerkki ovat puolestaan tutkimani suomen ajattelemisynonyymit.

Voin periaatteessa aivan yhtä hyvin sanoa

*ajattelevani asiaa,
mieltiväni asiaa,
pohtivani asiaa,
harkitsevani asiaa,
tuumivani asiaa,
tuumaavani asiaa,
punnitsevani asiaa,
puntaroivani asiaa,
kuin funtsivani asiaa,*

vaikka käytännössä väitöskirjassani esittämät tulokset osoittavat, että saattaisin todennäköisimmin *pohtia asiaa*, kun kohteenä oleva sana on luonteeltaan abstrakti.

Samaa merkitsevien rakenteiden tarkastelussa olen päätenyt todennäköisyyspohjaiseen kielikäsitukseen, jonka puolesta ovat puhuneet esimerkiksi Joan Bresnan (2007), Rens Bod, Jennifer Hay ja Stefanie Jannedy (2003) sekä useat muut — kyseinen näkemys ei kuitenkaan ole vielä kauttaaltaan hyväksytty. Todennäköisyyspohjaisen eli probabilistisen kielikäsituksen mukaan valinnat eri käyttöympäristössä — eli tiettyjen selittävien tekijöiden yhdistelmien vaikutuspiirissä — ovat pitkällä tähtäimellä suhteellisia osuuksia, jotka voidaan myös tulkita todennäköisyyksinä. Tämä siis vaikka yksittäiset valinnat yksin tarkasteltuna ovat tietenkin aina diskreettejä. Näinpä kielellisen järjestelmän toiminta — jota kuvataan jollakin teoriolla ja sen sisältämällä selittävillä tekijöillä — ei käytännössä aina tuota kategorisia, poikkeuksettomia valintoja kielenkäytössä. Sen sijaan kielellisen järjestelmän toiminta osoittaa sisäänrakennettua vaihtelua, joka on havaittavissa ajan myötä, riittävän kattavissa aineistoissa. Todennäköisyyspohjaisen käsityksen mukaan kielessä esiintyvä vaihtelu ei olekaan kielenulkoisten mielen prosessien tuottamia »häiriöitä» vaan erottamaton osa kieltä. Rakenteiden todennäköisyydet voivat myös muuttua ajan myötä, niihin vaikuttavien

▷

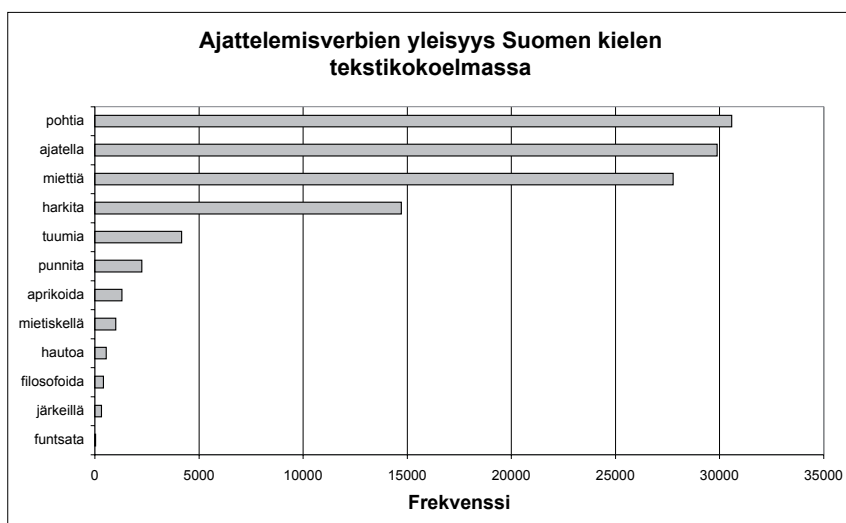
ympäristötekijöiden painoarvojen hienoisesti muuntuessa, mikä puolestaan tekee mahdolliseksi kielen kokonaisuudessaan muuttumisen pikku hiljaa.

Todennäköisyyspohjainen kielikäsitys ei kuitenkaan tarkoita sitä, etteikö kielessä voisi esiintyä käytännössä kategorisia valintoja. Tällöin vain kyseisen valinnan odotustodennäköisyys on 100 prosenttia, tai tarkemmin sanottuna hyvin lähellä yhtä. Esimerkiksi tästä otan infinitiivimuodon valinnan *alkaa*-verbin yhteydessä. Vaikka koulussa olemme oppineet, että meidän tulisi tällöin käyttää ensimmäistä infinitiiviä, vaikkapa *alkaa sataa*, arkipäiväisessä puhekielessä monet meistä käyttävät sen sijaan kolmatta infinitiiviä, siis *alkaa satamaan*. Kun muutama päivä sitten tarkastelin Jyväskylässä ilmestyvän Keskisuomalaisen vuoden 1999 vuosikertaa, sen Suomen kielen tekstikokoelmasta (2001) löytyvässä elektronisessa muodossa, löysin yhteensä 73 tapausta, missä *alkaa*-verbiä seurasi infinitiivimuoto. Näistä 72 kappaletta (eli 99%) noudattaa koulukieliopin mukaista muotoa, ja vain yksi kappale (siis 1%) norminvas-taista, *'sitä pitäisi alkaa kirjoittamaan'*. Tällöin voimme sanoa, että ainakin huolitellussa yleiskielessä, jota sanomalehtiteksti edustaa, ensimmäisen infinitiivin valinta *alkaa*-verbin yhteydessä on liki diskreetti, sillä sen todennäköisyyskin otoksen perusteella on lähes $P=1$.

Selvää vaihtelua voimme sen sijaan havaita suomen vertailurakenteissa, jossa vaihtohtona ovat joko partitiivimuotoista substantiivisia seuraava komparatiivinen adjektiivi, jo mainittu *Pekkaa parempi*, tai

komparatiivimuotoista adjektiivisia seuraava konjunktio *kuin* ja nominatiivinen substantiivi, siis *parempi kuin Pekka*. Näiden kahden kohdalla saatoinkin samaisessa Keskisuomalaisen aineistossa havaita, että partitiivimuotoja esiintyi 214 (eli 87%), esimerkiksi *elokuvaa suuremmat valokuvat, elämää suurempi teatterielämys, normaalia matalampi tulos, leukaakin kiinnostavampi kohta*, kun taas *kuin*-muotoja löytyi 32 kappaletta (eli 13%), esimerkiksi *arvokkaampi kuin ihmishenki*. Vaikka siis partitiivimuoto on selvästi yleisempi, *kuin*-muotoja esiintyy merkittävä määrä: kyseessä on selvä rakennevaihtelu.

En ole tarkastellut näiden nyt esittämieni kahden rakennevaihtelun käyttöympäristöissä esiintyviä tekijöitä, jotka voisivat selittää tehtyjä valintoja, vaan valinnut nämä ensisijaisesti esimerkkeinä rakennevaihteluista, jotka oletan useimpien kuulijoista tuntevan — ja myös tunnistavan niissä valitsevan periaatteellisen samamerkityksellisuuden ja valinnaisuuden. Väitöskirjassani esittämäni tutkimuksen varsinaisena kohteena on synonyymisten sanojen vaihtelu, jonka esimerkkitapauksiksi olen valinnut suomen ajattelemisverbit. Syynä tähän on, että ajattelemisverbit ovat sanakirjakuvausten perusteella helposti keskenään vaihdettavia. Näistä päädyin valitsemaan neljä yleisintä, nimittäin *ajatella, miettiä, pohtia* ja *harkita*; yleisyysarvioni perustin Suomen kielen tekstikokoelmaan (2001), jonka vajaan kahden sadan miljoonan sanan aineistossa näiden neljän yleisyysero muihin on ilmiselvä (Kuva 1).



Kuva 1. Ajattelemisverbien yleisyys Suomen kielen tekstikokoelmassa.

Ajattelemisverbien vaihtelun analyysis-
sa hyödyntämäni monimuuttujamenetelmä
on *polytominen logistinen regressio*. Mitä
nämä kolme termiä tarkoittavat?

Ensinnäkin *polytominen* merkitsee, että
vaihtoehtoisia valintoja (eli lopputulemia)
on enemmän kuin kaksi (nyt tarkastet-
tavassa väitöskirjassani neljä). Toiseksi
logistinen tarkoittaa että kyseisellä mene-
telmällä mallinnetaan *esiintymisosoituksia*
ja ennustetaan *odotustodennäköisyyksiä*.
Kolmanneksi *regressio* on tilastollisen
mallinnuksen laji, jossa mahdollisia lop-
putulemia selittävistä tekijöistä muodos-
tettua mallia (eli sitä kuvaavan yhtälön
muuttujatekijöille arvioitavia *vetokerto-
mia*) sovitetaan havaintoaineistoon siten,
että malli istuu havaintoihin mahdollisim-
man hyvin. Aineistoon hyvin istuva malli
ennustaa liki samat *esiintymisosoitukset* eli
odotustodennäköisyydet tietyllä selittävien
tekijöiden joukolla, kuin mikä on havaitta-
vissa aineistossakin — muttei kuitenkaan
liian hyvin, sillä mallin tulisi toimia myös
uudella aineistolla, jota ei käytetty mallin

alkuperäisessä sovittamisessa. Polyto-
missa logistisessa regressiomallinnuksessa
tekijöille arvioitavat vetokertoimet ilmai-
sevat, kuinka paljon tekijän esiintyminen
aineistossa lisää tai vähentää tietyn loppu-
tuleman esiintymisen todennäköisyyksiä
suhteessa muihin vaihtoehtoihin.

Väitöskirjassani kokeilemistani eri po-
lytomisista logistisista regressiomalleista
pätevin sisälsi kaikkiaan 48 eri selittävää
tekijää. Koska käyttämässäni aineistossa
ajattelemisverbien käyttöympäristössä
esiintyi kuitenkin kerrallaan korkeintaan
11, mutta keskimäärin vain viisi selittävää
tekijää, otan käytännön esimerkiksi pari ai-
toa virkettä ja niiden käyttöympäristössä
esiintyvät selittävät tekijät:

- (1) *Miten*[TAPA:GENEERINEN]
ajatellit[INDIKATIIVI+TOINEN PERSOO-
NA, PILOSUBJEKTI, TEKIJÄ:YKSILÖ]
erota[KOHDE:INFINITIIVI]
... jostain umpimielisestä SAKn
kannattajasta?
[sfnet]
[3066/politiikka_9967]

▷

P(*ajatella*|Konteksti)
 ~ 0.34 ($\approx 1492/3404$)
 + (21:1)·TAPA:GENEERINEN
 + (2:1)·INDIKATIIVI
 + (1:1.6)·TOINEN_PERSOONA
 + (1:1)·PIILOSUBJEKTI
 + (1:1.2)·TEKIJÄ:YKSILÖ
 + (6:1)·KOHDE:INFINITIIVI
 + (1.6:1)·[INTERNET-GENRE]
 =1.0

Tällä seitsemän selittävän tekijän yhdistelmällä *ajatella*-verbin odotustodennäköisyys $P=1.0$, eli tässä meillä on esimerkki käytännössä kategorisesta valinnasta kyseisessä käyttöympäristössä. Tässä muun muassa geneerisen tavan ilmaisun (*näin, noin, siten, miten*) esiintyminen kasvattaa *ajatella*-verbin esiintymistodennäköisyyden kahdenkymmenyhden suhde yhteen (21:1) ja infinitiivikin verbin kohteena kuuden suhde yhteen (6:1). Taasen toisen persoonan esiintyminen pienentää *ajatella*-verbin esiintymistodennäköisyyden viiden suhde kahdeksaan (5:8~1:1.6), mutta se ei oleellisesti pienennä ennustettua *ajatella*-verbin odotustodennäköisyyttä.

- (2) *Tarkastusviraston mielestä*[META-KOMMENTTI]
tätä ehdotusta[KOHDE:TEKO]
olisi[KONDITIONAALI + KOLMAS PERSOONA, PIILOSUBJEKTI]
syytä[VERBIKETJU:VÄLTTÄMÄTTÖMYYS]
pohtia tarkemmin[TAPA:RIITTÄVÄKSI-KOETTU].
 [766/hs95_7542]

P(*pohtia*|Konteksti)
 ~ 0.21 ($\approx 719/3404$)
 + (1:1.2) META-KOMMENTTI
 + (1.3:1) KOHDE:TEKO
 + (1:1.2) KONDITIONAALI
 + (1:1.1) KOLMAS_PERSOONA
 + (1:1.1) PIILOSUBJEKTI
 + (1:1) VERBIKETJU:VÄLTTÄMÄTTÖMYYS
 + (1:1.2) TAPA:RIITTÄVÄKSI-KOETTU
 =0.125

Toisessa, kahdeksan selittävää tekijää sisältävässä virkkeessä *pohtia*-verbin todennäköisyys onkin niin pieni kuin $P=0.125$.

Syynä tähän on, että kaikkien käyttöympäristössä esiintyvien selittävien tekijöiden vetokertoimet ovat hivenen *pohtia*-verbiä vastaan tai tilastollisesti ei-merkitseviä (käytännössä yhden suhde yhteen). Sitä vastoin samat ympäristötekijät suosivat huomattavasti enemmän *harkita*-verbiä, jonka odotustodennäköisyys olisi $P=0.725$; sitä vain ei kirjoittaja ollutkaan tällä kertaa valinnut. Tämä on mainio esimerkki rakenevaihteluiden todennäköisyydellisyydestä. Mainitut odotustodennäköisyydethän tarkoittavat, että kyseissä käyttöympäristössä noin viisi kertaa seitsemästä odottaisi löytyvän *harkita*, mutta kuitenkin yli kerran kahdeksassa myös *pohtia*. Kyseinen esimerkkivirke aineistossani sattui olemaan juuri se tapaus, jolloin harvinaisempi verbi sattui esiintymään.

Käyttämällä polytomista logistista regressiota saatoinkin arvioida kullekin neljästä tutkimastani ajattelemisverbistä vetokertoimet kaikille 48 selittävälle tekijälle, jotka lopulta sisällytin mallinnukseen. Näistä kiinnostavimpia ovat ne, jotka lisäävät kunkin verbin todennäköisyyttä esiintyä. Esitän näistä esimerkinomaisesti vain kaksi eli *pohtia*- ja *harkita*-verbin vetokertoimet, sillä erityisesti niillä on mielenkiintoinen yhteys kyseisten sanojen etymologiseen alkuperään.

Kun *pohtia*-verbin käyttöympäristössä esiintyy ajattelun kohteena suora sitaatti, kasvaa sen vetokerroin kahdeksan suhde yhteen. Jos taas tekijänä on ryhmä kuten *hallitus*, tai kohteena on abstraktio kuten *asia*, tai kieliopilliseen käyttöympäristöön liittyy paikka kuten *kokoushuone*, on vetokerroin *pohtia*-verbin puolesta vielä neljän suhde yhteen. Sanat ja muut inhimillisen kommunikaation välikappaleet ajattelun kohteina edelleen kasvattavat *pohtia*-verbin vetokerrointa kolmen suhde yhteen, kuten myös epäsuora kysymys kohteena. Ajankohdan täsmällinen ilmaus, kuten *nyt*

tai *kello 12*, tai alkamisen ilmaisu verbiketjussa, esimerkiksi *ryhtyä* tai *alkaa*, myös nostavat *pohtia*-verbin vetokertoimen kahden suhde yhteen. Kun *harkita*-verbin käyttöympäristössä esiintyy kohde, joka viittaa tekoon, kuten *lähtö* tai *päätös*, kasvaa *harkita*-verbin vetokerroin yhdeksän suhde yhteen. Ehdon ilmaisu *jos*-lauseena lisää vetokerrointa vähemmän mutta silti kolmen suhde yhteen. Ehdollisuuden merkintä puolestaan konditionaalimuotoisena, eli *harkitsisi* ja vastaavat muodot, kasvattaa verokerrointa vielä kahden suhde yhteen.

Tällaisten verbikohtaisten selittävien tekijöiden vetokertoimien perusteella saatoin luoda niille kullekin yleiset karakterisoinnit, jotka erottavat verbit toisistaan. *Ajatella* on ajallisesti määrittelemätön ja rajaamaton, pysyvämpiin aikomuksiin tai mielipiteisiin viittaava verbi, ja usein se ilmaisee yksittäisen ihmisen toista ihmistä tai ihmisiä koskevaa ajattelua. *Miettiä* puolestaan kytkeytyy myös yksittäisen ihmisen ajattelutoimintaan, mutta on ajalliselta kestoiltaan määritellympi. *Pohtia* taas istuu kuvaamaan yhteisön kollektiivista ajattelua, joka ei henkilöidy yksittäiseen ihmiseen ja joka usein koskee abstraktia asiaa. Viimeiseksi *harkita*-verbin kohdalla ajattelun kohde on usein tulevaisuudessa ajatteluhetkestä katsottuna, ja monasti on kysymyksessä päätös tai valinta, johon mahdollisesti liittyy toistettavuus ja ehdollisuus.

Nämä karakterisoinnit liittyvät mielenkiintoisella tavalla ajattelemisverbien historiallisiin alkuperiin, jotka liittyvät kansamme ja kulttuurimme agraaristaan, metsästykseseen, kalastukseen ja maanviljelyyn. *Pohtia*-verbin varhaisempi käyttö, joka tunnettiin ensisijaisena vielä viime vuosisadan alkupuolella, siis *erottaa jyvät akanoista*, liittyy elonkorjuuajan yhteisöllisiin toimiin. Tämä yhteisöllisyys on

edelleen säilynyt *pohtia*-verbin nykyisessä abstraktimmassa merkityksessä. Myös *harkita*-verbin kohdalla meidän on helppo nähdä sen varhaisemmassa käytössä — jonkin naaraamisessa veden pohjasta — sellaista perusteellisuutta ja toistoa, joka niin ikään löytyy sanan abstraktimmasta nykymerkityksestä. Nämä kaksi esimerkkiä kytkevät tämän lektioni yhdeksän vuotta taaksepäin ja osoittavat, miten silloiset pohdiskeluni kielen ja kulttuurin yhteydestä sekä kulttuurin kehityksen heijastumista kielen ja sen sanojen merkityksen muutokseen voivat käytännössä ilmetä.

Yhteenvetona totean, että käyttämässäni aineistossa esiintyy valitsemieni selittävien tekijöiden yhteydessä selvää vaihtelua neljän liki samaa merkitsevien ajattelemisverbien välillä. Aineistossa havaitut valinnat ovat merkittävässä määrin — mutteivät kokonaan — selitettävissä käyttämieni perinteiseen kielitieteelliseen analyysiin pohjautuvien selittävien tekijöiden avulla. Jäljelle jäävää vaihtelua ei käsittäakseni kuitenkaan voine selittää kattavasti vain lisäämällä selittävien tekijöiden joukkoa. Mitä tällaisen rakenne- tai sanavaihtelun tilastolliseen analyysiin ja mallintamiseen tulee, polytominen logistinen regressio sopii mielestäni siihen hyvin: sen tuloksia on suoraviivaista tulkita.

Tutkimukseni tulokset siis tukevat todennäköisyyspohjaista näkemystä vuorovaikutussuhteesta kielenkäytön sekä kielen kuvausmallien ja teorioiden välillä. Tämän näkemyksen mukaan harvat kielelliset valinnat ovat kategorisia tietyssä käyttöympäristössä, vaan useimmissa tapauksissa esiintyy eriasteista variaatioita mahdollisten lopputulemien suhteen. Tämä kielellinen vaihtelu tulee havaittavaksi lopputulemien suhteellisina osuuksina, kun käytössä on riittävän laajat ja kattavat aineistot. ■

▷

AINESLÄHTEET

- SFNET 2002–2003: Suomenkielistä Internetin sfnet-uutisryhmäfoorumilla lokakuusta 2002 huhtikuulle 2003 käytyä keskustelua (n. 100 miljoonaa sanaa). Koostanut Tuuli Tuominen ja Panu Kalliokoski, Helsingin yliopisto, tietotekniikkaosasto, ja Antti Arppe, Helsingin yliopisto, yleisen kielitieteen laitos. <http://www.csc.fi/kielipankki/>.
- Suomen kielen tekstikokoelma [Finnish Text Collection] 2001: Suomenkielistä sanomalehti-, aikakauslehti- ja kaunokirjallista tekstiä 1990-luvulta (n. 180 miljoonaa sanaa jakautuen 97 alikokoelmaan). Koostanut Kotimaisten kielten tutkimuskeskus, Helsingin yliopisto, yleisen kielitieteen laitos, ja Joensuu yliopisto, kielten laitos. <http://www.csc.fi/kielipankki/>.

MUUT LÄHTEET

- BOD, RENS – HAY, JENNIFER – JANNEDY, STEFANIE (toim.) 2003: *Probabilistic linguistics*. Cambridge: MIT Press.
- BRESNAN, JOAN 2007: Is syntactic knowledge probabilistic? Experiments with the English dative alternation. – Sam Featherston & Wolfgang Sternefeld (toim.), *Roots: Linguistics in search of its evidential base* s. 75–96. Berlin: Mouton de Gruyter.
- DIVJAK, DAGMAR – GRIES, STEFAN TH. 2006: Ways of trying in Russian: Clustering and comparing behavioral profiles. – *Corpus Linguistics and Linguistic Theory* 2:1 s. 23–60.
- GRIES, STEFAN TH. 2003: *Multifactorial analysis in corpus linguistics: A study of particle placement*. London: Continuum.
- PERKINS, REVERE D. 1992. *Deixis, grammar, and culture*. Amsterdam: John Benjamins.

ANTTI ARPPE *Univariate, bivariate, and multivariate methods in corpus-based lexicography: A study of synonymy*. Helsingin yliopiston yleisen kielitieteen laitoksen julkaisuja 44. Helsinki: Helsingin yliopisto. <http://urn.fi/URN:ISBN:978-952-10-5175-3>.

Yhteystiedot:

Yleisen kielitieteen laitos, PL 9 (Siltavuorenpenger 20 A), 00014 Helsingin yliopisto
Sähköposti: etunimi.sukunimi@helsinki.fi