

# Aamiaiskahvilasta ötökkätarjontaan

## Suomen kirjoitetun yleiskielen morfosyntaktisten yhdyssanarakenteiden produktiivisuus

LAURA TYYSSTERI

### Väitöksenalkajaisesityelmä Turun yliopistossa 5. syyskuuta 2015

Väitöstutkimuksessani olen selvittänyt, mitkä suomessa kieliopillisesti mahdollisista yhdysanarakenteista ovat tätä nykyä produktiivisia käytännön tasolla eli millaisia yhdysanarakenteita kielenkäyttäjät suosivat muodostaessaan uusia yhdysanoja. Käytännöllistä produktiivisuutta kartoittava tutkimus täydentää kielioppien ja sanastonkuvausten antamaa kuvaa kielestä.

Kieliopit ja sanastonkuvaukset ovat kuin lintukirjoja. Lintukirjoissa luetellaan, millaisia lintuja on olemassa. Lisäksi niissä kerrotaan, miten eri linnut käyttäytyvät ja missä ne pesivät. Kielioppien ja sanastonkuvausten tarkoitus puolestaan on kuvata kaikki ne rakenteet, joita kielessä esiintyy. Kuvaukseen kuuluu esimerkiksi rakenteiden ja niiden esiintymisympäristöjen esittely.

Suomen kielisysteemi rajoittaa yhdysanojen muodostamista ja sanoissa käytettävien rakenteiden valintaa vain vähän. Yhdysanan osina voidaan käyttää eri sanaluokkiin kuuluvia sanoja, osat voivat esiintyä sanoissa hyvin monenlaisissa eri muodoissa eikä leksikaalisten osien määrälle ole mitään periaatteellista rajoitusta. Kaikkia kieliopillisesti mahdollisia rakenteita ei kuitenkaan käytetä yhtä ahkerasti.

Koska kielioppien ja sanastonkuvausten tehtävä on kuvata kielen mahdollisuuksia, niissä esitellään myös paljon harvinaisia rakenteita. Aina kuvausten tekijöillä tai kellään muullakaan ei edes ole tarkkaa tietoa siitä, kuinka paljon mitään rakenteita todellisuudessa käytetään. Esimerkiksi suomen kielen yhdysanarakenteiden osalta tarkka tieto eri rakenteiden yleisyydestä on tähän asti puuttunut.

Tutkimukseni tulokset perustuvat lähes 30 000 uuden suomen kielen yhdysanan tilastolliseen ja laadulliseen analyysiin. Tutkimuksen kohteena oleva kielimuoto on kaikille kielenkäyttäjille yhteinen kirjoitettu yleiskieli. Tutkimuksessa käytetty aineisto

on peräisin Kotimaisten kielten keskuksen eli Kotuksen Nykysuomen sanastotietokannasta (NSST), johon Kotuksessa kerätään jatkuvasti uutta ja uudella tavalla käytettyä suomen kielen sanastoa. Kotuksen tietokantaan vuodesta 2000 vuoteen 2009 tehdyistä poiminnoista olen ottanut erilleen painetun median kielestä poimitut yhdyssanat ja rakentanut niistä oman tutkimustietokantani (Tyysteri 2009–2015). Tämän tutkimustietokannan pohjalta olen analysoinut suomen kielen uusien yhdyssanojen pituutta, osien määrää, osien muotoa, osien sanaluokkaa sekä näiden eri piirteiden keskinäisiä vaikutussuhteita.

Olen käyttänyt tutkimuksessani kaksivaiheista metodia: Metodien ensimmäinen askel on uudessa sanastossa havaittujen rakenteiden tyyppifrekvenssin tilastollinen analyysi. Toinen askel on varsinkin matalafrekvenssisten tai tilastollisessa analyysissä muuten poikkeavaksi osoittautuneiden rakenteiden kvalitatiivinen tarkastelu. Käytetty metodi on kehitetty tätä tutkimusta varten, sillä produktiivisuuden tutkimuksessa aiemmin käytetyt menetelmät eivät sellaisenaan sovi suomen kielen yhdyssanarakenteiden tutkimukseen, ja uusille tutkimustavoille on tilausta myös muussa produktiivisuuden tutkimuksessa. Metodien kehittäminen onkin yksi tutkimukseni keskeisistä tavoitteista.

Aineiston analyysin myötä eteeni piirtyi lintukirjamaisen luettelon sijaan kuva suuresta parvesta, jonka jäsenistä valtaosa on rakenteeltaan prototyypisiä ja keskenään samankaltaisia pikkulintuja. Suurin osa aineiston sanoista noudattaa kaikkein tavanomaisinta rakennetta: Niissä on kaksi leksikaalista osaa, jotka ovat molemmat substantiiveja. Alkuosa on nominatiivimuotoinen, eikä se seuraa jälkiosaa taivutuksessa. Keskimääräinen pituus on 11–13 merkin paikkeilla, ja muodostustapana on käytetty suoraan kahdesta sanasta yhdistämistä. Tätä prototyypisintä ja samalla yksinkertaisinta mahdollista rakennetta noudattavat esimerkiksi sellaiset sanat kuin *lintuvaroitus*, *kanabingo*, *lintulautta* ja *ideahautomo*.

Jo etukäteen oli selvää, että juuri tällaisia yhdyssanoja on suomen kielessä entuudestaan eniten ja että rakenteeltaan juuri tällaisia uusia yhdyssanoja myös muodostetaan suomen kieleen eniten. Tarkkaa tietoa prototyypisen rakenteen yleisyydestä ei kuitenkaan ollut.

Tutkimusaineiston sanoista yli 92 % on substantiiveja. Substantiivialkuosa on yli 80 %:ssa aineiston sanoista samoin kuin nominatiivimuotoinen alkuosa. Yli 83 % koostuu kahdesta leksikaalisesta osasta. Suurin osa prototyypisistä piirteistä on tutkimusaineistossa yleisempiä kuin yleiskielen sanakirjoissa. Prototyypisten tapausten lisäksi parveen kuuluu kuitenkin lähes kaikkien muidenkin lintukirjassa mainittujen lajien edustajia. Lisäksi joukossa keekoilee myös uusia vieraslajeja ja lintua tavalla tai toisella muistuttavia otuksia, joita ei aiemmin tehtyihin lintukirjoihin ole merkitty lainkaan. Suurimmat yllätykset parven jakaumassa olivat prefiksinkaltaisella alkuosalla alkavien yhdyssanojen suuri määrä sekä genetiivialkusten yhdyssanojen vähyyt.

Suomen kielessä ei ole vanhastaan katsottu olevan prefiksistysteemiä lainkaan, ja mallin prefiksinkaltaisten elementtien käyttöön on arveltu tulleen käännöslainauksen kautta muista kielistä (esim. Nuutinen 1989: 110–119). Prefiksinkaltainen alkuosa on esimerkiksi sellaisissa yhdyssanoissa kuin *eurokäännös*, *esilintu*, *antikaunis*, *epälaulaja* ja *ulkokana*. Prefiksinkaltaisia elementtejä ei ole mainittu useimmissa suomen kielen

yhdyssanoja käsittelevissä tutkimuksissa lainkaan, ja suomen yhdyssanoja tähän mennessä laajimmin käsitelleessä *Isossa suomen kieliopissakin* (ISK 2004) huomiot ovat vain hajanaisia.

Nykyään prefiksinkaltaisia elementtejä käytetään uusien yhdyssanojen muodostuksessa kuitenkin jo selvästi enemmän kuin esimerkiksi vanhastaan tärkeänä pidettyjä genetiivimuotoisia alkuosia. Tutkimusaineistossa prefiksinkaltaisten alkuosien osuus on peräti 9 % kaikista alkuosista, kun genetiivimuotoisten alkuosien osuus jää alle 4 %:iin. Uudessa yhdyssanastossa käytettävistä prefiksinkaltaisista alkuosista hieman puolet on kotoista ja hieman alle puolet vierasta alkuperää.

Kaisa Häkkinen ja Pia Björkqvist (1991: 52) ovat jo aiemmin huomauttaneet prefiksinkaltaisten yhdysosien käytön muodostavan yhdyssananmuodostuksessa selvän, sekä produktiivisuutensa että funktionsa puolesta yhä tärkeämmäksi muuttuvan alasysteemin. Viimeistään tämä tutkimus osoittaa, että prefiksinkaltaiset elementit ovat vallanneet paikkansa suomen kielen sananmuodostuksessa, ja ne on jatkossa syytä ottaa paremmin huomioon myös kieliopeissa ja sanastonkuvauksissa.

Toinen aiemmin huomiotta jäänyt tulokaslaji on *-inen*-loppuisista vierasperäisistä adjektiiveista käytettävä *-i*-loppuinen yhdysosamuoto. Sellainen on esimerkiksi sanoissa *aktiiviseniori*, *polaarilento*, *primitiiviliikunta* ja *virtuaalilentäminen*. Tällaisten muotojen käyttö on yhdyssanan alkuosana säännönmukaista, eikä *-inen*-loppuisista vierasperäisistä adjektiiveista käytetä lainkaan *-s*-loppuisista muotoa, joka on tuttu kotoisten *-nen*-loppuisten sanojen yhteydestä. *-s*-loppuinen muoto – joka on yleisin yhdysosamuotomme – esiintyy esimerkiksi sellaisissa sanoissa kuin *erikoismuna*, *muinaistulokas*, *satunnaistalvehtija* ja *yhteismuninta*.

Kuten todettua, yksi suurimmista yllätyksistä oli vanhastaan tärkeinä pidettyjen genetiivialkuisten yhdyssanojen vähyys. Voi olla, että genetiivialkuja vältetään uusissa sanoissa esimerkiksi siksi, että niitä pidetään usein oikeinkirjoituksen kannalta hankalina. Useissa genetiivialkuisissa ilmauksissa on mahdollista sekä yhteen että erikseen kirjoittaminen sen mukaan, kuinka merkitykseltään kiteytyneenä kirjoittaja pitää ilmausta. Esimerkiksi *pesänputsari*-yhdyssanan osat voitaisiin joissain konteksteissa kirjoittaa myös erilleen toisistaan. Toinen mahdollinen syy genetiivialkuisten yhdyssanojen vähäisyydelle on kielissä yleinen rakenteiden yksinkertaistumisen tendenssi sekä se, että valmiiksi yleisin nominatiivialkuinen rakenne valtaa alaa muilta alkuosamuodoilta.

Tutkimusaineistossa myös monet sellaiset sanat, joissa genetiivialku tuntuisi odotuksenmukaiselta, esiintyvät nominatiivialkuisina. Tällaisia sanoja ovat esimerkiksi *kanaihoinen*, *jalkalämmitin*, *tapahtumajärjestäjä* ja *islampelko*. Esimerkiksi sanoista *trendinhaastaja*, *tyvenkohottaja* ja *uskonsoturi* sen sijaan esiintyy rinnakkain sekä genetiivi- että nominatiivialkuinen muoto.

Genetiivialkuisia yhdyssanoja ei vielä voida pitää uhanalaisina, mutta lähes sukupuuton partaalla näyttää olevan ainakin kaksi kieliopeista ja sanastonkuvauksista tuttua yhdyssanarakennetta. Aiemmassa tutkimuksessa on usein todettu, että yhdyssanan alkuosa voi suomen kielessä esiintyä periaatteessa missä sijamuodossa tahansa, joskin varsinkin partitiivin ja monikon nominatiivin esiintymiseen liittyy joitain jälkiosan rakennetta koskevia rajoituksia. Tämän tutkimuksen lähes 30 000 yhdyssanan aineis-

tossa muu sijamuoto kuin nominatiivi tai genetiivi esiintyy kuitenkin vain 28 yhdyssanassa. Kaikki nämä tapaukset ovat joko aiemmista yhdyssanoista johtamalla tai sanaliitosta tiivistymällä muodostuneita (esim. *maahanimeyttämö* ja *maallemuuttaja*), eikä rakennetta voida nähdäkseni enää pitää ainakaan suoran yhdistämisen kannalta produktiivisena suomen kirjoitetussa yleiskielessä.

Toinen uhanalainen ryhmä ovat osiltaan kongruoivat yhdyssanat, joissa myös alkuosa taipuu samaan tapaan kuin jälkiosa (esim. *hyvä omatunto : hyvällä omallatunnolla*). Tällaisia osiltaan kongruoivia yhdyssanoja on tutkimusaineistossa vain kuusi kappaletta. Aiemmissä tutkimuksissa kongruenssin on todettu olevan katoamassa sellaisitakin yhdyssanoista, joissa sitä on vanhastaan esiintynyt (Lehtinen 1984; Eronen 1996). Voikin olla, että tulevaisuudessa tämä jo harvinaiseksi käynyt laji katoaa kokonaan.

Vaikeasti ymmärrettäviksi syytetyt ja yleiseksi luullut pitkät yhdyssanat paljastuivat tutkimuksessa luultua harvinaisemmiksi. Aiemman tutkimuksen mukaan suomen kielen yhdyssanoissa on harvoin enemmän kuin viisi leksikaalista osaa. Tämän tutkimuksen perusteella osia ei kuitenkaan yleensä ole neljää enempää, ja neliosaisiakin on vain alle prosentti tutkimusaineistosta.

Suomea vieraana kielenä opiskelevia on toisinaan jopa suoranaisesti peloteltu suomen pitkillä sanoilla. Senkaltaiset pitkät ja vaikeatajuiset sanahirviöt kuin *kestomagneettitahtikonekäyttö* ja *lähialueministerivaliokunta* ovat kuitenkin vain poikkeustapauksia. Kielenoppijalle voikin kertoa helpottavana tietona, että itse asiassa suomen kielen yhdyssanat eivät ole niin pitkiä, kuin usein väitetään.

Pitkien yhdyssanojen ymmärtäminen helpottuu, kun niiden rakenteen hahmottaa ja ne osaa pilkkoa osiin. Tämä tutkimus tarjoaa esimerkiksi opettajille paljon tietoa siitä, millaiset rakenteet ylipäänsä ovat produktiivisia suomen kielen uudessa yhdyssanastossa. Opetuksen lisäksi tämän tutkimuksen tuloksia voidaan hyödyntää kieliteknologiassa, jossa niitä voidaan käyttää apuna kehitettäessä työkaluja kielen automaattiseen prosessointiin. Esimerkiksi kaikissa tilanteissa oikein toimivien oikolukuohjelmien luominen on mahdotonta, jos myös uudissanastossa todennäköisesti vastaan tulevia rakenteita ei tunneta.

Kuten tässäkin esitelmässä käyttämistäni esimerkkinsanoista käy havainnollisesti ilmi, yhteen yhdyssanaan voidaan tiivistää valtavasti informaatiota. Erilaisten morfosyntaktisten yhdyssanarakenteiden suosio kertoo siitä, millä kieliopin keinoilla tämä informaatio on kielenkäyttäjien mielestä tarkoituksenmukaisinta tiivistää. Aina uutta sanaa muodostaessaan kielenkäyttäjä valitsee tilanteeseen sopivimman rakenteen useista eri vaihtoehdoista. Kielenkäyttäjien – yleensä tiedostamatta – tekemät valinnat heijastavat esimerkiksi sitä, millaisia rakenteita pidämme riittävän helppoina sekä tuottaa että ymmärtää.

Mitä enemmän yhdyssanaan pakataan leksikaalisia osia, sitä yksityiskohtaisempi ja tarkempi sen merkityksestä saadaan. Viestin perillemeno voidaan varmistaa myös käyttämällä alkuosassa sellaista sijamuotoa, joka osoittaa osien välillä olevan merkityssuhteen yksiselitteisesti. Esimerkiksi sanoilla *maahanmuuttaja*, *maassamuuttaja* tai *maastamuuttaja* on selvästi vähemmän erilaisia tulkintamahdollisuuksia kuin nominatiivialkuisella sanalla *maamuuttaja*.

Se, että kielenkäyttäjät kuitenkin useimmissa tapauksissa valitsevat yksinkertaisimman mahdollisen rakenteen, kertoo siitä, käytettävyys ja nopea viestintä on yleensä yksiselitteisyyttä tärkeämpää. Useimmissa tapauksissa yksinkertainen rakenne riittää yhdyssanan sisältämän informaation tarkoituksenmukaiseen pakkaamiseen. Ymmärrämme toisiamme, vaikka kieli ei olisi yksiselitteistä.

Kaikki aiemmin tehdyissä lintukirjoissa mainitut yhdyssanarakenteet ovat suomen kielessä edelleen mahdollisia: niin kauan kuin yksikin sana noudattaa tiettyä rakennetta, se voi myös toimia analogisena mallina uudelle muodosteelle. Käytännössä osaa vanhoista rakenteista ei kuitenkaan enää käytetä kovin aktiivisesti uusien sanojen muodostuksessa. Niiden sijaan on saatu uusia rakenteita, joista selvästi käytetyimpänä erottuvat prefiksinkaltaiset alkuosat. Kaiken kaikkiaan voin väitöstutkimukseni perusteella sanoa, että suomen kielen uudet yhdyssanat ovat rakenteeltaan homogeenisempi joukko, kuin kielioppeja ja sanastonkuvauksia lukemalla voisi olettaa.

## Lähteet

- ERONEN, RIITTA 1996: Monimuotoiset yhdyssanat. – *Kielikello* 4/1996 s. 12–14.
- HÄKKINEN, KAISA – BJÖRKQVIST, PIA 1991: Varhaisnykysuomen yhdyssanat. – *Sananjalka* 33 s. 31–55.
- ISK = HAKULINEN, AULI – VILKUNA, MARIA – KORHONEN, RIITTA – KOIVISTO, VESA – HEINONEN, TARJA RIITTA – ALHO, IRJA 2004: *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seuran toimituksia 950. Helsinki: Suomalaisen Kirjallisuuden Seura.  
Verkkoversio: <http://scripta.kotus.fi/visk/etusivu.php>.
- LEHTINEN, RAIJA 1984: Omastatunnosta omatuntoon? – *Virittäjä* 88 s. 517–518.
- NSST = Kotimaisten kielten keskuksen Nykysuomen sanastotietokanta. Päivitetään jatkuvasti. Tiedot ajalta 1.1.2000–1.9.2009, haettu 2.10.2009. Helsinki: Kotimaisten kielten keskus. Saatavuus: Arkistotietokantaa voi käyttää Kotimaisten kielten keskuksen intranetissä, mutta käyttö vaatii luvan.
- NUUTINEN, OLLI 1989: Käännöslainat. – Jouko Vesikansa (toim.), *Nykysuomen sanavarat* s. 110–126. Porvoo: WSOY.
- TYYSSTERI, LAURA 2009–2015: Tutkimustietokanta. NSST:sta 1.1.2000–1.9.2009 poimittujen tietojen pohjalta koostettu julkaisematon yhdyssanatietokanta. Turku: Suomen kieli, Turun yliopisto.

**Laura Tyysteri:** *Aamiaskahvilasta ötökkätarjontaan. Suomen kirjoitetun yleiskielen morfosyntaktisten yhdyssanarakenteiden produktiivisuus*. Annales Universitatis Turkuensis C 408. Turku: Turun yliopisto 2015. Väitöskirja on luettavissa osoitteessa <http://www.doria.fi/handle/10024/113113>.

Kirjoittajan yhteystiedot:  
etunimi.sukunimi@outlook.com