

JAKSOTUSTESTIN TULOKSIA

Yrjö Arajuuri *Kieli, kirjoitus ja lukija. Tekstin kopiointiprosessi kielen rakenteen valottajana.* Suomalaisen Kirjallisuuden Seuran Toimituksia 660. SKS, Helsinki 1997. 244 s. ISBN 951-717-970-7.

Yrjö Arajuuren väitöskirja on empiirinen tutkimus, joka perustuu kokonaan koehenkilöillä suoritettuun tekstin kopiointitestiin. Testi on tekijän itsensä kehittämä ja toteuttama. Koehenkilöt (yhteensä 711 ylioppilastasoista henkilöä) kopioivat seinälle heijastetun lyhyehkön tekstin siten, että merkitsivät vinoviivan tekstiin siihen kohtaan, jossa joutuivat katsomaan mallitekstiä kopiointia jatkaakseen. Nämä vinoviivojen väliset jaksot, joita tekijä nimittää tulkintajaksoiksi, ovat analyysin kohteina. Testijärjestelyt ja -proseduurin tekijä on hoitanut ilmeisen korrektisti.

Eri tekstejä ja samalla eri testejä on 16. Kussakin on keskimäärin 45 sanaa, ja kaikkien yhteinen sanamäärä on 719. Kukin koehenkilö on osallistunut vain yhteen testiin kahta tapausta lukuun ottamatta, joissa samat henkilöt osallistuivat samanaikaisesti kahteen testiin. Tutkimuskohteina olevia tulkintajaksoja on aineistossa yhteensä 8 212.

Teoreettisia peruskäsitteitä ja teoreettista viitekehystä väittelijä esittelee kahdeksan sivun verran. Taustateoriakseen hän ilmoittaa kognitiivisen kieliopin. Alumpana hän esittää lähtökohdikseen lukutapahtumatutkimusten tuloksia. Testien suunnittelun lähtökohdana on ollut se tieto, että lukija ottaa lukutapahtumassa katseellaan kielelliseen tulkintaansa ja pikamuistiinsa tietynmittaisen osan riviä. Tutkimuksella on siis yhteys muistin ja lukemisen psykologiaan, vaikka tekijä itse sanoo kyllä — ja toistaa useaan otteeseen —, ettei työ pyri olemaan (psy-

kologinen) lukemistutkimus eikä painopiste ole psykolingvistiikassakaan vaan lingvistiikassa: »Se ei pyri jäljittämään tekstin kielellistä prosessointia vaan keskittyy siihen, millä tavoin kielelliset ilmiöt korreloivat sen kanssa, miten koehenkilö jaksottaa tekstin sitä lukemisen jälkeen kopioidessaan.»

Työ ei ole teoreettinen, vaan se on puhtaasti empiirinen, kokeellinen tutkimus. Tämänkaltainen väitöskirja on kielitieteessä harvinainen. Se on aineistokeskeinen, mutta se ei sisällä todellista kielenkäyttöaineistoa. Se pyrkii teoreettiseen päämäärään, mutta se ei ole teoreettinen. Se ei kuulu mihinkään kielitieteen pääalaan tai -suuntaan (eikä sen toki tarvitsekaan kuulua). Kokeellista kielen kohdistuvaa tutkimusta on meillä harastettu lähinnä vain fonetiikassa, psykolingvistiikassa ja logopediassa.

Väitöskirjan perusasetelma ja testi on varsin yksinkertainen ja teoreettinen tavoite suhteellisen kapea-alaiseksi rajattu. Keskeisenä tutkimusongelmana ja tavoitteena on sen selvittäminen, voidaanko testin avulla todistaa, onko kielessä sanaa pitempiä vakiintuneita yksiköitä ja jos on, niin millaisia, tai — päinvastaisesti tarkastellen — määräytyvätkö jaksot tällaisten yksiköiden perusteella. Erityisesti tarkastellaan lauseja virkekokonaisuuksia, lausekkeita ja idio-meja.

Onko tekstin jaksotuksessa sitten kysymys enemmän mekaanisesta muistista vai lingvistisestä rakenteellisesta hahmotuksesta ja tulkinnasta? Mm. seuraavat kohdat kertovat, että lähtökohdat ja kopiointijakso-

jen syntymisen syyt ovat ei-lingvistisiä:

[Normaalissakin lukutapahtumassa] »Jaksojen syntymisen syiksi esitetään pääasiassa kaksi inhimillistä tekijää: näkökenttämme on rajallinen, ja pikamuistin kapasiteetti on pieni» (s. 23); »Voidaankin olettaa, että tuo testitehtävän edellyttämä kirjoitustauko tulee pidettäväksi, kun lukijan pikamuistin kapasiteetti ei enää anna mahdollisuutta jatkaa lukemista. – – Varsinkin sellaisten koehenkilöiden, jotka ovat läpi testin käyttäneet hyvin lyhyitä jaksoja, voisi olettaa lukeneen tekstiä hyvin mekaanisesti, lähes sana sanalta.» (S. 70.); »Tässä esitetyt hajahavainnot osoittavat, että jaksoilla on taipumus muodostua kutakuinkin samanmittaisiksi, lähestyä jonkinlaista optimituutta» (s. 118–119).

Tästä lähtökohdasta katsoen jaksojen rajat syntyvät kopioidessa mekaanisen satumanvaraisesti eivätkä kielellisen rakenteen mukaisesti. Yhden koehenkilön kommentit (s. 207) kertovat, että sisältöä ei kopioidessa juurikaan memoroida. Parhaiten omaksutaan kestonmuistissa olevat kokonaisuudet, kuten idiomit. Vaikka virkkeen raja on vahva tulkintajakson raja, jakson ei tarvitse noudattaa sitä, esim. »/Kolmetoista ihmistä pelastettiin. Helikopteri/» (s. 98, 142). Samoin jakson raja voi olla keskellä sanaa (s. 103). Tämän voi tulkita siten, että pyrkimyksenä on kopioida tekstiä kerralla mahdollisimman paljon, ja pysähdys tulee siihen kohtaan, jossa muisti ei enää riitä. Tässä on tietenkin yksilöllisiä eroja, ja sivulla 58 on esimerkki hyvämuistisen koehenkilön pitkistä jaksoista. Näkömuistilla on varmasti merkittävä osuus.

Testeistä kuudella on julkilausuttuna tavoitteena selvittää, miten jokin seuraavista tekijöistä lukutapahtumassa vaikuttaa jaksojen pituuteen (ei laatuun): syntaksin ja koherenssin puuttuminen, sanaston abstraktisuus ja konkreettisuus, runsas pisteiden esiintyminen tai välimerkittämyys ja rivin-

loppuinen sananjako. Näillä testeillä ei siis pyritä jaksojen kielellisen yksikköluonteen selvittämiseen vaan erilaisten kielellisten ja ortografisten tekijöiden vaikutuksen mittaamiseen. Lisäksi koko aineistosta on havainnoitu tekstin alun ja lopun sekä rivijaon vaikutusta. Mielenkiinnon kohteena on siis muistijakson pituus, jolla sinänsä ei ole lingvististä relevanssia. Kaikilta näiltä osin työ liittyy sellaiseen lukemis- ja luettavuustutkimukseen, jossa on lingvistikseen relevanttia lähinnä vain se tieto, että rakenteellisesti yksinkertainen, koherentti, sanastoltaan konkreettinen ja välimerkitty teksti on helpompi omaksua (jolloin myös jaksot jäävät muistiin pitempinä).

Tämän edellä mainitun pitkän sivujuonteisen vaikutustutkimuksen status suhteessa päätavoitteeseen ei tule selkeästi esille. Väitöskirjan tehtävänmäärittely jää kaksinaiseksi ilmeisesti siksi, että testit oli suunniteltu ja toteutettu toisenlaisesta, lähinnä luettavuustutkimuksellisesta viitekehuksesta ja tavoitteesta käsin ja kognitiiviset peruskäsitteet ja sen mukainen viitekehys on keksitty vasta myöhemmin. Näin nämä kaksi tavoitetta eivät tekstityksessä muodosta koherenttia kokonaisuutta, ja se näkyy myös kirjan rakenteessa.

Kognitiivisen teorian pohjalta asetettua päätavoitetta palvelee testeistä seitsemän. Niiden avulla pyritään selvittämään idiomien ja sen korvaavan ilmauksen, yhdyssanojen, pitkän attribuuttikasauman ja pitkän virkkeen jaksotusta. Koska niissä on kysymys kielellisistä rakenteista, niiden jaksotuksesta ja jaksojen laadusta, ne ovat jakson yksikköluonteen kannalta relevantteja testejä. Kolmella testillä ei ollut spesifistä tavoitetta.

On ilmiselvää, että aineistossa on lingvististen perusteiden mukaan syntyneitä, tulkintaa edellyttäviä jaksoja. Selvimpiä ovat tapaukset, joissa alkuperäinen ilmaus on kopiointiohjeesta huolimatta korvattu

uudella sama- tai lähimerkityksisellä ilmauksella, esim. kun *omaan maahansa nähden* saa muodon *oman maansa suhteen*. Ne osoittavat, ettei tekstin käsittely ole ainakaan kaikissa kohdissa mekaanista kopiointia. Näistä syistä tekijä nimittää aineistonsa (memorointi- tai kopiointi)jaksoja tulkintajaksoiksi. Koska mukana on joka tapauksessa mekaanisia jaksotuksia, joissa ei ole tulkinnasta kysymys, nimitys ei sovi kaikkiin tapauksiin. Tekijä sanookin määrittelyssään, että nimitys ei implikoi, että jaksotus määräytyisi pelkästään tulkinnan perusteella, mutta sanoo kuitenkin toisaalta liian yleistävästi, että jaksosten erottaminen ei ole sattumanvaraista (s. 28).

Olellainen kysymys tämän jälkeen on se, mitkä kaikki jaksot ovat tulkinnan perusteella syntyneet ja mitkä ovat sattumanvaraisia ja mikä on niiden määrällinen suhde. Tärkeintä on siis jaksosten laadullinen analyysi. Mutta siitä ei ole kattavaa kuvausta. Laadullinen variaatio jää epämääräiseksi. Paljon tekstiä on käytetty jaksosten pituuden variaation käsittelyyn (joka kertoo jostain lukemis- ja kopiointitapahtumasta), mutta laadun kuvaus olisi kiinnostavampaa. Tähän ei aineisto tällaisenaan anna mahdollisuuksiakaan. Olisi kannattanut tehdä läpikotainen syntaktinen analyysi (hierarkia sulkein ilmaistuna), jolloin olisi ollut helppo tietokoneella tulostaa kaikki jaksotyypit ja niiden jakauma. Olisi voinut täsmällisesti nähdä ja päätellä, mikä osuus on yksikön luonteisilla jaksoilla. Liitteenä oleva aineisto paljastaa ainoastaan rajakohtien summauksen eikä sitä, millaisia jaksoja yksittäisissä testisuorituksissa on. Kun koko aineiston kuvauksessa pääasiana on siis jaksosten pituus, jaksotyypit jäävät näkymättömiin.

Kuinka vahva merkitys yksikköydellä aineiston perusteella on? On koko joukko erilaisia jaksoja, jotka eivät muodosta tai joihin ei sisälly minkäänlaista kielellistä

yksikköä. Niiden suhteellista osuutta ei ole esitetty. Yksikköys on tietysti suhteellinen ja löyhärajainen käsite, mutta esimerkiksi seuraavat summatuista tekstiliitteistä poimimani jaksot, joita kukaan ei ollut edelleen jakanut, ovat varsin löyhiä (vrt. kokonaislistaan s. 166–168): *Jo tänään on, ja on, Ja siellä, Tosin tällainen, ja toteamaan, On myös, sellaisenaan käyttää loppuun, Mikäli arvonmääritys*. Seuraavat vahvempien rajojen edessä tai jälkeen esiintyvät jaksot on koehenkilöiden suuri enemmistö jättänyt jakamatta (numero ilmaisee ne muutamia tapauksia, joissa on merkitty raja numeron ilmaisemaan kohtaan): *meille 1 on, elintärkeätä 3 se, kuvioihim 1 me. 1 Siellä, Mutta 1 yrittänyttä, että 4 vanha, On 2 ravintoarvoltaan, Tai 4 oikein, sekä 4 rapeita, joskaan 5 hän, kuvan 5 kuin 5 Dragsfjärdin*.

Seuraavankaltaiset ovat sellaisia päinvastaisia tapauksia, joissa on runsaasti (ehkä noin puolella ja joissain tapauksissa selvällä enemmistöllä koehenkilöistä) merkittyjä rajoja, vaikka sanat rajan kahden puolen kuuluvat varsin kiinteästi yhteen: *sijoitettava 18 tulevaisuuteen, elinympäristömme 20 suojelun, varmasti 17 elintärkeätä, jälkeläisillämme 36 on, on 18 lähdeittävä, Siellä 34 missä, tekee meille 15 Kotisaari, käy 16 puhuminen, varannot 35 kestävät 26 lyhytaikaisesti, nousu 18 on 19 ainakin 15 muutamaksi 15 vuodeksi, määritysmenetelmien 18 soveltamista, ei 1 olisi 23 mahdollista, samanlaisten 25 tuontitavaroiden, jälleenyntihinnasta 29 käsin*.

Näissä viimeksi mainituissa on sellaisia, joita tekijä pitää vakiintuneina sanaa laajempina yksiköinä. On tietysti selvää — kuten olen edellä jo koko aineiston osalta yleisesti todennut —, että tällaisessa testissä myös kiinteitä yksiköitä on jaettu sattumanvaraisesti. Ovatko esimerkiksi yllä mainitut luvut sattumanvaraisia? Mikä on erilaisten jaettujen ja jakamattomien yksiköiden tai »tulkinnallisten» ja »sattumanvaraisten»

rajojen määrällinen suhde tai erityyppisten jaksojen pituuksien suhde, ja mitä se todistaa?

Tekijä on verrannut (yksikönluonteiseksi katsomansa) idiomin sisältävien jaksojen keskimääräistä pituutta koko aineiston keskiarvoiseen jaksonpituuteen tai vastaavasti lausekkeen sisäisten rajojen määrää keskiarvoiseen sananvälisten rajojen määrään (ei sen sijaan lausekkeen sisältävän jakson pituutta keskiarvoiseen jakson pituuteen tai muuntotyypisten jaksojen pituuteen tai pelkästään lausekkeesta muodostuvien jaksojen määrää ja pituutta muihin verrattuna) ja todennut, että tällaiset yksiköt jakautuvat selvästi keskiarvoa harvemmin. Tekijä on päässyt tähän tulokseen poimimalla aineistosta parhaat ja selvimmät tapaukset päältä ja ottamalla lausekkeista huomioon vain kaksisanaiset. Laskutavat ja päätelmät ovat oikeita, mutta edellä mainitsemani läpikotaisempi vertailu olisi antanut täydellisemmän ja monipuolisemman kuvan koko aineistosta, kaikista jaksotyypeistä.

Idiomit, postpositio- ja substantiivilausekkeet sekä verbiketjuista liittotempukset ovat selvimpiä yksikköjä sillä perusteella, että ne säilyvät parhaiten jakautumattomina. Verbilauseke on kokonaisuutena paljon hatarampi. Subjekti ja predikaatti eivät yleensä muodosta jaksoa, mutta kiinteyteen vaikuttaa verbin nominaalisten osallistujien ilmauksellinen pituus. Aineistoa ei kuitenkaan ole ollut riittävästi kaikkien lauseketyyppien tai toisaalta erilaisten kollokaatioiden käsittelyyn. Verbin valenssin huomioon ottaminen tai yleensä verbin nominaalisten osallistujien ja myös niiden semanttisen roolin huomioon ottaminen riittävän aineiston valossa vaatisi yksikköluonteen asteittaisuuden selvittämiseksi paljon laajempaa kokeellista tutkimusta.

Tekijä olisi voinut pohtia myös ns. fiksaatiopisteiden merkitystä sen tutkimustiedon pohjalta, että katse pyrkii mieluummin

pysähtymään tekstissä esiintyvien lyhyiden rakennesanojen kohdalle. Miten hänen aineistossaan rajat osuvat, millaisten konstituenttien välissä ne ovat? Lisähuomion kiinnittäminen rajakohtiin olisi ilmeisesti valaissut jaksonmuodostusongelmaa ja jaksojen laatua. On oletettavissa, että informaatioltaan painavimmat (sisältö)sanat sijoittuvat yleensä eri jaksoihin, niin että raja sijoittuu usein tällaisen sanan eteen tai jälkeen (vaikkapa juuri kevyemmän muotosanan kohdalle). Tematiikalla luulisi olevan merkitystä.

Olisi ollut kiinnostavaa jossain muodossa verrata jaksotusta puheen jaksotukseen, onko niissä yhtäläisyyksiä ja korrelaatiota, ja siihen, millaisia mahdollisia yksikköjä puheen rytmijaksot ja puhetahdit ovat, vaikka tämä onkin kyllä jo oman tutkimuksen aihe. Valmista litteroitua vertailuaineistoa ja teoriapohjaa olisi.

Kokonaisuutena voi sanoa, että väitöskirjatyön lähtökohta ja kokonaisuasetelma on innovatiivinen, mutta teoreettisesti kirja jää kevyeksi. Kytkeä kognitiiviseen teoriaan jää viittauksenomaiseksi. Lähdekirjallisuuden käyttö jää niukaksi, ja se on selvä puute. Keskeinen ongelma on, mitä relevantteja tuloksia testeistä saadaan ja mikä on tulosten merkitys. Testit antavat viitteitä siitä, että tulkintajaksojen pituus riippuu tekstin rivijaosta, välimerkeistä, sijainnista tekstissä (alussa, lopussa, sisällä), sanojen pituudesta ja koheesiosta, sanaston abstraktiotasosta ja tuttuudesta, syntaktisesta rakenteesta, kollokaatioiden vaikiintuneisuudesta ja idiomeista. Toisin sanoen mainitut seikat vaikuttavat memorointiin ja sitä kautta luettavuuteen. Näitä sekä aineistoa, aineiston käsittelyä, jaksotusta ja sen variaatiota, jaksoja ja niiden pituutta selostetaan ja kuvaillaan puolen kirjan verran hyvin seikkaperäisesti. Teksti on ulko-kohtaista deskriptiota — hyvin helppolu-kuista —, ja se sisältää toistetta ja yksinker- ▶

taisen epäkiinnostaviakin sekä päätavoitteen kannalta perifeerisiä asioita.

Päätuloksina korostetaan sitä, että kielessä on sellaisia melko vakiintuneita yksiköitä, kuten lausekkeita, jotka eivät tavallisesti jakaudu eri jaksoihin vaan jotka memoroidaan yleensä kokonaisuuksina. Voi kysyä, kuinka uusi asia se on. Mutta tärkeää on joka tapauksessa ollut todistella asiaa kokeellisesti. Kiinnostavinta antia ja kokonaan uutta tieteellistä tietoa kirjassa on erilaisten yksiköiden kiinteysasteen erittely. Tulkintajaksojen yksityiskohtaisempi laadullinen erittely ja tyypittely sekä määrälliset vertailut eivät ole kuitenkaan systemaattisia vaan ovat jääneet tärkeimpien yleisten huomioiden varaan. Toisaalta aineisto osoittautuu riittämättömäksi pitämälle menevien päätelmien ja täsmällisem-

pien typologioiden kannalta. Kvantitatiiviset havainnot on suhteutettu keskiarvoiseen jaksonpituuteen. Olisi voinut käyttää sofistikoituneempiakin menetelmiä. Olisi ollut mahdollista mitata eri tekijöiden vaikutus jakson muodostumiseen. Keskiarvoiseen jaksonpituuteen vertaaminen oikeuttaa kyllä niihin päätelmiin, joita väitöskirjassa esitetään, ja sen perusteella voidaan uskoa, että ehjimmät sanoja suuremmat tulkinnalliset kokonaisuudet ovat pisteiden ja pilkkujen erottamat sekä idiomit ja substantiivi- ja postpositiolausekkeet. ■

PAULI SAUKKONEN

*Kotimaisten kielten tutkimuskeskus,
Sörnäisten rantatie 25, 00500 Helsinki*
Sähköposti:
Pauli.Saukkonen@Domlang.Fi