

Väitöstutkimus morfologisesta produktiivisuudesta

Alexandre Nikolaev: *Suomen nominien taivutusjärjestelmän produktiivisuudesta. Korpuslingvistisiä ja kokeellisia tutkimuksia.* Väitöskirja, Itä-Suomen yliopiston filosofinen tiedekunta, Joensuu 2011. 260 s.

Alexandre Nikolaevin yleisen kielitieteen alaan kuuluva väitöskirja *Suomen nominien taivutusjärjestelmän produktiivisuudesta: Korpuslingvistisiä ja kokeellisia tutkimuksia* tarkastettiin Itä-Suomen yliopiston filosofisessa tiedekunnassa Joensuussa 1.4.2011. Nikolaevin lektio aiheesta ”Substantiivien taivutusjärjestelmän produktiivisuudesta” on julkaistu *Virittäjän* verkkoliitteessä 2/2011.

Kyseessä on artikkeliväitöskirja, jonka kuusi alkuperäisartikkeliä ovat yhteispi-tuudeltaan 120 sivua. Artikkeleista Nikolaev on kirjoittanut yhden yksin (11 sivua), ja muut viisi artikkeliä ovat yhteisartikkeleita, joiden kirjoittajien joukossa joka artikkelissa on Nikolaevin ohjaaja Jussi Niemi. Kolmessa artikkelissa kirjoittajat ovat Nikolaev ja Niemi kahdestaan, yhdessä lisäksi yksi kirjoittaja ja vielä yhdessä lisäksi kuusi muuta kirjoittajaa. Artikkeleista kolme on ilmestynyt *Virittäjässä*, yksi julkaisussa *Puhe ja kieli*, yksi kansainvälisessä kokoomateoksessa ja yksi oli tarkastuksen yhteydessä arvioitavana julkaisemista varten. Julkaisuforumit ovat hyviä, ja useimmissa tapauksissa artikkelit ovat käyneet läpi refereer-arvioinnin.

Yhteisartikkelit eivät suomalaisissa kielitieteen artikkeliväitöskirjoissa ole olleet yleisiä. Ottaen huomioon artikkeleiden lyhyehkön yhteen lasketun sivumäärän ja monet kirjoittajat, herää kysymys, onko väittelijän itsenäinen panos riittävä. Nikolaev tekee kuitenkin riittävästi selkoa omasta panoksestaan, joka kohdistuu erityisesti laajojen korpusanalyysien tekemiseen sekä relevanttien lingvististen teorioiden (kuten optimaalisuusteorian) ja psykolingvististen kokeellisten menetelmien integratiiviseen soveltamiseen. Juuri tämä tekeekin väitöskirjasta riittävän yleislingvistisen. Toinen osoitus Nikolaevin omasta panoksesta on pitkä (melkein ylipitkä eli 101-sivuinen) johdanto- ja yhteenvetoluku, joka kiinnostavasti nivoo yhteen vuosikymmenen ajan kestäneen väitöskirjatyon eri vaiheet, menetelmät ja tulokset ja lisäksi luo katsauksen suomen taivutusjärjestelmän kuvauksen vaiheisiin Petraeuksesta alkaen.

Lektion substantiiveihin perustuva nimi kuvaa väitöskirjan sisältöä osuvammin kuin yhteenvedon mainitsemat nomininit. Käytännössä Nikolaev käsittelee substantiiveja; nominien alaluokistahan kaksi (pronominit ja osin lukusanat) ovat itse asiassa täysin epäproduktiivisia.

Artikkelit käsittelevät suomen taivutusmorfologian kuvausta ja prosessointia sekä teorioiden että metodien näkökulmasta. Koeasetelmiin minulla ei ole oleellista huomauttamista. Nikolaevin lukeneisuus ja teoreettis-metodinen hal-

linta on kokonaisuutena ottaen kohdallaan. Kvantitatiivisia menetelmiä hän soveltaa asiantuntevasti, mikä osin kokeellisessa työssä tietenkin on välttämätöntä. Lingvistisen teorian näkökulmasta kummastuttaa, ettei tekijä kertaakaan viittaa yhteen morfologisen produktiivisuusteorian perusteeseen, Laurie Bauerin jo vuonna 2001 ilmestyneeseen kirjaan *Morphological productivity*, joka käsittelee monipuolisesti produktiivisuuden yleistä ongelmaa ja sen pitkää tutkimushistoriaa.

Kokeellisia tutkimusasetelmia lukuun ottamatta suurin osa väitöskirjan empiirisestä aineistosta on peräisin Kielipankin (www.csc.fi) suomenkielisistä korpuksista (laajimmillaan yli 130 miljoonaa sananmuotoesiintymää), jotka voittopuolisesti koostuvat 1990-luvun sanomalehtiaineistoista. Esitarkastusvaiheessa huomautin, että tällaisten automaattisen morfologisen analyysin läpikäyneiden mutta koodaukseltaan tarkistamattomien korpusten luotettavuus tulisi perusteellisesti arvioida. Pistokokeita tehdessäni olin huomannut muun muassa, että Nikolaev tuolloin ilmoitti *veri*-substantiivilla olevan aineistossa 10 monikon genetiivin esiintymää *verten*. Tarkistaessani nämä osoittautuivat kuitenkin kaikki *varten*-postposition väärin kirjoitetuiksi muodoiksi. Lopulliseen väitöskirjaan (johdanto s. 19) tämä yksityiskohta on korjattu ja Nikolaev on lisännyt alaviitteen, missä hän yleisesti huomauttaa painovirheiden ja koodausvirheiden olemassaolosta. Silti lukija toivoisi saavansa tarkempia tietoja käytetyn korpuksen luotettavuudesta juuri Nikolaevin tutkimusongelman kannalta. Lisäesimerkkinä voisi mainita sananmuodon *untēn*. Sitä löysin käytetyistä korpuksista nyt tarkistettaessa 73 esiintymää

(Nikolaevilla 77 kpl). Kourallinen on saksan joko oikein kirjoitettuja *untēn*-adverbin tai väärin kirjoitetun *unter*-preposition esiintymiä, lopuista melkein kaikki leksikaalistuneen idiomin *untēn mai/lla, -lle* esiintymiä. Pelkkä luku 73 (77?) ei siten anna realistista kuvaa *untēn*-muodon morfologisesta asemasta, koska oikeiden ja ei-idiomaattisten esiintymien määrä on korkeintaan 6–7.

Useissa artikkeleissaan Nikolaev käsittelee kielenkäyttäjien mentaalisia morfologisia representaatioita. Aineistoon liittyen lukijaa jää vaivaamaan se, ettei Nikolaev kunnolla pohdi, kuinka suurella varmuudella Tekstipankin kirjoitettujen tekstien morfologisen analyysin pohjalta voidaan päätellä kielenkäyttäjien sisäisiä representaatioita, jotka ensisijaisesti perustuvat varhaislapsuudessa omaksumtuun puheeseen ja sen rutiineihin. Tämä on Per Linellin (2005) mainitseman kielitieteen ”written language biasin”, kirjoitetun kielen suosimisen ilmentymä.

Vanhin ja Nikolaevin yksin kirjoittama on artikkeli ”Eräiden suomen kielen taivutustyyppien produktiivisuudesta” (*Puhe ja kieli* 22, 2002: 113–124). Kirjoittaja tarkastelee produktiivisuuden omaksumista kahdessa produktiivisessa versus epäproduktiivisessa paradigma-parissa (*varis – kirves; lasi – kivi*) sekä kolmanneksi kvantitatiivisen astevaihtelun osalta, missä siis vaihtelullisuus on produktiivista (*hattu : hatun*). Epäsanojen taivutuskokeisiin osallistui kolmeen ikäryhmään kuuluvia (3–5-, 10–11- ja 13–14-vuotiaita) syntyperäisiä suomalaisia ja venäläisiä lapsia. Syntyperäiset suomenkieliset valitsivat odotuksenmukaisesti produktiivisen taivutusstrategian, vaikka 3–5-vuotiaat jättivätkin melkein kolmanneksen epäsanoina taivuttamatta. Venäläisillä tendenssi on vielä vahvempi.

Perustulokset ovat uskottavia ja mielenkiintoisia. Yritys selittää kvantitatiivisen astevaihtelun produktiivisuutta tavumäärän vakioimisella ei silti vakuuta, muun muassa siksi, että useimmissa taivutusmuodoissa tavumäärä onkin eri kuin yksikön nominatiivissa ja genetiivissä sekä monikon nominatiivissa (vrt. kaksitavuiset *hattu* : *hatun* : *hatut*, mutta kolmitavuiset *hatussa* : *hatuissa* : *hatuista* : *hatulla* : *hatuilla* : *hatulta* : *hatuilta* : *hatulle* : *hatuille* jne.)

Artikkelissa ”Suomen nominien taivutuksesta: rytmi-, sivupaino ja agglutinaatiohypoteesien testausta” (*Virittäjä* 109, 2005: 530–553) Nikolaev ja Jussi Niemi analysoivat uudelleen suomen taivutusopin klassista ongelmaa, kolmi- ja useampitavuisen nominien ns. vapaata vaihtelua monikon genetiivissä ja partitiivissa. Arto Anttila sovelsi väitöskirjassaan (1997) ensimmäisenä optimaalisuusteoriaa näiden ongelmien analyysiin käyttäen 1,3 miljoonan kokoista sananmuotoaineistoa. Uutta Nikolaevilla ja Niemellä on erityisesti 109 miljoonan sananmuodon tekstipankkiaineiston käyttö, mikä tarjoaa entistä luotettavamat tiedot vaihtelevien muotojen esiintymisyleisyydestä. Käytössä olevat rajoitteet ovat ”vältä kahden tasavahvuisen tavun vierekkäisyyttä”, ”vältä heikkoja sivupainollisia tavuja” ja ”vältä vartalonmuutoksia suffiksaatiossa”. Optimaalisuusteorian mukainen rajoitteiden kilpailu osoittautuu hyvin toimivaksi menetelmäksi ennustaa pintavaihtelun preferenssejä. Malli ontuu vain *paperi*-tyyppisissä sanoissa.

Artikkelissa ”Nominien paradigmaattistuminen suomessa: Millä rakenteellisilla ehdoilla kielenkäyttäjät sitovat potentiaaliset nominin taivutusluokkiin?” (*Virittäjä* 110, 2006: 46–69) Nikolaev ja

Niemi vertaavat *patsas*- ja *keskus*-taivutustyyppien produktiivisuusominaisuuksia leksikaalisesti CD-Perussanakirjan ja tekstitaajuudeltaan eli lemmafrekvenssiltään Kielipankin laajan aineiston avulla. Leksikaalisesti *patsas*-sanoja on 440 ja *keskus*-sanoja 2 628, eli jälkimmäisiä on melkein kuusi kertaa enemmän, kun taas lemmataajuuksien eli tekstifrekvenssien suhteet ovat päinvastaiset mediaanilla mitattuna (*patsas*-sanojen mediaanisanan lemmafrekvenssi on 231, kun vastaavan *keskus*-sanana lemmafrekvenssi on 40). Tämä on yleinen lainalaisuus: epäproduktiiviset sanat ovat käytöltään yleisempiä ja juuri sen takia ne säilyvät käytössä. Kummankin taivutustyyppin kolmitavuiset sanat pyrkivät olemaan keskitavultaan vahvoja (morilla mitattuna), ja laitavat ovat heikkoja. Erityisen vahvaksi taivutustyyppin merkiksi osoittautuvat viimeiset kaksi segmenttiä eli *-As* ja *-Us*. Epäsanojen taivutustestissä koehenkilöt suosivat odotuksenmukaisesti vahvasti *-Us*-paradigmaa. Kriittisenä huomautuksena voi todeta, että *-As*-sanoja tietenkin syntyy produktiivisesti johtamalla (vaikkapa *trendi+käs*). Tekijät pitävät yhtenä tuloksena sitä, että koehenkilöt eivät ole mieltäneet epäsanoina erisnimiksi eivätkä siksi yleisesti taivuttaneet *Carlos* : *Carlos+in*. Mutta tämä on todennäköisesti seuraus koeasetelmasta, koska koehenkilöt saivat tehtäväksi sijoittaa epäsanat objektimuotoon lausekehukseen *Minä otan ___*, eikä tällainen konteksti ole luonteva objektina olevan sanan erisnimitulkinnalle.

Artikkelissa ”Suomen nominien taivutusjärjestelmän produktiivisuuden indekseistä” (*Virittäjä* 112, 2008: 518–544) Nikolaev ja Niemi yleistävät edellisen artikkelin keskeiset tulokset koko suomen nominien taivutusjärjestelmään. Tekijät

valaisevat mielenkiintoisin kvantitatiivisin testein, minkälaisia riippuvuuksia ja osittaisia poikkeamia on pienen leksikaalisen frekvenssin – suuren käyttötaajuuden, ja toisaalta suuren leksikaalisen frekvenssin ja pienen käyttötaajuuden välillä. Erityisen merkittäväksi osoittautuu hapaksien eli korpuksessa vain kerran (tai erittäin harvoin) esiintyvien sanojen määrä. Produktiivisuustarkasteluissa liian vähälle huomiolle jäävät silti yhdyssanat. Vaikka kivettyneissä ja epäproduktiivisissa taivutustyypeissä onkin vain vähän perussanoja, nämä perussanat saattavat silti toistua sadoissa produktiivisesti muodostetuissa yhdysanoissa, jotka voivat olla erittäin yleisiä. Vaikkapa sana *lapsi* on taivutuksellinen uniikkitapaus, mutta se esiintyy korpuksessa 620 eri yhdyssanan perusosana, siis *sijoituslapsi*, *muslimilapsi*, *aviolapsi* jne. – Ei tunnu oikealta väittää *risti*-paradigmasta, että se olisi ”lähes agglutinatiivinen”. Monikon *i/e*-vaihtelun lisäksi tämän tyyppin sanat ovat tietenkin astevaihtelunkin alaisia, ja astevaihtelu on mitä epäagglutinatiivisin.

Artikkelissa ”Impaired attention to phonological input, dysfunctional ’lexical frequency counter/s’ and abnormal grammatical morphology: A possible causal chain?” (Zlatev ym. toim. 2009: 246–260) Niemi, Nikolaev ja Kenneth Hugdahl selvittävät tapaa, millä spesifistä kielihäiriötä potevat koehenkilöt prosessoivat eräitä keskeisiä taivutustyyppejä, kun he kohtaavat tuntemattomia epäsanonoja. Jossain määrin hypoteettiseksi selitykseksi tarjoutuu kyvyttömyys tajuta syötteen frekvenssieroja.

Vertaisarviointiin jätetyssä artikkelissa ”Paradigm complexity and perception of isolated lexemes: Effects of number of stem allomorphs on performance

in lexical decision and event-related potential experiments” Nikolaev, Ari Pääkönen, Jussi Niemi, Mikko Nissi, Eini Niskanen, Mervi Könönen, Esa Merivaala ja Hilikka Soininen selvittävät kokeellisesti, onko *lasi*- ja *vesi*-paradigmojen eräiden sananmuotojen todellisessa prosessoinnissa eroja. Aivovastemittaukset osoittavat odotuksen mukaisesti, että *vesi*-muotojen prosessointiin kuuluu enemmän resursseja (”left anterior negativity”). Tekijöiden sinänsä uskottava tulkinta on, että epäproduktiiviset muodot ovat laajemman, monen polun verkoston osia kuin produktiiviset muodot. Yksikön nominatiivilla ei ole mitään erikoisstatusta oletettuna ”perusmuotona”. Toisaalta epäproduktiivisten muotojen nopeampi tunnistaminen tulkitaan suuremman esiintymistäajuuden aiheuttamaksi. Pienenä huomautuksena voidaan todeta, että *vesi*-paradigma ei suinkaan ole teoreettisestikaan tarjolla sellaisten lainojen kuten *ledi*, *brändi* taivutusmalliksi, koska ei ole olemassa näin taipuvia **-di*-loppuisia nominatiiveja. Parempi ja omaperäinen esimerkki olisi *kasi* ’8’, joka taipuu *kasi+n* : *kasi+a* eikä **kade+n* : **kat+ta* kuten *käsi* (eikä toisaalta myöskään **kase+n* : **kas+ta* kuten *jousi*).

Yhteenvetoluvussa Nikolaev käsittelee uutena asiana mm. suomen substantiivien taivutusjärjestelmän informaatiopitoisuutta käsitteiden ”entropia” ja ”informaatio” avulla. Molemmat käsitteet ovat tunnetusti vaikeaselkoisia. Ne liittyvät toki toisiinsa, mutta niiden samastaminen (esim. s. 37) yksinkertaistaa liikaa. Tässä Nikolaev laskee tiettyjen paradigmaattisesti keskeisten taivutusmuotojen entropian päätyen siihen, että ”eniten informaatiota sisältää” yksikön partitiivi, koska sen entropian arvo on suurempi kuin yksikön nominatiiv-

vin ja genetiivin sekä monikon partitiivin ja inessiivin. Mutta suurta entropiaa ja korkeaa informaatiopitoisuutta voidaan myös pitää toistensa vastakohtina. Suuri entropia eli haje on yllätyksellisyyden, rakenteen puutteen ja epäjärjestyksen mitta, kun taas korkea informaatiopitoisuus viestii ennustettavuutta ja rakenteen eriytyneisyyttä. Sen takia informaation yksi luonnehdinta onkin, että se on negatiivista entropiaa (negentropiaa, vrt. Niiniluoto 1989: 18–22). Yksikön partitiivin muodosta käsin on helpompi ennustaa esimerkiksi yksikön nominatiivi kuin vaikkapa yksikön genetiivistä tai monikon inessiivistä käsin.

Kaikkiaan Alexandre Nikolaevin väitöskirja tarjoaa mielenkiintoisia, empiirisesti perusteltuja näkökulmia taivutusmorfologian produktiivisuuteen ja morfologisen prosessoinnin klassisiin ongelmiin.

FRED KARLSSON
etunimi.sukunimi@helsinki.fi

Lähteet

- ANTTILA, ARTO 1997: *Variation in Finnish phonology and morphology*. Väitöskirja, Stanford University, Department of Linguistics.
- BAUER, LAURIE 2001: *Morphological productivity*. Cambridge: Cambridge University Press.
- LINELL, PER 2005: *The written language bias in linguistics. Its nature, origins and transformations*. London: Routledge.
- NIINILUOTO, ILKKA 1989. *Informaatio, tieto ja yhteiskunta. Filosofinen käsitteanalyysi*. Helsinki: Valtion painatuskeskus.
- ZLATEV, J. – JOHANSSON FALCK, M. – LUNDMARK, C. – ANDRÉN, M. (toim.) 2009: *Studies in language and cognition*. Newcastle: Cambridge Scholars Publishing.