

# Sanojen synteettisyys suomen kielessä

ANNELI PAJUNEN

## 1 Synteettisyyden määrittelemisestä

Sanojen synteettisyys määritellään kielissä morfosyntaktisesti. Morfosyntaktinen jaottelu perustuu abstraktin (= kieliopillisen) merkityksen koodauseroihin ja niitä usein vastaaviin rakenteellisiin kompleksisuuseroihin (kompleksisuudesta ks. esim. Kusters 2003). Erojen perusteella kielet voidaan jakaa eri tyyppisiin jatkumolla analyttinen–synteettinen–polysynteettinen. Puhtaassa analyttisessä tyypissä yksi vapaa morfeemi eli itsenäinen sana ilmaisee joko yhden leksikaalisen kategorian eli merkityksen (konkreettinen merkitys kuten substantiivi 'talo') tai yhden kieliopillisen merkityksen (abstrakti merkitys kuten luku partikkelisanalla ilmaistuna). Synteettisessä kielessä sana ilmaisee leksikaalisen merkityksen ja tyypillisesti sanan alkuun, keskelle tai loppuun liittynyt sidonnainen morfeemi yhden tai useampia kieliopillisiä merkityksiä. Polysynteettisessä kielessä yksi sana ilmaisee useampia leksikaalisia ja/tai kieliopillisiä merkityksiä ja myös leksikaalinen merkitys voidaan ilmaista sidonnaisella morfeemilla. Jos kielten rakenteellista, tarkemmin sanarakenteen, kompleksisuutta ilmaistaan edellisellä jaottelulla, on tapana sanoa, että analyttisten kielten sanassa olevien leksikaalisten tai kieliopillisten kategorioiden määrä on keskimäärin yli yhden, synteettisten kielten yli kahden ja polysynteettisten kielten yli kolmen (Greenberg 1960: 185, 194). Kyse on luonnollisesti jatkumosta, jolla kielet eivät edusta puhdasta tyyppiä; esim. nominitaivutus ja verbitaivutus voivat yksittäiskielessä erota arvoiltaan.

Suomen kieli luetaan vanhastaan synteettisiin kieliin, ja suomessa leksikaalisen ja kieliopillisen merkityksen koodausero on selvä. Suomen sanojen keskimääräinen synteettisyyden aste on Korhosen (1969: 32) mukaan 2,22, kun otetaan huomioon sekä derivaatio että taivutus. Synteettisenä kielenä suomessa on sekä agglutinoivia että fusionaalisia piirteitä eli tunnusten läpinäkyvyys ja sanojen segmentoitavuus vaihtelee. Viime aikoina kuitenkin suomen kielen synteettisyys tai sen aste on kyseenalaistettu ja verrattu suomea englannin ja ruotsin kaltaisiin kieliin, joissa sekä nominien että verbien taivutus on vähäistä tai olematonta (ks. Bickel & Nichols 2005; Dahl 2008). Arvioin tässä artikkelissa näitä eri näkemyksiä määrittelemällä synteettisyyden (yleistään *category per word* eli cpw-arvon) eri laskutapoja tarkemmin ja laskemalla eri laskutapojen mukaisia arvoja. Osoitan, että uudessa synteettisyyskeskustelussa suomen

sanojen cpw-arvoja on joko mitattu virheellisesti tai kielten vertailussa on sekoitettu eri laskutapoja. Tosin on syytä pitää myös mielessä, että yksittäisten mittarien antama tieto maailman kielistä on vajaata. Esim. sanarakenteen kompleksisuus ja toisaalta lekseemiin koodatun merkityksen määrä voi kielessä olla komplementaarisessa suhteessa (vrt. esim. Gil 2008; Riddle 2008).

Kirjallisuudessa on esitetty useita synteettisyyttä ilmaisevien cpw-arvojen laskutapoja. Käsittelen tässä suomen sanojen avulla sanaluokittaisen mahdollisen taivutuksen (= paradigmaattinen taso), sanaluokittain maksimaalisen taivutuksen (= syntagmaattinen taso) sekä tekstitasolta keskimääräisen taivutuksen ja vertailen saatuja tuloksia. Olen tutkinut mahdollista taivutusta intuition avulla ja kieliopista. Maksimaalista taivutusta olen tutkinut isoista sanomalehtikorpuksista<sup>1</sup> (ks. Pajunen 2002; 2003); myös esimerkit on poimittu näistä. Keskimääräinen taivutus on laskettu Lauseopin arkiston (= LA:n) miljoonan sanan korpuksesta, jossa on sekä puhutun että kirjoitetun kielen aineistoa. Puhutun aineisto koostuu aluemurteista. Sanomalehtikorpuksukset on koodattu automaattisesti, LA:n aineisto manuaalisesti.

## 2 Suomen sanaluokat ja mahdollinen taivutus

Suomen pääsanaluokkia ovat substantiivit, adjektiivit ja verbit ja pienluokkia mm. pronominit, numeraalit ja konjunktiot. Pääsanaluokat ovat kielissä hyvin yleisiä ja ne voidaan määritellä usealla kriteerillä, joista mainittakoon fonotaktinen, fonologinen, morfologinen, syntaktinen, semanttis-ontologinen ja pragmaattinen. Pienluokat ovat kielikohtaisempia ja ne määritellään tyypillisesti muodon ja distribuution perusteella. Sanaluokkien väliset erot ovat liukumonkaltaisia, eivät jyrkkiä, mutta sanaluokan keskeiset jäsenet pystytään yleensä määrittelemään paitsi funktion myös sanarakenteen (johtamaton vs. johdos), taivutuksen maksimaalisuuden vs. minimaalisuuden ja distribuution perusteella (ks. Pajunen 1994; 1998a; 1998b). On myös yleistä, että tietyille sanaluokalle tyypillisessä diskurssifunktiossa käytetyt sanat ovat tunnusmerkittämpiä (eivät sisällä esim. muuntavia johtimia yhtä paljon) kuin sanaluokalle epätyypillisessä diskurssifunktiossa käytetyt sanat (ks. esim. Croft 1991: 55, 67). Esim. substantiivi on useammin johtamaton silloin, kun sillä viitataan objektien semanttiseen luokkaan kuin silloin, kun sillä predikoidaan tiloista ja tapahtumista (vrt. *tyttö, tyttö-ile-; isä, isä-tyt-* 'tulla isäksi'). Määrittelen seuraavassa sanaluokittain ilmaistavia kieliopillisia merkityksiä (taulukko 1) ja arvioin tutkimukseen kuuluvia sanaluokkia niiden mahdollisen taivutuksen (taulukko 2) kannalta.

Osa kieliopillisista merkityksistä ilmaisee sanaluokittain tyypillisiä inherenttejä ominaisuuksia, substantiivien osalta mm. esineisyyttä ja luonnollista luokkaa, adjektiivien osalta mitattavuutta ja verbien osalta muutosta ajassa ja teon luonnetta (ks. taulukko 1, vrt. Itkonen 2008a: luku VIII). Inherenttejä ominaisuuksia ei suomessa aina ilmaista tunnuksella eli formatiivilla, vaikka ne aina vaikuttavat sanaluokan sanojen käyttäytymiseen. Esim. substantiivien konkreettisuus vs. abstraktius rajaa sen, min-

1. Kiitän Sanoma Oy:tä luvasta käyttää *Helsingin Sanomien* aineistoa.

käläisiä ominaisuuksia tai tekoja niistä voidaan predikoida. Jaollisuus ja jaottomuus edelleen vaikuttavat sijavalintoihin. Muutosta koodaa verbiin mm. tempus, aspekti ja modus. Verbin inherenttejä ominaisuuksia koodaavat formatiivit kietoutuvat nykysuomessa osaksi yhteen; muutoksen tapa ilmaistaan usein välillisemmin (esim. johdoksilla). Aspektia ilmaisevat tempuksen lisäksi mm. sijat. Puheakteista käsky ja kysymys voidaan suomessa ilmaista formatiivilla, väitettä, faktisuutta ja affirmaatiota ei ilmaista, negaatio ilmaistaan mutta ei yksin taivutuksella. Modaliteetin ilmaisussa on käytössä monia eri keinoja, myös formatiiveja.

**Taulukko 1.**

Pääsanaluokkiin tyypillisesti koodattavia kieliopillisia merkityksiä Itkosen mukaan (2008a: luku VIII).

<b>Inherentit ominaisuudet</b>
<b>Substantiivit:</b> esineisyys, luonnollinen luokka (konkreetti, abstrakti, jaollinen, jaoton, elollinen, eloton)
<b>Verbit:</b> puheakti, affirmaatio ja negaatio, modaliteetti, tempus, aspekti
<b>Adjektiivit:</b> mitattavuus (komparaatio ja intensifikaatio)
<b>Kontekstuaaliset ominaisuudet</b>
tuttu ja uusi informaatio, deiksis
<b>Relationaaliset ominaisuudet</b>
persoona, luku, luokka, omistus
<b>Rektionaaliset ominaisuudet</b>
lauseenjäsenet ja semanttiset roolit
<b>Argumenttirakenteen muunnokset</b>
passiivi, kausatiivi, mediumi, refleksiivi, resiprokaali jne.

Kontekstuaalisia merkityksiä, informaation kulkua, ei ilmaista suomessa formatiiveilla mutta aineiston murrepuheessa käytetään yleisesti pronomineja ja pronominaaleja (*se* N; *semmoinen* N tms.) tutun ja uuden informaation erotteluun. Välillisesti informaation kulun merkitseminen näkyy pronomien ja siten myös taivutuksen määrän noususta.

Relationaaliset merkitykset ilmaistaan usein kahdessa sanassa kongruenssin eli tunnusten osittaisen tai täydellisen toiston avulla (pääsana-attribuutti-, subjekti-predikaatti- ja predikaatti-kongruenssi; omistuskongruenssi). Relationaalisista merkityksistä luku ja omistus koodautuu suomessa substantiiviin, persoona verbiin, luokkaa ei koodata formatiivilla ainakaan läpinäkyvästi. Luku ja omistus laukaisevat kongruenssin, persoonan merkintä sekoittuu paikoin luvun merkintään. Ominaisuuksiin ei inherentisti liity useutta, joten luvun tunnus adjektiivissa on aina kongruenssin tuoma; useimmiten myös adjektiivin sijatunnus on kopioitu toisesta sanasta. Kongruenssi nostaa adjektiivien cpw-arvoa selvästi. Persoonapronomineissa ja demonstratiivipronomineissa ei ilmaista lukua nominatiivisijassa taivutusformatiivilla vaan vartalonvaihtelulla (*minä* vs. *me*: mutta *me-i-ssä*) toisin kuin muissa pronomineissa (*joka* vs. *jo-t-ka*).

Persoonatunnukset koodaavat finiittiverbiin tiedon sekä tekijäpersoonasta että tekijöiden useudesta (*minä mene-n : me mene-mme*) mutta 3:nnen persoonan tunnukset eivät ole läpinäkyviä (esim. *tekevät*). Tyypillisesti subjektina toimiva substantiivi koodaa tekijöiden useuden ja predikaatti kopioi tämän tiedon. Murteiden kanta vaihtelee. Esim. lounaismurteissa yksikön ja monikon 3. persoona koodataan eri tavalla (*men : menit*), savolais- ja hämäläismurteissa monikko voi jäädä ilmaisematta. Tekojen useutta tai toistuvuutta ei suomessa ilmaista taivutuksella (vaan esim. johtimilla). Omistus koodaa omistajan persoonaa ja useutta etupäässä substantiiveihin; omistettavan useutta ei koodata, siten ilmauksissa *minun koirani : meidän koiramme* voidaan viitata kummassakin joko yhteen tai useampaan koiraan.

Kongruenssin vaikutus näkyy siis (kopioitujen) formatiivien määrän nousuna. Aineisto ei mahdollista kongruenssin laskemista erikseen, joten sitä ei voi ottaa huomioon erillisenä kieliopillisena merkityksenä (vrt. englantia, ks. Bickel & Nichols 2005: 95; ks. myös Greenberg 1960: 187). Tämä olisi perusteltua mm., koska kongruenssilla on oma selvä syntaktista yhteenkuuluvuutta ilmaiseva ja prosessointinopeutta lisäävä funktionsa (ks. Vainio, Hyönä & Pajunen 2008).

Rektionaaliset merkitykset ilmaistaan suomessa pitkälti sijataivutuksella mutta lisäksi mm. sanajärjestyksellä. Suomessa kaikki nominit osallistuvat sijataivutukseen mutta ainoastaan substantiivien sijataivutus on maksimaalista, so. ne taipuvat eri sijafunktioissa laajimmin ja myös määrällisesti eniten (tarkemmin Pajunen 1994). Sijakongruenssi nostaa erityisesti kieliopillisten sijojen määrää, koska mm. adjektiivattribuuteista suurin osa taipuu juuri kieliopillisissa sijoissa. Argumenttirakenteen muutosta ilmaisee suomessa sekä taivutuksen *t*-passiivi että derivaatio (esim. transitiivisuusvaihtelut) ja sijavaihtelut (esim. kontaktiverbien partonominen ja konatiivinen vaihtelu, vrt. *ampua karhua, karhuun (päin), karhu*, ks. Pajunen 2001: 238–).

Lisäksi suomessa ilmaistaan diskurssin sävyä kliiteillä (*-hAn, -kin, -kAAAn, -pA ja -s*); sävy käsitetään tässä funktioksi, jota voidaan ilmaista useammallakin formatiivilla. Vaihtoehtoinen analyysi edellyttäisi kliittien funktioiden tarkempaa analyysiä ja tuloksena mahdollisesti olisi korkeampia cpw-arvoja. Kliiteistä *-kAAAn* ilmaisee lisäksi negatiota. Sävykliitti liittyy eri sanaluokkia edustaviin sanoihin mutta ei aina konjunktioihin (vrt. *että-s, koska-s, kun-han* mutta *\*kun-kin, \*että-hän*). Tutkimuksen aineistossa on alueellisia eroja etenkin *-kin, -s* ja toisaalta *-hän* kliittien käytön välillä. Sävykliitteihin luetaan tässä myös kieltoverbiin liittyvä *-kä*, jonka funktiot ovat lähinnä sidostavia. Sävykliitit voivat osin kombinoitua keskenään ja myös kysymyskliitin kanssa; tosin *kin-* ja *hAn-*kliitit eivät korpusten mukaan kombinoitua, *kAAAn-*kliitti yhdistyy muihin rajatuimmin.

Sanaluokat määritellään suomessa pääasiassa morfologisen kriteerin eli sanojen taivutuksen mukaan (Penttilä 1963: 327–328; ISK 2004: 426). Sijataivutukseen osallistuvat ovat nomineja ja tempuksen, moduksen ja persoonan mukaan taipuvat ovat verbejä. Taipumattomat sanat ovat partikkeleja. Tämän tutkimuksen perusaineisto on valittu tämän jaottelun mukaisesti. Nomineihin lukeutuvista mukana ovat substantiivit (appellatiivit ja propriit), adjektiivit, pronominit (mutta eivät pronominaalit) ja partisiipit eri funktioissaan. Verbien taivutuksen ja distribuution perusteella erottuvia alaluokkia ovat finiittimuodot, kieltoverbit ja kieltoverbiin liittyvät kieltoimuodot sekä infinitiivit;

näistä finiittiryhmä taipuu sanatasolla eniten ja on siten tässä ensisijainen. Infinitiivien arvot perustuvat Pajusen ja Palomäen (1984; 1985) laskelmiin. Partikkelit eivät ole tutkimuksessa mukana. Myös adverbihin liittyy suomessa usein nominitaivutustunnuksia mutta ne voidaan synkroniselta kannalta katsoen jättää analysoimattakin.

Sanaluokkia voidaan tarkastella myös taivutuskategorioittain, jolloin kiinnitetään huomiota sanaluokalla ilmaistavien morfologisten erojen määrään; taustalla on luonnollisesti semanttisia ominaisuuksia (vrt. taulukot 1 ja 2). Voidaan myös tarkastella taivutuselementtien distributioeroja, jotka heijastavat etenkin diskurssifunktioita; ne paljastavat esim. nominien sisäisiä eroja (tarkemmin ks. kaaviot 2, 3 ja 6). Taulukon 2 taivutus summaa kaikki ne merkitykset (= paradigmaattinen taivutus), jotka suomen kielessä voidaan liittää ao. sanaluokkiin; sulut tarkoittavat, että taivutus koskee rajattua alaryhmää tms.

*Taulukko 2.*

Paradigmaattinen taivutus suomessa.

Sanaluokittain mahdolliset taivutuskategoriat aineistossa	yht.
Substantiivit: (komparaatio), luku, sija, omistus, sävy, kysymys	6
Pronominit: (luku), sija (omistus), sävy, kysymys	5
Adjektiivit: (komparaatio), luku, sija, (omistus), sävy, kysymys	6
Partisiipit: pääluokka, (komparaatio), luku, sija, (omistus), sävy, (kysymys)	7
Verbit: (pääluokka), tempus, modus, persoona, sävy, kysymys	6
Infinitiivit:(pääluokka), infiniittisyys, sija, (omistus), sävy, kysymys	6

Sanojen synteettisyyttä mittaavat indeksit perustuvat paradigmaattiseen taivutukseen, koska se rajaa sanan (tai sanaluokan) mahdollisen taivutuksen. Yksittäiset tietyn sanaluokan sanat eivät kuitenkaan salli kaikkia koko sanaluokalle mahdollisia formatiiveja (vrt. sulut taulukossa 2), vaan yksittäisten sanojen taivutettavuus vaihtelee. Jos synteettisyyttä arvioidaan tekstisanojen ja formatiivien suhteena, ongelmaan ei tarvitse kiinnittää huomiota. Jos sen sijaan pyritään löytämään sanaluokan se sanamuoto, jonka cpw-arvo on korkein (ks. Bickel & Nichols 2005: 94–95), on erotettava mahdollinen ja todella esiintyvä taivutus.

Bickel ja Nichols (mp.) esittävät useista kielistä verbien maksimaaliseen taipumiseen perustuvan cpw-arvon ja vertailevat kieliä keskenään arvojen perusteella. Valittavasti Bickelin ja Nicholsin esittämät arvot ja niiden laskutapa ovat ongelmallisia (ks. tarkemmin Itkonen & Pajunen 2008; Itkonen, Pajunen & Suomi 2009). Ensiksikin 'kategoria' tarkoittaa yleensä merkitysyksikköä, mutta Bickelillä ja Nicholsilla (ja yleisimminkin *World Atlas of Language Structures* -teoksessa eli WALS:ssa) se tarkoittaa muotoyksikköä, so. affiksia tai kliittiä. Toiseksi sanamuoto, jonka cpw-arvo on mahdollisimman korkea, eli käytännössä maksimaalista taivutusta edustava muoto, ei anna kielen (keskimääräisestä) synteettisyydestä edustavaa kuvaa. Esimerkiksi suomessa sija liittyy joka substantiiviin ja sen vaikutus substantiivien keskimääräiseen cpw-arvoon on siten 1. Ruotsissa sijoja on vain yksi, genetiivi, ja se liittyy korkeintaan joka kym-

menenteen substantiiviin, jolloin sen vaikutus keskimääräiseen cpw-arvoon olisi korkeintaan 0,1. Maksimaalinen cpw-arvo ruotsin substantiiveilla on kuitenkin Dahlin (2008: 547) mukaan 3 (*hund-ar-na-s* koira+MON+MÄÄR+GEN). Kolmanneksi Bickelin ja Nicholsin cpw-arvo ilmaisee oikeastaan agglutinaation eli taivutuksen läpinäkyvyyden astetta, ei taivutuksen määrää sinänsä. On huomattava, että agglutinaatio on syntetisyyks-käsitteen alakäsite, ei synonyymi.

Neljänneksi cpw-arvoa ilmaisee WALS:ssa ei vain yhdessä verbimuodossa samalla kertaa esiintyvien formatiivien lukumäärä vaan eräiden kielten yhteydessä myös (verbi)taivutuksessa yleensä esiintyvien (paradigmaattisten) vaihtoehtojen lukumäärä. Tämä selittää englannin ja suomen sijoittamisen samaan cpw-arvo-luokkaan eli 2–3: englannin arvo perustuu paradigmaattiseen arvoon, suomen (virheelliseen) maksimaaliseen cpw-arvoon. Tosiasiassa englannin verbissä voi kerrallaan olla vain yksi taivutusformatiivi (*work-s* tai *work-ed*), vain paradigmaattiselta kannalta vaihtoehtoja on kaksi. Mainittakoon, että esim. Haiman (1985: 165) määrittelee englannin askeettiseksi, lähes analyttiseksi kieleksi.

Suomessa verbissä voi olla viisikin formatiivia yhtä aikaa (ks. seuraava luku) ja paradigmaattisesti arvioiden vaihtoehtoja on kuusi (ks. taulukko 2). Toisin sanoen paradigmaattiseen taivutukseen perustuva arvo on aina korkeampi tai vähintään yhtä korkea kuin syntagmaattiseen, ja kieliä ei voi vertailla näitä kahta laskutapaa sekoittaen. On myös syytä huomata, että WALS:n cpw-arvon teoreettisesti alin arvo ei ole yksi – kuten Greenbergillä (1960: 185) – vaan nolla. Määrittelen seuraavaksi tarkemmin tiettyjen sanaluokkien maksimaalisen eli syntagmaattisen taivutuksen suomessa.

### 3 Sanaluokkien syntagmaattinen taivutus suomessa

Periaatteessa on helppo luetella ne kieliopilliset merkitykset, joita suomen kielessä ilmaistaan tunnuksilla eli formatiiveilla. Tämä ei vielä riitä, koska merkinnän systemaattisuus vaihtelee ja koska merkintä ei toteudu sanaluokittain aina samassa laajuudessa. Pohdin kieliopillisen merkityksen ilmaisun ongelmallisuutta eri tavoin ja määrittelen samalla sanaluokan jäsenten mukaan hieman vaihtelevan syntagmaattisen cpw-arvon tässä tutkimuksessa mukana oleville suomen sanaluokille.

Arvojen laskemiseen ja formatiivien määrittelyyn vaikuttavia tekijöitä ovat ainakin seuraavat kolme: 1) Yksi merkitys voi koostua kontrastoivista sarjoista (kuten eri sijat, persoonat jne.) ja sarjan kaikkia jäseniä ei aina ilmaista erillisellä formatiivilla vaan esim. nollalla tai kumulatiivisesti. Esim. sijamerkityksistä nominatiivi ilmaistaan nollalla, vastaavasti luvusta yksikkö. 2) Toisaalta esim. muodolla *anta-a* on perinteisen kieliopin mukaan merkitykset 'aktiivin indikatiivin preesens, yksikön kolmas persoona'; lisäksi muoto ilmaisee vielä merkityksiä 'faktinen', 'affirmatiivinen', 'väite'. Ts. käytössä oleva formatiivi on (erittäin) kumulatiivinen. 3) Kolmanneksi jonkin muodon tunnus voidaan tulkita eri tavoin analyysistä riippuen. Erityisesti diakroninen ja synkroninen analyysi tuovat eri ratkaisuja. Havainnollistan eroja parilla esimerkillä; huomattakoon, että puhutun aineistoni kannalta diakronian ja synkronian raja on yleisikieltä horjuvampi.

Kielihistoriaan perustuvan analyysin mukaan esim. imperatiivin tunnus sisältää (mahdollisesti preesensia ilmaisevan) *k*-aineksen (ks. Rapola 1966: 300; Lehtinen 2007: 132), joka yksikön 2. persoonan muodoista on sellaisenaan kadonnut (*tee* ja *anna*) mutta aiheuttaa puheessa (yleensä) rajageminaation (*tees se, annat tänne*). Muissa persoonissa edustuu ns. 1. imperatiivin *kA*- (*tehkää(mme), tehkää(tte)*) tai ns. 2. imperatiivin *kO*-tunnus (*tehköön, tehkööt*) (ks. esim. Penttilä 1963: 217–218). Imperatiivin persoonamuodoista yksikön 2. persoona on tunnuksen, muissa persoonissa tilanne vaihtelee. Nykykielen kannalta voi sanoa, että rajageminaatio ilmaisee sekä imperatiivia että persoona (Leskinen 1970: 215; ISK 2004: 145). Voidaan myös katsoa, että tunnuksena toimii rajageminaatio yhdessä vokaalivartalon astevaihtelun luonteesta riippuen heikko- (*tee', anna'*) tai vahva-asteisen (*hyppää'*) muodon kanssa.

Passiivin finiittimuodot katsotaan suomessa persoonattomiksi (esim. *lue-taan, men-tiin*, ks. esim. Penttilä 1963: 460). Nykykielen kannalta tunnus on kokonaisuus (mahdollisesti kumulatiivinen), mutta historiallisesti se voidaan eritellä passiivin (*t*)*t*-tunnukseen ja persoona ilmaisevaan vokaalin pidentymään; alkuaan (*t*)*t*-aines palautuu kausatiivin tunnukseen ja vokaalin pidentymä 3:n persoonan possessiivisuffiksiin (ks. Lehtinen 2007: 131–132). Esim. *Iso suomen kielioppi* (2004: 137) erottaa passiivin finiittimuodoista edelleen kaksi tunnusta. Yksi peruste ratkaisulle on, että modustunnus voi jakaa passiivin tunnuksen (*men-tä-isi-in*) kahtia.

Arvioin seuraavaksi suomen taivutussanaluokille ns. maksimaalisen cwp-arvon. Käsittelen ensin nominisanaluokat ja palaan verbien yhteydessä ns. nollan ongelmaan tarkemmin.

Kaikkia substantiiveille, adjektiiveille tai verbeille ominaisia tai mahdollisia kielipiillisiä merkityksiä ei voi yhdessä sanassa ilmaista samanaikaisesti (vrt. taulukko 2). Esim. substantiivit eivät periaatteessa komparoidu mutta eräät etenkin spatiaalista merkitystä ilmaisevat voivat komparoitua. Nimeävät substantiivit eli propriit eivät komparoidu lainkaan (ainakaan siten, että proprinluonteisuus säilyisi). Jotkin substantiivit ovat myös inherentisti yksiköllisiä (esim. abstraktisubstantiivit usein), joten niiden taipuminen luvussa on kyseenalaista tai harvinaista. Korpuksista ei löydy esimerkkejä komparoituneista substantiiveista (lukuun ottamatta leksikaalistunutta sanaa *vanhemmat*); tämä tosin voi johtua myös koodausrajoituksista. Esim. sana *etelä* esimerkissä *tuhansia kilometrejä etelämpänä* on korpuksessa analysoitunut adjektiiviksi. Substantiivien maksimaalinen syntagmaattinen cwp-arvo vaihtelee siten 3–5 välillä (*vanhe-mm-i-lle-ni-kaan, mikroauto-i-lla-an-ko; Salmi-ko-han*). Pronomineista luku(formatiivi) ei esiinny persoona- ja demonstratiivipronominien kaikissa muodoissa, vaan luku voidaan ilmaista myös vartalon konsonantivaihtelulla. Eräät pronominit sallivat omistusliitteet (*kaikk-i-a-an, oma-a-nsa, itse-ä-än*) mutta komparoitumista ei korpuksen pronomineissa esiinny. Sävy- ja kysymyskliitit esiintyvät pronomineilla yhdessä (*tä-stä-kö-hän, mi-stä-kö-hän syystä*). Pronominien maksimaalinen cwp-arvo on 3–4.

Ns. johtamattomat perusadjektiivit, jotka tyypillisesti ilmaisevat suhteellisia ominaisuuksia, komparoituvat mutta eivät salli omistusliitettä; näiden adjektiivien tekstifrekvenssit ovat korkeat, vaikka niiden lekseemimäärä suhteessa johdettuihin on matala. Absoluuttista materiaalista ominaisuutta ilmaisevat substantiivikantaiset adjektiivit eivät sen sijaan komparoidu (*kankainen* mutta *\*kankaisempi*) ja toisaalta osa

substantiivikantaisista, nimittäin ns. suhteutusadjektiivit sallivat omistuksen ilmaisen (*kaltaise-ni*, *mittaise-ni*). Siten adjektiivien maksimaalinen cpw-arvo vaihtelee kolmesta viiteen (*vaikea-mp-i-a-kin*, *viisaa-mp-i-a-ko*). Kirjoitetun kielen korpuksista ei tosin löydy adjektiiviesimerkkejä, joissa olisi sekä sävy- että kysymyskliitti yhtä aikaa (vrt. *aikaise-mp-i-na-kin vuosina* tai *pitkä-n-kö teetaun*); tällaisen kyllä pystyy konstruoimaan (esim. *pitkä-kö-hän sen pitäisi olla*; *pide-mm-i-ssä-kö-hän ne viat vain olisiivatkin*). Puhutun kielen aineistosta kliittejä ja niiden yhdistelmiä löytyy enemmän (ks. kaaviot 4 ja 5).

Partisiippien taiputusmahdollisuudet ovat suuret jo niiden monifunktioisuuden takia. Esim. muoto *aute-tta-v-i-i(n)sa-kin* sisältää passiivin, 1. partisiipin, monikon, illatiivin ja 3. persoonan omistusliitteen tunnuksset sekä *kin*-kliitin. Funktioittain tarkasteltuina vähimmin taipuvat liittomuodot (luku ja sija: nominatiivi, translatiivi tai – etenkin murteissa – essiivi) ja eniten erilaisiin lauseenvastikkeisiin tms. kuuluvat partisiipit (*näh-tä-vä-kse-en*, *sala-t-u-mp-i-a-kin tietoja*: passiivi, partisiippi, (komparaatio,) luku, sija, (omistus,) (kliitti). Passiivin 2. partisiipin muodoista ei löydy aineistosta sellaista osunaa, jossa olisi erillisenä sekä luvun että omistuksen tunnus. Välillisesti tietysti persoona ilmaisee lukua (*palattua-ni* mutta *palattua-mme*). Komparaatio esiintyy vain eräissä funktioissa (vrt. ed. ja *neuvoja koke-nee-mm-i-lta-an*: 2. partisiippi, komparaatio, luku, ablatiivi, omistus). Sävykliittejä esiintyy partisiipeissa mutta kysymyskliiteistä ei löydy esimerkkejä. Maksimaalinen cpw-arvo vaihtelee partisiipeissa neljästä kuuteen funktiosta riippuen.

Verbien cpw-arvon määrittely on nomineja ongelmallisempaa. Perinteisen näkemys mukaan paitsi indikatiivin myös konditionaalien ja potentiaalien verbimuodot analysoidaan preesensmuotoisiksi; nykykielessä preesensin tunnus on nolla (ks. esim. Penttilä 1963: 472). Tempus ilmaisee suomessa kahtiajakoa mennyt (preteriti) : menemätön (preesens), ja semanttiselta kannalta suomen potentiaali ja konditionaali ilmaisevat juuri menemätöntä aikaa. Silti tempus ja modus tulkitaan toisissa yhteyksissä morfologiselta kannalta toisensa poissulkeviksi (ks. ISK 2004: 139): yhdessä verbin taiputusmuodossa olisi vain joko tempuksen tai moduksen tunnus. Muotoilu johtaa helposti päättelemään, että jos suomessa oletetaan indikatiivin preesens, sen tunnus täytyisi tulkita nollaksi (koska formatiivia ei ole). Konditionaalien ja potentiaalien yhteydessä tämä nolla ei enää olisikaan nolla vaan sitä ei ole lainkaan. Ts. etenkin verbien cpw-arvojen laskemiseksi ns. nollan asemaa on vielä pohdittava erikseen.

Yleensä on totuttu ajattelemaan, että yhtä merkitystä vastaa yksi muoto. Esim. WALs:ssa rakennetaan pitkälti periaatteelle, että jokaista merkitystä ilmaisee ns. vartavastinen (*dedicated*) formatiivi.<sup>2</sup> Periaate on suomen kannalta kovin rajaava, koska suomessa tunnuksset ovat melko harvoin vartavastisia. Esimerkiksi tunnuksset *n*, *k*, *j* ja *t* ovat monikäyttöisiä; vartalovokaalin pidentyminen liittyy paitsi persoonan mm. omistuksen ja illatiivin merkintään. Täytyy siis jotenkin löytää tasapaino tunnusten varta-

---

2. Tämä näkemys tekee agglutinoivasta tyypistä eräänlaisen ihanteen. Näkemys on ongelmallinen mm., koska puhtaat agglutinoivat tyypit ovat vähissä. Toisaalta syntyy myös mielikuva, että kielen muutos, joka usein hävittää agglutinaation edellyttämiä läpinäkyviä morfeemirajoja, olisi jotenkin häiriöluonteinen eikä normaali ilmiö.



vastisuudelle vs. nolllille; nollien kontrolloimaton lisääminen johtaa ilman muuta mielivaltaisuuteen.

Itkosen (2008a: 65–68; 2008b) mukaan taivutuskielessäkin voi ns. kieliopillista perusmerkitystä ilmaista pelkkä taipumaton sanajuuri tai -vartalo; lausetasolla voidaan puhua myös (lause)muodon peruskäytöstä (*basic use*). Useampi kuin yksi (perus)merkitys voi jäädä vaille eksplisiittistä ilmaisua. Tämä on ymmärrettävää ja lisäksi on kielien perimmäisen funktionaalisuuden mukaista, että (perus)käytössä esiintyvä muoto on ”ekonominen”: sillä on korkea frekvenssi ja se on mahdollisimman lyhyt. Suomessa eri taivutuskategorioissa nolllalla merkitään nominatiivi (sijat), yksikkö (luku), aktiivi (pääluokka), preesens (tempus) ja indikatiivi (modus). Persoonista kaikki voidaan merkitä formatiivilla finiittiverbien preesensmuotojen yhteydessä, mutta yksikön 3:nnesta tunnus puuttuu eräissä taivutustyypeissä (esim. *saada*, *haravoida* ja eräät supistumaverbit). Imperfektissä yksikön 3. persoonassa merkitään juuri nolllalla (ja lisäksi sanahahmolla ja siihen sisältyvillä muutoksilla). – Sananloppuisen *-k:n* kato on Suomessa vaikuttanut useamman formatiivin häviämiseen (imperatiivi, preesens, latiivi, kieltoisuuden preesens jne.). Murteista etenkin itäisissä savolaismurteissa *-k* voi edelleen olla säilyneenä (ks. esim. Rapola 1966: 301–302); puhutussa aineistossani *-k:lla* on monenlaista edustumaa.

Puheakteista väite ilmaistaan nolllalla. Myös positiivi voidaan katsoa komparaation perusmerkitykseksi, mutta olisi jo keinotekoista väittää, että komparaation perusmerkitys ilmaistaan nolllalla. Konkreettisista esineistä puhuttaessakaan ei ole koko ajan tarpeen tehdä oletuksia niiden omistajista, vaan (eroavan) omistussuhteen ilmaisu on lisätiedon luonteista (*talo* mutta *taloni*).

Nolllalla ilmaistavien merkitysten onkin ts. oltava jollain lailla fokaalisia: nolllalla ilmaistaan tiettyjä keskeisiä merkityksiä; muiden vähemmän fokaalisten merkitysten (esim. faktiivinen, affirmaatio jne.) läsnäolo kuuluu tutkimuksen implisiittisiin taustaoletuksiin (Itkonen 2008a: 167; 2008b). Keskeisten merkitysten tekstifrekvenssit ovat korkeita; näin on myös LA:n aineistossa (ks. Pajunen & Palomäki 1984; 1985).

Merkityksen keskeisyyden lisäksi nolllalle voidaan antaa kaksi muuta ehtoa (ks. Haas 1958; Korhonen 1969: 22; Itkonen 2008a: 69–74): 1) nollan pitää vaihdella formatiivin kanssa. Näin on esim. suomen 3:nnen persoonan muodoissa (*anta-a* : *saa-Ø*) ja imperatiivin yhteydessä (*tee:Ø* tai *tee'*). Nykykielen kannalta katsoen myös *Ø*-nominatiivi vaihtelee formatiivin kanssa; esim. ISK:n (2004: 109) mukaan nominatiivi on päätteetön vain yksikössä, joten monikon *t* lienee kumulatiivinen (sija ja luku). Preesensin muodoissa voi kielihistorialliselta kannalta edelleen nähdä jäänteitä vanhasta *k*-tunnuksesta monikon 1. ja 2. persoonan muodoissa (ks. Lehtinen 2007: 126), mutta nykykielen kannalta nolla ei varioi formatiivin kanssa. Toisaalta 2) nollan pitää kontrastoida saman paradigman toisia merkityksiä ilmaiseviin muotoihin. Persoonatunnusten välillä on paitsi variaatiota myös kontrastia (yksikön 1. *n* : 2. *t* : 3. *V-V/Ø*), samoin imperatiivissa (*tee'/tee* : *tehkää*) ja sijasyteemissä (nominatiivi *Ø/t* ja muut sijat). Kontrasti löytyy myös pääluokan, luvun, aikamuotojen, moduksen, puheaktin ja affirmaation vs. negaation yhteydessä mutta nolla ei varioi formatiivin kanssa aktiivin, yksikön, indikatiivin, väitteen ja myönteisen merkityksen merkinnässä. Kun mainitut kolme kriteeriä – fokaalinen eli keskeinen merkitys, variaatio ja kontrasti – yhdistetään, on syytä ottaa

suomen cpw-arvojen laskuissa huomioon seuraavat nollat: yksikön 3. persoona, imperatiivin 2. persoona, nominatiivi ja mahdollisesti preesens.

Verbien cpw-arvon määrittely suomessa on vaikeaa juuri nollien ja toisaalta kumulatiivisten formatiivien takia. Yhdessä ja samassa verbissä voi esiintyä pääluokan, tempuksen tai moduksen ja persoonan formatiivi; toisaalta preesensin nolla-tunnus voi liittyä kaikkiin moduksiin. Lisäksi kaikissa finiittimuodoissa voi esiintyä sekä sävyettä kysymyskliitti. Siten verbin maksimaalinen cpw-arvo vaihtelee neljästä viiteen (*voi-ta-isi-in-ko-han, kysy-tt-i-in-kö, ol-isi-n-ko-han*).

Infinitiiveissä on aina infiniittisyyden tunnus (*-a-*, *-e-*, *-ma-* tai *-minen*) sekä sijatunnus; tosin 1. infinitiivin lyhyempi muoto tulkitaan nykyisin sijattomaksi (eikä laatiivi-sijaksi). 1., 2. ja 3. infinitiivi voivat kaikki olla passiivissa, 1. infinitiivi vain murteissa (tarkemmin hämäläismurteissa). Infinitiivit sallivat vain tietyt sijat ja rajaavat myös passiivin esiintymää sijoittain; esim. 3. infinitiivi on passiivissa vain instruktiiivissa (*vanha Moskova piti hävite-ttä-mä-n*). Omistus voi esiintyä 1. infinitiivin pidemässä muodossa ja 2. infinitiivin inessiivissä ja instruktiiivissa sekä 3. infinitiivin abessiivissa (*teh-dä-kse-en, tehd-e-ssä-än, näh-t-e-si, näke-mä-ttä-än*). Lisäksi infinitiivit sallivat sekä kysymys- että sävykliitin (*ryhty-ä-kö, kylkiä kuvaa-ma-lla-ko; haparoid-e-n-kin, hajot-e-ssa-an-kin*). Infinitiivien maksimaalinen cpw-arvo vaihtelee kolmesta kuuteen infinitiivistä ja sen sijamuodosta riippuen.

Sekä nominien että verbien yhteydessä voisi vielä huomioida erikseen sävykliittien *-kin* : *-kAAAn* kyvyn ilmaista polaarisuutta (*kysyy-kin* : *(ei) kysy-kään*) ja toisaalta kongruenssi-ilmiöt. Ne muuttavat lähinnä paradigmaattista kuvaa. Esim. kielteinen *-kAAAn* ei näytä kombinoituvan muiden kliittien kanssa. Toisaalta kongruenssitulkinta laskisi mm. adjektiivien cpw-arvoa, ei nostaisi sitä, jos kongruenssi-formatiivi tulkit-taisiin yhdeksi (toisin kuin luku ja sija). Synkroniselta kannalta katsoen sanahahmojen vaihtelut (astevaihtelu, vokaali- vs. konsonanttivartalo jne.) kantavat myös kielipiilisiä merkityksiä yhdessä formatiivien tai nollan kanssa (vrt. T. Itkonen 1976; Paunonen 1976); niitä ei kuitenkaan pysty laskuissa ottamaan jälkikäteen huomioon. Vastaa-vasti murteiden vaihtelevat käytännöt on tässä yhteydessä sivuutettava; ne edellyttäisi-vät erillistä analyysia.

#### 4 Suomen pääsanaluokkien keskimääräiset teksti-cpw-arvot

Kielen synteettisyyttä ilmaisevia cpw-arvoja lasketaan eri tavoin, kuten jo todettiin. Yksi tapa on laskea yhdessä sanamuodossa esiintyvien formatiivien maksimaalinen lukumäärä (vrt. ed. luku), jolloin kyse on pelkistetyimmillään agglutinaatioindeksistä (vrt. Greenberg 1960). Laskutapa on käytössä tämän hetkisessä typologisessa tutkimuksessa (ks. esim. WALS: luku 22) mutta – kuten edellä todettiin – aina ei tehdä riittävän selvää eroa yhtäältä merkityksen ja muodon ja toisaalta paradigmaattisen ja syntagmaattisen tason välillä puhumattakaan kielianalyysien vaihtelevasta tasosta (tarkemmin ks. Itkonen & Pajunen 2008). Suomesta kirjallisuudessa on esitetty sattumanvaraisia muka syntagmaattisen laskutavan mukaisia arvoja, ja on väitetty mm., että suomen verbien maksimaalinen cpw-arvo olisi 2–3 tai jopa vain 2 (Bickel & Nichols

2005: 96; Dahl 2008: 549; Jääskeläinen 2009). Äkkinäinen voisi tämän hyväksyä (tempus/modus ja persoona), mutta tarkemmin ajatellen aika monta formatiivivaivaa, mm. esimerkiksi *teh-tä-isi-in-kö-hän* kolme. – Mainittakoon, että Bickelin ja Nicholsin argumentaatio perustuu suomen osalta S2-alkeisoppikirjaan; muissa mainituissa lähteissä tyydytään toistamaan WALSn kanta.

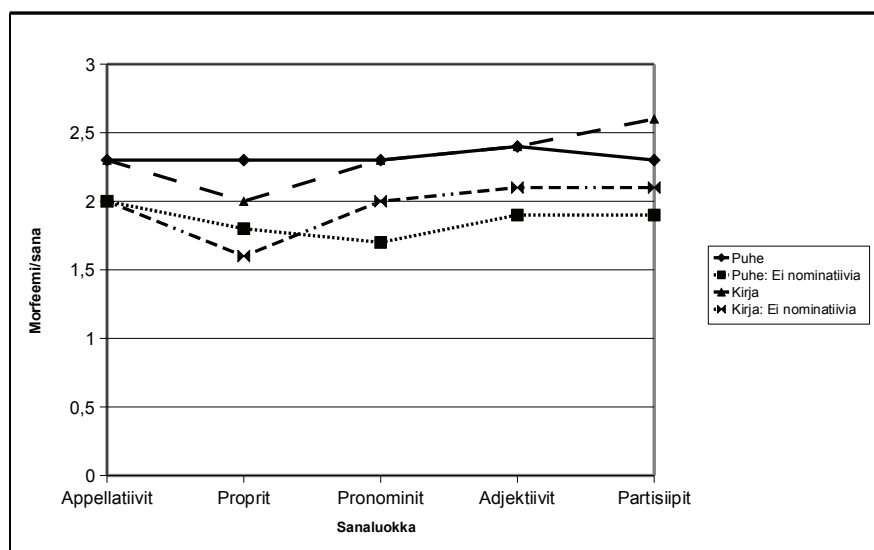
Toinen, ja selvästi parempi tapa arvioida kielen synteettisyyttä on jakaa tunnus-ten esiintymät ja sanojen esiintymät keskenään (Greenberg 1960; Korhonen 1969: 32); nämä laskelmat tehdään yleensä juoksevasta tekstistä. Nimitän keskimääristä cpw-arvoa teksti-cpw-arvoksi ja tarkastelen tässä yhteydessä teksti-cpw-arvoja valituissa suomen sanaluokissa.<sup>3</sup> Laskelmissa on aina mukana – Greenbergin tapaan – leksikaalisten kategorioiden minimi, joka on yksi; muuten ero analyttisiin kieliin ei säily. Analyttisissä kielissä näet tämä yksi voi olla joko leksikaalinen tai kieliopillinen. Aineisto ei anna mahdollisuuksia derivaation tai yhdyssanojen määrien huomioimiseen. Tämä näyttää häivyttävän puhutun ja kirjoitetun kielen sanarakenteiden kompleksisuuseroa, joka heijastuu mm. sanapituuksiin. Sanapituudet ovat puhutussa aineistossa keskimäärin kolmisen merkkiä kirjoitettua alemmat (ks. Pajunen & Palomäki 1984: 70–71).

Lauseopin arkiston aineistossa on n. miljoona sanamuotoa. Aineistosta on tätä tutkimusta varten poistettu keskeneräiset sanat ja tulkinnaltaan epävarmat analyysit, joten puhutun kielen aineistossa on n. 887 000 tekstisanaa ja kirjoitetun n. 160 000 tekstisanaa. Kummastakin aineistosta on laskettu sekä appellatiivien että proprien, pronomien, adjektiivien, partisiippien ja verbien teksti-cpw-arvot. Joitain tarkistuksia on tehty Pajunen ja Palomäen laskelmista, jotka perustuvat otokseen LA:n aineistosta. Esitän tulokset kaavioissa 1–8.

Nominisanaluokissa teksti-cpw-arvoon on laskettu seuraavia funktioita ilmaisevat formatiivit: komparaatio, monikko, sija, omistus, sävy, kysymys ja passiivi, joka esiintyy partisiippien yhteydessä. Samat arvot on kuvattu myös jättämällä nominatiivi-nolla huomioimatta (ks. kaavio 1).

Teksti-cpw-arvon vaihteluväli on 2,3–2,4 puhutun kielen nominiluokissa (siis yksi leksikaalinen morfeemi ja 1,3–1,4 kieliopillista morfeemia / sana). Kirjoitetun kielen aineistossa vaihtelu on hieman suurempi: 2–2,6. Appellatiivien, pronomien ja adjektiivien teksti-cpw-arvo on puhutussa ja kirjoitetussa sama, partisiippeja taivutetaan kirjoitetussa kielessä enemmän kuin puhutussa (2,3 vs. 2,6), propreja vähemmän (2,0 vs. 2,3). Kun arvoista karsitaan nominatiivin vaikutus, puhutun kielen nominien cpw-arvot putoavat odotetusti kirjoitetun kielen vastaavia alemmiksi propreja lukuun ottamatta. Syy on ilmeinen: puhutussa kielessä on enemmän nominatiivisijaisia peruslauseenjäseniä, erityisesti pronominisubjekteja ja syntaktinen rakenne on vähemmän kompleksinen eli tässä hierarkkinen. On hieman odotuksenvastaista, että puhutun ja kirjoitetun kielen nominien taivutus ei eroa enempää, mutta tästä voinee päätellä, että juuri sananmuodostus vaikuttaa kirjoitetun kielen korkeampiin sanapituuksiin eikä niinkään taivutus.

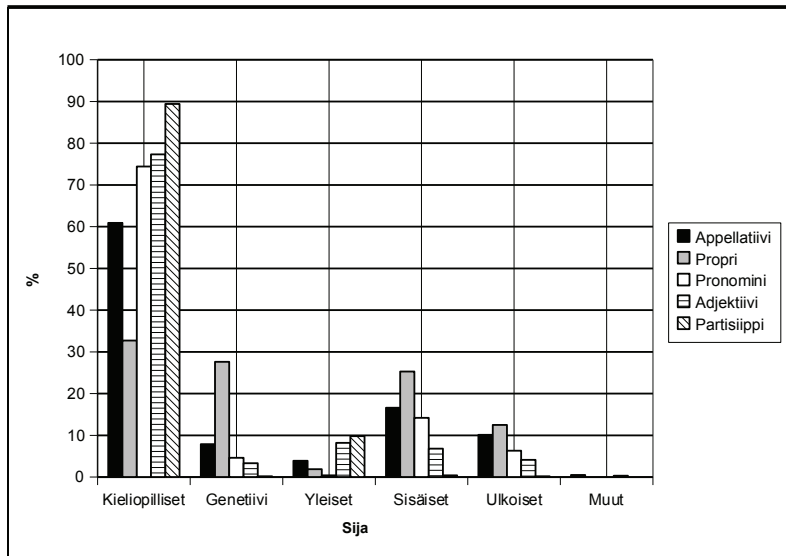
3. Laskelmat on alkuaan tehty toista tutkimusta varten, ja sanaluokkien valinta on silloisten tavoitteiden rajaama.



**Kaavio 1.**

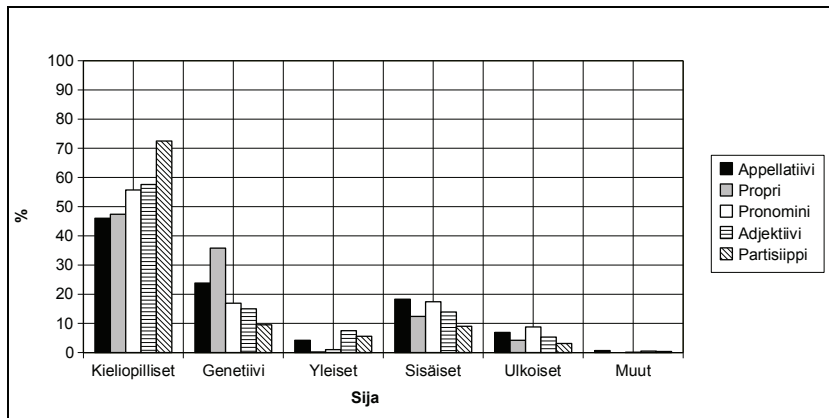
Teksti-cpw-arvot LA:n aineiston nomineissa. LA:n puhutun kielen aineistossa sanoja on 887 307, nomineja 347 117, kirjoitetun kielen aineistossa sanoja on 161 301, joista nomineja 110 615.

Eri nominityypit eivät kuitenkaan taivu sijoissa tasaisesti (ks. kaaviot 2 ja 3). Puhutun kielen appellatiivit taipuvat odotuksen mukaisesti kaikissa sijoissa, muista nominityypeistä marginaaliset sijat (eli abessiivi, komitatiivi ja instruktiivi) lähes puuttuvat (vrt. myös kaavioiden 4 ja 5 muut sijat). Appellatiivien ns. maksimaalinen sijataivutus heijastuu mm. siitä, että appellatiiveista nominatiivisijaisia on vain n. 30 %, pronomineista ja adjektiiveista n. puolet ja partisiipeista jo n. 85 %. Adjektiivien sija määräytyy joko rektion kautta (essiivi ja translatiivi, ks. kohta yleiset kaavioissa 2 ja 3), kopulakonstruktion tai sitten kongruenssin perusteella (etenkin subjektin ja objektin määrittäminen). Partisiipit toimivat joko liittomuotojen osina tai tilanilmaisuuksina; essiivi esiintyy suhteessa yleisimmin juuri partisiipin sijana. Propreilla viitataan paitsi tekijään tai kohteeseen, omistajaan tai lokaatioon. Kirjoitetussa kielessä appellatiivien sijajakauma on vielä puhuttuakin tasaisempi; nominatiivin osuus on vain 25 %. Molemmissa ryhmissä marginaalikoodaus muissa kuin substantiiveissa perustuu LA:n osin kielihistoriallisiin koodauskriteereihin (ks. Ikola 1985).



Kaavio 2.

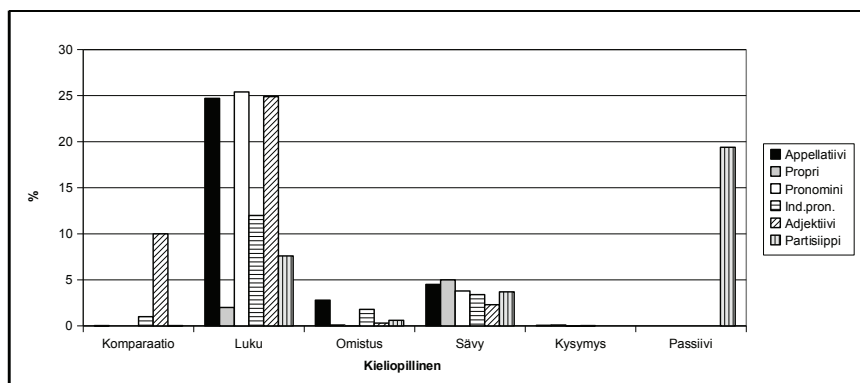
Nominien sanaluokka- ja sijajakauma LA:n puhutun kielen aineistossa. Appellatiivit  $n=130\,807$ , propriit  $n=12\,509$ , pronominit  $n=165\,451$ , adjektiivit  $n=25\,786$ , partiisiipit  $n=25\,853$ .



Kaavio 3.

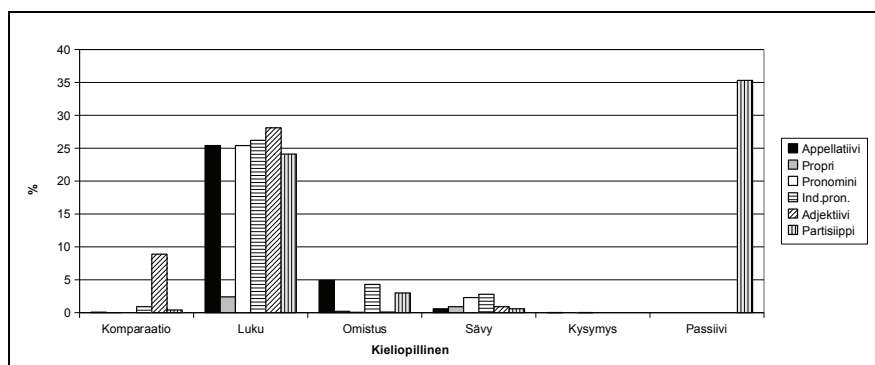
Nominien sanaluokka- ja sijajakauma LA:n kirjoitetun kielen aineistossa. Appellatiivit  $n=63\,085$ , propriit  $n=9\,694$ , pronominit  $n=16\,002$ , adjektiivit  $n=13\,480$ , partiisiipit  $n=11\,509$ .

Nominin taipuvuus sijassa, joten sijaformatiivin (mukana nominatiivi) vaikutus teksti-cpw-arvoon on yksi. Puhutussa kielessä muut nominintaivutusluokat nostavat cpw-arvoa vaihdellen 0,2–0,4 ja kirjoitetussa 0,4–0,6. Tarkennan näiden vaikutusta edelleen (ks. kaaviot 4 ja 5). Kaavioiden lukuohje on seuraava: komparaation tunnus liittyy n. prosenttiin kaikista indefiniittipronomineista, n. 10 prosenttiin kaikista adjektiiveista jne.



**Kaavio 4.**

Eräiden formatiivien esiintyminen LA:n puhutun kielen nominien alaluokissa (% sanaluokan sanojen määrästä). Appellatiivit n=130 807, propriit n=12 510, pronominit n=146 636, indefiniittipronominit n=18 825, adjektiivit n= 25 786 ja partisiipit n=25 853.

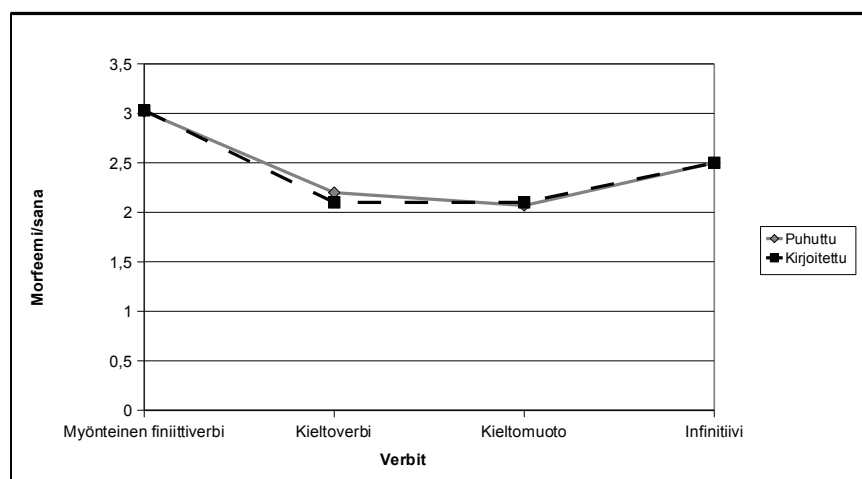


**Kaavio 5.**

Eräiden formatiivien esiintyminen LA:n kirjoitetun kielen nominien alaluokissa (% sanaluokan sanojen määrästä). Appellatiivit n= 63 087, propriit n=9 694, pronominit n=11 703, indefiniittipronominit n=4 368, adjektiivit n=13 480 ja partisiipit n= 11 509.

Komparaation tunnus liittyy aineistossa adjektiiveihin ja indefiniittipronomineihin ja kirjoitetussa kielessä myös muutamaan partisiippiin. Monikossa on keskimäärin neljännes appellatiiveista, persoona- ja demonstratiivipronomineista sekä adjektiiveista. Propriit ovat molemmissa aineistoissa useimmiten yksikössä. Puhutussa kielessä indefiniittipronominien ja partisiippien yksiköllisyys on vallitsevaa mutta kirjoitetussa kielessä nämä nominityyppit eivät eroa yleistendenssistä. Omistusmuodot esiintyvät samoissa nominityypeissä mutta ovat kirjoitetussa kielessä yleisempiä. Sävyformatiivi esiintyy aika tasaisesti puhutun kielen nominityypeissä, useimmiten substantiiveissa, mutta kirjoitetussa kielessä sävyformatiiveja käytetään paljon vähemmän. Kysymysformatiivit ovat nomineissa molemmissa aineistoissa harvinaisia. Passiivin partisiipit ovat kirjoitetussa kielessä paljon puhuttua yleisempiä; ero on yli 15 prosenttiyksikköä.

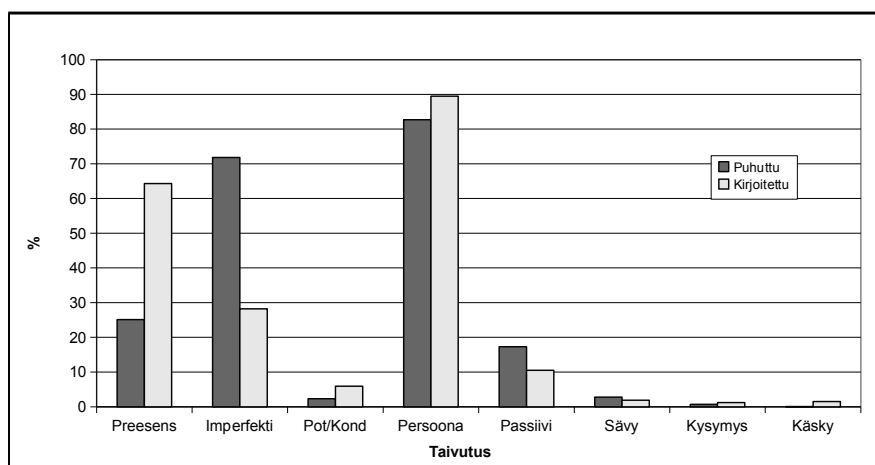
Verbien teksti-cpw-arvo on laskettu myönteisistä finiittiverbeistä koko aineistosta ja osasta aineistoa on tiedot kieltoverbeistä ja kieltomuodoista sekä infinitiiveistä (ks. kaavio 6). Tämän otoksen sanamäärä on puhutusta 200 000 ja kirjoitetusta kielestä 20 000. Myönteisten finiittiverbien teksti-cpw on puhutussa 3,02 (1 leksikaalinen ja 2,02 kieliopillista morfeemia) ja kirjoitetussa aineistossa lähes sama (3,03). Aineiston kielteiset finiittiverbit koostuvat kieltoverbistä (*ei, älä*) ja siihen liittyvästä leksikaalisen verbin kieltomuodosta (*ota, ottane, ottaisi, oteta*); edellisiä on aineistossa enemmän kuin jälkimmäisiä, koska kieltomuoto ilmaisee vain menemätöntä aikaa. Kieltoverbiin liittyviä formatiiveja ovat persoona ja kysymys- ja sävykliitit; kliiteistä *kä* liittyy ainoastaan kieltoverbiin. Kieltoverbin modus ilmaistaan leksikaalisesti. Kieltomuotoon koodataan pääluokka, (presens)tempus (muut aikamuodot ilmaistaan partisiippimuodoilla), modus, kysymys- ja sävykliitit. Kieltoverbin ja kieltomuodon teksti-cpw-arvo on molemmissa aineistoissa hieman päälle kahden. Infinitiiveihin liittyy aina infinitiivitunnus; muita mahdollisia formatiiveja ovat passiivi, sija, omistus, kysymys- ja sävykliitit. Infinitiivien teksti-cpw-arvo 2,5.



#### Kaavio 6.

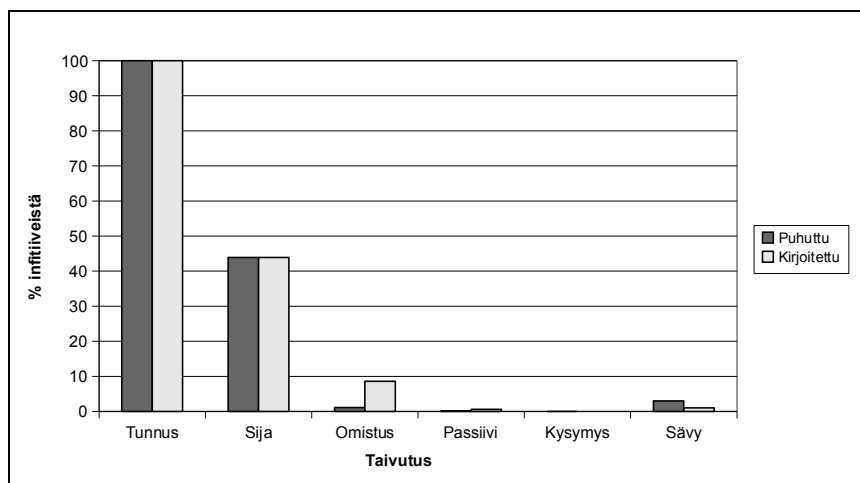
Teksti-cpw-arvot LA:n puhutun ja kirjoitetun kielen aineiston verbeissä. Myönteiset finiittiverbit puhutussa  $n=128\,444$  ja kirjoitetussa  $23\,368$ , kieltoverbit  $n=25\,659$  ja  $2\,850$ , kieltomuodot  $n=9\,027$  ja  $2\,079$  sekä infinitiivit  $n=4\,234$  ja  $619$ .

Tarkastelen vielä myönteisiä finiittiverbejä (kaavio 7) ja infinitiivejä (kaavio 8) tavutuskategorioittain (formatiivi x %:ssa sanoja).



**Kaavio 7.**

Taivutusformatiivien määrä LA:n aineiston myönteisissä finiittiverbeissä. Finiittiverbejä puhutun kielen aineistossa 128 444, kirjoitetun kielen aineistossa 23 368.



**Kaavio 8.**

Taivutusformatiivien määrä LA:n puhutun ja kirjoitetun kielen infinitiiveissä. Infinitiivit puhutun kielen aineistossa n=4 234, kirjoitetun kielen aineistossa n=619. Laskelmat on tehty Pajusen & Palomäen (1984; 1985) aineistosta.

Myönteisissä finiittiverbeissä taivutusformatiivit ovat seuraavat: preesens (Ø nykykielessä), imperfekti, potentiaali ja konditionaali, persoona (yksikön ja monikon 1–3), passiivi, sävy ja formatiiveilla merkittävistä puheakteista kysymys ja käsky (ks. kaavio 7). Preesensmuoto on laskettu vain indikatiiveihin, mutta preesensin ”oikea” määrä saadaan laskemalla yhteen indikatiivin preesens sekä potentiaali ja konditionaali (puhuttussa n. 27 % ja kirjoitetussa n. 73 %). Passiivimuodoista ei ole eroteltu passiivin ja persoonan tunnuksia, mutta jos näin haluttaisiin tehdä, persoonasarake nousisi 100 %:iin



molemmissa aineistoissa. Sävykliittejä ei käytetä finiittiverbeissä yhtä usein kuin nomineissa; sen sijaan kysymyskliitit ovat juuri verbeille ominaisia, vaikka niiden määrä aineistossa ei ole suuri. Puhutussa ja kirjoitetussa aineistossa finiittiverbien taivutuksessa ei taulukon perusteella ole kovin suuria eroja tempusta lukuun ottamatta. Tempusmerkinnän jakauma on lähes vastakkainen; kyse on tekstilajierosta. Ero heijastuu myös persoonamerkintään, koska imperfektissä yksikön 3. persoona on tunnuksen eli Ø. Jos nollien vaikutus teksti-cpw-arvoihin halutaan häivyttää, olisi puhutun aineiston persoonamerkintää kaavion 7 tietojen perusteella liikaa. Kirjoitetun kielen aineistossa persoonamerkinnän osuus olisi paremmin kohdallaan, koska preesensissä yksikön 3. persoona merkitään formatiivilla. Teksti-cpw-arvoissa puhutun ja kirjoitetun kielen ero tasaantuisi, koska nollat jakaantuvat eri aineistoihin. – Mainittakoon vielä, että kielteisten finiittimuotojen osuus kaikista finiittiverbeistä on puhutussa aineistoissa n. 17 % ja kirjoitetussa n. 11 %.

Infinitiiveissä on aina infiniittisyyden tunnus (100 %; ks. kaavio 8). Sijasarakkeesta voi päätellä, että yli puolet infiniitiveistä on 1. infiniitin latiiveja (*tehdä, olla*); latiiivia ei ole laskettu infiniitin sijaksi. Omistusformatiivi on puhutussa aineistossa vähässä käytössä mutta kirjoitetussa infiniitiveistä lähes 10 % on omistusmuotoisia. Passiivi- ja kysymysformatiivi ovat vähässä käytössä, sävykliittiä esiintyy puhutussa aineistossa kirjoitettua enemmän.

Suomen deklinoivissa ja konjugoivissa sanaluokissa teksti-cpw-arvot vaihtelevat kahden ja kolmen välillä. Puhutun ja kirjoitetun kielen välillä ei aineistoissa ole suuria eroja. Jos samasta aineistosta lasketaan kaikkien tekstisanojen cpw-arvot, jäädyään selvästi alle kahden. Puhutun kielen aineiston kaikkien sanojen teksti-cpw-arvo on myös selvästi alempi kuin kirjoitetun ja niin ikään murteiden välillä on eroja (ks. Pajunen ja Palomäki 1984: 45). Eroihin näyttäisi eniten vaikuttavan taipumattomien sanojen määrä otoksesta; näitä on LA:n korpuksessa selvästi enemmän puhutun kuin kirjoitetun kielen aineistossa. Murteiden osalta taipumattomia sanoja on eniten lounaismurteiden aineistossa. Teksti-cpw-arvoja ei kannatakaan laskea puhutun kielen kokonaisuusmäärästä taipumattomien sanojen vaihtelevan ja arvoja vinouttavan määrän takia.

## 5 Koonti

Kielten synteettisyyttä voidaan laskea usealla tavalla ja synteettisyyttä ilmaisevia mittareita on vastaavasti useita. Kielten vertailuun käy luonnollisesti vain samoin periaattein toimivat mittarit.

Olen edellä kuvannut suomen keskeisten sanaluokkien synteettisyyttä mahdollisen eli paradigmaattisen ja syntagmaattisen eli todella esiintyvän taivutuksen avulla. Edellinen arvioi synteettisyyden rajoja ja jälkimmäinen tuo eri tuloksia mm. sen mukaan, miten suhtaudutaan ns. vartavastisiin ja kumulatiivisiin formatiiveihin ja ns. nolliin. Edelleen syntagmaattisen taivutuksen avulla voidaan laskea joko sanaluokan maksimaalista tai tekstuaalista cpw-arvoa. Edellinen luku antaa tulokseksi nominiluokkien tai verbien eniten taipuvan sanan formatiivimäärän, jälkimmäinen näiden sanaluokkien keskimääräisen taipumisen määrän. Jos nämä laskelmat perustuvat vain varta-

vastisiin formatiiveihin, kyse on agglutinaatioindeksistä. On laskentaperusta mikä tahansa, suomen kielestä viimeaikaisessa tyologisessa kirjallisuudessa esitetyt cpw-arvot ovat virheellisiä.

Suomen nominien ja verbien maksimaalinen cpw-arvo on viisi; arvo on sama, vaikka laskennassa edellytettäisiin joka kieliopilliselle morfeemille ilmitunnus eli laskettaisiin agglutinaation astetta. Englannin maksimaalinen cpw-arvo on yksi. Keskimääräinen teksti-cpw vaihtelee suomessa kahdesta kolmeen sekä nominien että verbien alaluokissa; englannissa vastaava arvo on Greenbergin (1960: 193) mukaan 1,68 (luku sisältää myös derivaation). Suomen verbien teksti-cpw-arvo on nomineja korkeampi hieman epäodotuksenmukaisesti. Tilanne on kielissä yleinen mutta suomessa edellyttää tiettyjen nollaformatiivien sisällyttämistä laskelmiin. Myös puhutun ja kirjoitetun kielen teksti-cpw-arvojen samankaltaisuus on hieman vastoin odotuksia mutta tulosten perusteella näyttää siltä, että pelkästään taivutuksen määrä ei erota puhuttua ja kirjoitettua kieltä selvästi toisistaan; erot näkyvät paremmin myös funktiot huomioitaessa.

Suomen verbien paradigmaattinen cpw-arvo on kuusi, englannin kaksi. On siis helppo todeta, että suomen ja englannin samaistaminen synteettisyysasteen perusteella WALS:in esittämään tapaan on perusteetonta. Vertailun lopputulos ei silti tarkoita, että suomen kieli olisi maailman muihin kieliin verrattuna erityinen tai eksoottinen (ks. tarkemmin Pajunen 2002).

## Lähteet

### Aineisto

LA =Lauseopin arkisto. N. miljoonan sanan manuaalisesti analysoitu, aluemurteita edustava aineisto; täydennyksenä edustavasti valittu kirjoitetun kielen aineisto. Suomalaisen ja yleisen kielitieteen laitos, Turun yliopisto.

PAJUNEN, ANNELI 2002: SS2000-korpus. *Savon Sanomien* vuoden 2000 aineisto. FDG-analysoitu (Connexor) ContextMill-tietokanta (Virtanen).

— 2003: HS2000-korpus. Vuosien 2000 ja 2001 *Helsingin Sanomien* aineisto. FDG-analysoitu (Connexor) ContextMill-tietokanta (Virtanen).

### Muut lähteet

BICHEL, BALTHASAR – NICHOLS, JOHANNA 2005: Inflectional synthesis of the verbs. – Martin Haspelmath, Matthew S. Dryer & Bernard Comrie (toim.), *The World atlas of language structures* s. 94–97. <http://www.wals.info>.

CROFT, WILLIAM 1991: *Syntactic categories and grammatical relations*. Chicago: The University of Chicago Press.

DAHL, ÖSTEN 2008: Kuinka eksoottinen kieli suomi on? – *Virittäjä* 112 s. 545–559.

GIL, DAVID 2008: How complex are isolating languages? – Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (toim.), *Language complexity* s. 109–131. Amsterdam: John Benjamins.

- GREENBERG, JOSEPH H. [1954] 1960: A quantitative approach to the morphological typology of languages. – *International Journal of American Linguistics* 26 s. 178–194.
- HAAS, WILLIAM 1958: Zero in linguistic description. – W. Haas (toim.), *Studies in linguistic analysis* s. 33–53. London: Blackwell.
- HAIMAN, JOHN 1985: *Natural syntax*. Cambridge: Cambridge University Press.
- HASPELMATH, MARTIN – DRYER, MATTHEW S. – COMRIE, BERNARD (toim.) 2005: *The world atlas of language structures* [= WALS]. Oxford: Oxford University Press.
- IKOLA, OSMO (toim.) 1985: *Lauseopin arkiston opas*. Lauseopin arkiston julkaisuja 1. Turku.
- 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.
- ITKONEN, ESA 2008a. *Maailman kielten erilaisuus ja samuus*. I osa. Yleisen kielitieteen julkaisuja 12. Turku: Turun yliopisto.
- 2008b. The blind spot of contemporary typological linguistics: The role of zero. Plenary talk given at LCM3 (Odense, July 2008). <http://users.utu.fi/eitkonen/Typology.htm>.
- ITKONEN, ESA – PAJUNEN, ANNELI 2008: *A few remarks on World Atlas of Language Structures*. <http://users.utu.fi/eitkonen/Typology.htm>.
- ITKONEN, ESA – PAJUNEN, ANNELI – SUOMI, KARI 2009: Maailman kielten luokituksen ongelmia. [Arvostelu teoksesta *The World Atlas of Language Structures*. Toim. Martin Haspelmath, Matthew S. Dryer & Bernard Comrie. Oxford University Press. 2005.] – *Virittäjä* 113 s. 285–296.
- ITKONEN, TERHO 1976: Syntaktisten vaikutusyhteyksien luonteesta. – *Virittäjä* 80 s. 52–81.
- JÄÄSKELÄINEN, ANNI 2009: Suomi on yhtä aikaa kumma ja tavis. – *Tiede* 2/2009 s. 38–41.
- KORHONEN, MIKKO 1969: Die Entwicklung der morphologischen Technik im Lappischen. – Tapani Salminen (toim.) 1996, *Typological and historical studies in language by Mikko Korhonen* s. 17–143. Suomalais-Ugrilaisen Seuran toimituksia 223. Helsinki: Suomalais-Ugrilainen Seura.
- KUSTERS, WOUTER 2003: *Linguistic complexity. The influence of social change on verbal inflection*. <http://www.loty.let.uu.nl/>.
- LEHTINEN, TAPANI 2007: *Kielen vuosituhanet. Suomen kielen kehitys kantaauralista varhais-suomeen*. Tietolipas 215. Helsinki: Suomalaisen Kirjallisuuden Seura.
- LESKINEN, HEIKKI 1970: *Imperatiivin muodostus itämerensuomalaisissa kielissä*. I Suomi. Suomi 115:2. Helsinki: Suomalaisen Kirjallisuuden Seura.
- PAJUNEN, ANNELI 1994: Adjektiivikategorian universaaliudesta. – *Virittäjä* 98 s. 513–542.
- 1998a: Adjectives in spoken language discourse. – *Word* 49 (3) s. 341–368.
- 1998b: Pääsanaluokkien eriytymättömyydestä uralilaiskielissä. – Anneli Pajunen (toim.), *Kieliopillistumisesta, analogiasta ja typologiasta* s. 59–109. Helsinki: Suomalaisen Kirjallisuuden Seura.
- 2001: *Argumenttirakenne. Asiantilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Suomi 187. Helsinki: Suomalaisen Kirjallisuuden Seura.
- 2002: Suomen kieli maailman kielten joukossa. – *Virittäjä* 106 s. 563–569.
- PAJUNEN, ANNELI – PALOMÄKI, ULLA 1984: *Tilastotietoja suomen kielen rakenteesta* 1. Kotimaisten kielten tutkimuskeskuksen julkaisuja 30. Helsinki: Kotimaisten kielten tutkimuskeskus.
- 1985: *Tilastotietoja suomen kielen rakenteesta* 2. Kotimaisten kielten tutkimuskeskuksen julkaisuja 31. Helsinki: Kotimaisten kielten tutkimuskeskus.
- PAUNONEN, HEIKKI 1976: Allomorfiin dynamiikkaa. – *Virittäjä* 80 s. 82–105.

- PENTTILÄ, AARNI 1963: *Suomen kielioppi*. Porvoo: WSOY.
- RAPOLA, MARTTI 1966: *Suomen kielen äännehistorian luennot*. Suomalaisen Kirjallisuuden Seuran toimituksia 283. Helsinki: Suomalaisen Kirjallisuuden Seura.
- RIDDLE, ELIZABETH M. 2008: Complexity in isolating languages. Lexical elaboration versus grammatical economy. – Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (toim.), *Language complexity* s. 133–151. Amsterdam: John Benjamins.
- VAINIO, SEPPÖ – HÖYNÄ, JUKKA – PAJUNEN, ANNELI 2008: Processing modifier-head agreement in reading. Evidence for a delayed effect of agreement. – *Memory and Cognition* 36 s. 329–340.
- WALS = *The World atlas of language structures*. Toim. Martin Haspelmath, Matthew S. Dryer & Bernard Comrie 2005. <http://www.wals.info>.

## SUMMARY

### Word syntheticity in Finnish

The languages of the world are often compared according to the morphosyntactic complexity in a word, i.e. the amount of lexical and grammatical information encoded in a word. In analytic languages, one free morpheme expresses one lexical or grammatical meaning, while in synthetic languages one free morpheme expresses one lexical meaning and bound morphemes attached to it then express one or more grammatical meaning (e.g. case, tense, mood etc.). In polysynthetic languages both free and bound morphemes can express lexical meanings and there can be many bound morphemes in a single word. Category-per-word (or CPW) value expresses the amount of morphemes typical in a language. The CPW value in Finnish is said to be over two, which means that Finnish is a synthetic language; according to Greenberg (1960) the CPW value of English is 1,68, which means that English is an analytic language. Lately, the CPW value of Finnish has been questioned (see Balthasar & Nichols 2005, Dahl 2008).

This article shows that there are three different ways to define CPW values, the first being possible (paradigmatic) inflection in a word class, the second maximal (syntagmatic) inflection in a word form, and the third based on the inflection of word forms in a text corpus. The article demonstrates that while the possible inflection of Finnish verbs consists of six grammatical categories (voice, tense, mode, person, tone, speech act), the maximal inflection of a Finnish verb is only four or five (e.g. voice, tense, speech act, tone or tense, mood, person, speech act, tone).

Finally, the textual CPW value given for verbs is three; this value is based on a million-word corpus from the Finnish syntax archives. The conclusion is that English and Finnish CPW values can be the same only if the possible English verb inflection is compared to the average Finnish verb inflection. This, however, is not a valid mode of comparison.

Kirjoittajan yhteystiedot (address):  
suomen kieli  
33014 Tampereen yliopisto  
etunimi.sukunimi@uta.fi