

VERBISANASTON UUDISTUMINEN¹

Anneli Pajunen, Kieli- ja käännöstieteiden laitos, Tampereen yliopisto

Artikkeli käsittelee verbilekseemien uudistumista suomen kielessä. Avoimet luokat uudistuvat oletuksen mukaan nopeasti, mutta suomen kielessä oletus ei päde verbisanaluokasta. Tulokset pohjautuvat korpus tutkimukseen, jossa on käytetty toisaalta 31 milj. sanaesiintymän sanomalehtiaineistoa (Helsingin Sanomien vuosikerrat 2000 ja 2001) ja toisaalta nyky-suomen sanakirjojen aineistoa (Nyky-suomen sanakirja 1951–1961; Uudissanasto 80; CD-Perussanakirja 1997). Nyky-suomen sanakirjan ja sanomalehtikorpuksen verbilekseemeistä 80 % on yhteistä, esiintymistä yhteinen aines kattaa yli 99.5 %. Huomattava osa korpuksen uudesta aineksesta on frekventatiivijohdoksia tutuista verbikannoista. Sanakirjoista oikeasti puuttuva aines koostuu puhekielisyyksistä, lainasanoista ja uusia teko- ja valmistusprosesseja nimeävistä muutosverbeistä. Sanarakenteeltaan uudisverbit ovat kompleksisia, mukana on sekä johdoksia että yhdysverbejä, sanojen merkkimäärä on vastaavasti perussanastoa korkeampi.

Avainsanat: verbileksikko, uudismuodoste, frekvenssi, hapax legomenon -lekseemi.

LESIKAALISET MORFEEMIT JA UUDISTUMINEN

Leksikaaliset morfeemit erotellaan kieliopillisista tyyppillisesti seuraavien ominaisuuksien avulla: a) puheenaihe määrää tietyn leksikaalisen morfeemin eli lekseemin esiintymisen, ei syntaksi, b) lekseemeillä on suhteellisen konkreettinen merkitys, c) yksittäiset lekseemit esiintyvät melko harvoin verrattuna kieliopillisten morfeemien esiintymisiin, d) lekseemejä on paljon ja uusia syntyy jatkuvasti. Leksikaaliset morfeemit jaotellaan avoimiin ja suljettuihin sanaluokkiin, ja juuri avoimen luokan jäsenillä, kuten substantiiveilla ja verbeillä, on edellä mainitut ominaisuudet. Avointen ja suljettujen luokkien määrittely tällä tavoin on tavanomaista ja toistuu alan

käsikirjoissa (ks. esim. Hudson, 2000: 61–62; Schachter, 1985; leksikaalisesta ja kieliopillisesta merkityksestä ks. Itkonen, 2001: 20–28). Tavoitteeni on tässä yhteydessä arvioida kriteereistä viimeistä eli uudistumisvältettä suomen verbisanaston osalta.

Avointen luokkien jäsenten välisiä uudistumiseroja ei kirjallisuudessa ole tietääkseni käsitelty, mutta mahdolliset erot antaisivat aineksia pohtia paitsi verbien luonnetta myös substantiivi- ja verbisanaston kielitypologisia eroja. Verbit ovat substantiivien tavoin universaalisanaluuokka mutta tiedetään, että maailmassa on kieliä, joissa verbit yllättäen kuuluvat suljettuun luokkaan. Tunnetuin tällaisista kielistä on kalam (ks. Pawley, 1987) mutta ilmiö on tuttu muistakin papualaiskielistä ja lisäksi australialaiskielistä; Everett (<http://www.linguistlist.org>, viitattu 20.2.2005) lukee joukkoon myös eteläame-

Kirjoittajan yhteystiedot: Anneli Pajunen
Kieli- ja käännöstieteiden laitos,
Kanslerinrinne 1, 33014 Tampereen yliopisto
Sähköposti: anneli.pajunen@uta.fi

¹ Kiitän Sanoma Oy:tä oikeudesta käyttää Helsingin Sanomien aineistoa.

rikkalaisen piraha-kielen. Palaan lopussa lyhyesti siihen, miten suomen verbisanaston uudistumisnopeus auttaa eroja pohtimaan.

Arvioin verbien uudistumista seuraavasti: Tarkastelen seuraavassa luvussa, miten sanaston uudistumista tutkitaan, mitä oletuksia uudistumisnopeudesta on esitetty ja miten sanan uutuus määritellään. Uutuuden määrittelyä pohdin tarkkaan erityisesti kahdesta syystä. Ensinnäkin suomen tapaisessa kielessä uudisverbit ovat tyypillisesti johdoksia, ja koska niiden kantasana on useimmiten kielen vanha sana, johdoksen uutuus voi olla suhteellista. Toiseksi tutkimuksen aineisto on suuresta, automaattisesti jäsennetystä korpuksesta, joka rajaamattomana ja tarkistamattomana tuottaisi paljon myös tulkinnanvaraisia uudisverbejä. Oman aineistoni ongelmia ja edustavuutta käsittelen etenkin seuraavan luvun loppujaksoissa. Analyysiluvussa keskityn aineiston lekseemien ja esiintymien suhteeseen vertaamalla ensin kielen vakiintuneeseen ainekseen uudisverbejä, jotka merkki merkiltä arvioituna puuttuvat vertailumateriaalina olevasta Nykysuomen sanakirjasta. Erikseen käsittelen sanakirjan ja korpuksen yhteistä ainesta sekä niitä uudisverbejä, jotka puuttuvat sanakirjasta myös silloin, kun uudissanastatus määritellään tiukemmin. Loppulukuun kokoan tuloksia.

SANASTON TUTKIMINEN

Sanaston uudistumisen tutkimuksessa on kaksi päämetodia (ks. esim. Bauer, 2001: 143–161). Ensimmäinen vertaa kahta ajallisesti eriaikaista aineistoa toisiinsa. Vanhempana aineistona on yleensä laaja deskriptiivinen sanakirja ja uudempana sähköinen korpus. Toinen mittaa erilaisten produktiivisuusmittarien avulla morfologista produktiivisuutta. Kummassakin metodissa ns. hapax legomenon -lekseemit (= HL-lekseemit) eli yhden esiintymän sanat ovat tärkeässä ase-

massa. Vertailussa suurin osa uudismuodosteista on HL-lekseemeitä. Produktiivisuuden mittauksessa HL-tyyppien määrä suhteutetaan tokeneihin. Lasketaan esim. tietyn johtimen tuottamat HL-lekseemit ja jaetaan tulos kaikilla johtimen tuottamilla esiintymillä.

Omassa verbiaineistossani uudismuodosteilla on odotuksen mukaan vähän esiintymiä, ja suuri osa niistä on juuri HL-lekseemejä. Huomio kiintyy niin ikään tiettyihin uusiin johdoksiin, mutta en puutu johtimien produktiivisuuteen tässä yhteydessä, koska mittarien käyttöön liittyy monenlaisia ongelmia (ks. esim. Bauer, 2001). Keskityn yksin ajallisesti eriaikaisten aineistojen, nimittäin sanakirjan eli Nykysuomen sanakirjan (= NS) ja korpuksen eli Helsingin Sanomien (= HS) vuosien 2000 ja 2001 aineiston vertaamiseen.

On huomattava, että HL-lekseemien määrä nousee luotettavasti vasta, kun otoskoko on riittävän suuri, joten verbien uudistumisen tutkimuksessa kysymykseen tulee vain iso massakorpus (vrt. Baayen & Renouf, 1996: 74–76). HS-aineiston sanaesiintymien määrä on yli 31 miljoonaa ja luku lienee riittävä, koska saamani tulokset ovat hämmästyttävän samansuuntaiset kuin vastaavat Savon Sanomien aineistosta laskemani.

Oletukset sanaston uudistumisnopeudesta

Yleisoletus on, että kielen sanasto muuttuu ja uudistuu nopeasti verrattuna esim. kieliopillisiin morfeemeihin. Ero avointen ja suljettujen luokkien jäsenten välillä oletetaan vastaavaksi. Tämä yleisoletus onkin nähdäkseni kiistaton. Sen sijaan avointen sanaluokkien, kuten substantiivien ja verbien, välisiä keskinäisiä eroja ei tunneta eikä niistä tietääkseni ole käyty keskustelua. Pääasiassa vain substantiivi- ja adjektiivisanaston uudistumista näkyvät tutkitun (tai niiden uudis-

tumisesta on käytännön kokemusta) ja tulokset on ilmeisesti yleistetty verbeihin. Verbien osalta kirjallisuudesta kuitenkin löytyy joitain hajamainintoja, ja kommentoin niistä muutamia.

Baayen ja Renouf (1996) ovat tutkineet englannin *-ly*, *-ness*, *-ity*, *un-* ja *in-*johtimilla muodostettuja uudissanvoja. He toteavat, että aineistona olevassa Times-korpuksessa näistä produktiivisin eli *-ly* tuottaa uuden johdoksen keskimäärin yhden päivässä, *-ness* ja *un-* hieman vähemmän. Loput eli latinalaisperäiset johtimet eivät ole kovin produktiiveja. Baayen ja Renouf (1996: 85) mainitsevat ohimennen, että verbijohdoksien syntyminen saattaa olla vaikeampaa. He tosin liikkuvat hataralla pohjalla arvellessaan tämän johtuvan siitä, että johdin muuttaa tai modifioi (kanta)verbin argumenttirakennetta. Näin on monesta syystä. Ensinnäkin johtaminen (tai verbimorfologia) on standarditapa muuntaa kantaverbin argumenttirakennetta (ks. esim. Payne, 1997: 196); esimerkiksi transitiivi–intransitiivi-muunnos on kielissä hyvin yleinen. Toiseksi kielitypologiset erot johtamisen produktiivisuudessa ovat valtavat (ks. Itkonen, 2001: 29–31), joten vaikka johdoksia ei syntynyt analyttisessä englannissa, synteettisissä ja polysynteettisissä kielissä tilanne on aivan toinen. Kolmanneksi myös muut kuin verbi-johtimet muuntavat tai modifioivat kantasansan rakennetta ja merkitystä.

Bauer (2001: 35) toteaa, että uusien sanojen muodostuminen on niin nopeaa, että jokainen sanakirja on jo ilmestyessään vanhentunut. Vastaava kommentti löytyy sanakirjakontekstista usein (vrt. esim. Uudissanasto 80:n esipuhe; Kolehmainen, 1991; Grönroos, 2004). Bauer ei lainkaan viittaa sanaluokkaeroihin uudismuodosteita rajaavana tekijänä mutta hän arvelee, että vaikka yhdyssanojen muodostus on niin innovatiivista, että sanakirjoista puuttuu niistä aina huo-

mattava osa, johtamalla muodostettuja sanoja ei vastaavasti puuttuisi. Tämäkään väite ei ole kielitypologiselta kannalta pätevä.

Eve ja Herbert Clark (1979) osoittavat, että englannissa verbien johtaminen substantiiveista on produktiivista (esim. *a blanket* – *to blanket*). Eve Clark (1993: 198–199) toteaa, että englannin pääasiallinen keino johtaa uusia verbejä on nolladerivaatio ja että sitä käytetään laajasti; varsinainen johtaminen on englannissa vähäistä. Näyttäisi siis siltä, että myös verbisanaston uudistuminen olisi merkittävää ainakin englannissa. Clarkin (1993: 213–217) mukaan esim. unkarissa ja hepreassa uudisverbin johtaminen substantiivista näyttää rajatummalta kuin germaanisissa tai slaavilaisissa kielissä. Kumpikaan lähde ei kuitenkaan pohdi uudismuodosteiden käyttöä ja vakiintumista, vaikka Clark (1993: 218) kyllä toteaa, että lasten uudismuodosteet perustuvat yksittäiskielissä yksinkertaisiin muotoihin, läpinäkyviin merkityksiin ja mallin produktiivisuuteen.

Näyttää siis siltä, että mikäli kirjallisuudessa kommentoidaan verbisanaston uudistumista, kommentit keskittyvät yksittäiskielen tapaan muodostaa uudisverbejä. Uudistumisen mahdollisia sanaluokkaeroja ei pohdita. Kielitypologiselta kannalta katsoen kommentit tuntuvat vinoutuvan englannin ominaisuuksia painottaviksi.

Sanan uutuus

Perinteellisen kielitieteen kanta sanaston uudistumiseen on esitetty Paulin (1880/1975) luvussa IX *Urschöpfung*; uudemman version tarjoaa Erkki Itkonen (1966: 392–395). Perinteellisen kannan mukaan on selvää, että jokaisessa kielessä on ollut sellainen varhainen vaihe, jossa sanajuuret on 'luotu'. Sekä elottomaan että elolliseen historiaan sovelletavan ns. uniformitarisen periaatteen mukaisesti menneisyys ei voi olla laadullisesti toi-

senlaista kuin nykyisyys, ja siksi myös nykyään täytyy olla mahdollista luoda uusia sanoja 'tyhjistä'. Käytännössä tällaista *Urschöpfungia* esiintyy kuitenkin – rajoitetusti – vain deskriptiivisanastossa ja slangissa. Sanaston muissa osissa siihen ei ole tarvetta, koska uudistuminen (joka sinänsä on tietysti välttämätön) voidaan hoitaa muita keinoja käyttäen. Näitä ovat toisaalta merkityksen muutos ja toisaalta uuden johdoksen, yhdyssanan tai lainasanan muodostaminen. Sivuutan merkityksen muutoksen tässä yhteydessä, koska aiheena on sanaston uudistuminen eli uusin lekseemien synty.

Suomen tapaisessa kielessä uudet kannat ovat tyypillisimmin lainoja. Derivaatio on suomessa pääasiallinen tapa muodostaa uusia verbejä ja vaikka nykyisin syntyy myös yhdysverbejä, frekvenssien tasolla tämä osoitautuu aika marginaaliseksi ilmiöksi. Omasa aineistossani yhdysverbien osuus kaikista finiittisistä verbiesiintymistä onkin reippaasti alle puoli prosenttia. – Edelleen on mahdollista tehdä ero niiden uusien sanojen välillä, jotka nimeävät uusia käsitteitä, ja niiden, jotka vain muokkaavat vanhojen merkitystä tai muuttavat sanaluokkaa.

Uutuus ei siis rajoitu kannan uutuuteen. Suomen tapaisessa kielessä uudissanat ovat tyypillisimmin sanarakenteeltaan kompleksisia johdoksia. Verbiaineistoni HL-lekseemien keskimääräinen merkkipituus onkin odotuksen mukaan selvästi suurempi (eli kymmenen merkkiä) kuin yleisempien verbien vastaava merkkipituus (seitsemän merkkiä). Ero vastaa keskimäärin yhtä verbijohdinta (esim. *-ttA-*, *-UtU-*). Kompleksisten sanojen uutuutta on hankala arvioida, koska niiden kanta on todennäköisesti jo tuttu. Lisäksi tiedetään, että sanakirjoihin ei koskaan oteta kaikkia sanoja. Valinta voi perustua yleisyyteen tai olla systemaattista esimerkiksi siten, että vältetään transparentteja johdoksia, joissa johdin modifioi ennustet-

tavalla tavalla kannan merkitystä. Esimerkkinä voi mainita suomen frekventatiivijohdokset (*sitereailla*, *tööttäillä*) tai *minen*-johdokset (*sitereaileminen*; *tööttäileminen*), joita sanakirja-aineistoissa on vain sattumanvaraisesti verrattuna johtimien produktiivisuuteen. Vastaavasti sanakirjoista usein puuttuvat transparentit teettokausatiivijohdokset (esim. HS-aineiston *kohotuttaa* ja *päällystytää* puuttuvat sekä NS:stä että CD-Perussanakirjasta = CD-PS).

Toisaalta johdon transparenttius ja johdoksen merkityksen kompositionaalisuus eivät suoraan ole olleet hylkäämiskriteerejä. Esim. kompleksisia *u*-johdoksia on sanakirjoissa paljon. Ne ovat produktiivisuuden kannalta nykyisinkin mielenkiintoinen ryhmä, koska niiden ala on laajenemassa uusiin verbiluokkiin. Muodostustavaltaan ne silti ovat läpinäkyviä ja merkitykseltään kompositionaalisia (vrt. NS *puolustaa* > HS *puolustua*, NS *koristaa* > HS *koristua*, NS *kuvittaa* > HS *kuvittua*).

Uudissanojen yhteydessä tehdään yleensä jaottelu mahdollisiin sanoihin (*potential*, *probable*) ja todella esiintyviin aktuaalisiin sanoihin (*actual / occurring / existing*). Esimerkiksi Bauer (2001: 36) katsoo, että sana on olemassa (*existing*) siitä hetkestä lähtien, kun se on ensi kerran muodostettu (*coined*). Se vakiintuu (*established*) vasta, kun se on tuttu tarpeeksi suurelle osalle kielenkäyttäjää. – Bauerin jaottelu olemassa oleviin ja vakiintuneisiin on välttämätön mutta ei täysin riittävä. Synteettisen kielen massakorpusta analysoidessaan törmää näet muodosteihin, jotka voivat olla joko tekstissä oikeasti esiintyviä tai sitten jäsentimen tuottamia.

HS-aineistossani verbit *salakuljettua*, *kännyköityä* ja *sunnuntaistua* voi luokitella luokkaan olemassa olevat (löydös on) mutta eivät vakiintuneet (vain yksi esiintymä ja intuitiivisesti arvioiden uudenkuuloinen). Esimerkit *messuilla* (< *messut*) ja *sisäsiivuilla*

(< *sisäsivut*) ovat sen sijaan jäsentimen virheellisesti verbiksi kantasanaistamia substantiiveja. Ongelma on nyt mahdollisen uudisaineksen erottelu todella olemassa oleviin ja vain näennäisesti olemassa oleviin; eronteko ei voi perustua pelkästään intuitioon. Mainittu *sisäsivuilla* on kyllä intuitiivisestikin arvioiden selvä epäšana, mutta sekä aineistossa esiintyvä *kännyköityä* että esiintymätön *messuilla* vaikuttavat mahdollisilta. Tyypillistä lisäksi on, että aineistossa aidosti esiintyvä *salakuljettua* on epäodotuksenmukaisempi (koska antikausatiivijohdoksia syntyy liikuttamisverbeistä harvoin, ks. Pajunen, 2001: 184) kuin esiintymätön *messuilla*. Intuition avulla ei myöskään voi erotella esim. produktiivisilla (tai produktiivistumassa olevilla) johtimilla tuotettuja johdoksia. Jäsennin tuottaa esimerkiksi johtimesta (-inen >) -*ista* paljon suomen kielen kannalta mahdollisia muotoja (*edullistaa, tasaistaa*), samoin eri kausatiivijohdista (*arkiinnuttaa, päätöstää*); nämä muodot joutuu aina erikseen tarkistamaan kontekstin avulla.

Baayenin ja Renoufin (1996: 75, viite 5) mukaan ihmisillä ei olekaan luotettavaa tietoa rakenteeltaan kompleksisen, vähän käytetyn sanan uutuuden asteesta, koska kantasanan tuttuus vaikuttaa häiritsevästi. Siksi he omassa tutkimuksessaan ovat määritelleet uutuuden operationaalisesti. He ovat poimineet englannin *ly*-esiintymät automaattisesti n. 3,5 vuoden Times-aineistosta (80 milj. esiintymää) ja saaneet tulokseksi suuren määrän *ly*-sanoja. Heidän oletuksensa on, että jos aineiston yhdellä lisäkeruukuukaudella vielä löytyy uusia yhden esiintymän *ly*-sanoja, ne mitä todennäköisimmin ovat kielenkäyttäjän kannalta uusia muodosteita. Tällaisia uusia *ly*-sanoja he sitten löytävätkin viimeiseltä keruukuukaudelta yhden päivää kohden.

Baayenin ja Renoufin soveltama kartuntaerottelu ei ole johtimilla tms. rajoittamattomaan aineistoon suoraan sovellettavis-

sa, koska se valikoisi mm. harvakseltaan sanomalehtikielessä käytetyt deskriptiiviverbit uudissanoiksi. Ylipäätään metodi saataisi morfologisesti rikkaan kielen kohdalla johtaa hankalasti arvioitavaan lopputulokseen. Sanakirjoissa on esim. frekventatiivijohdoksia vähän mutta HS2000-korpuksessa verbileksemeistä vajaa 20 % on frekventatiivijohdoksia. Baayenin ja Renoufin kriteerillä esim. *kroolaillla* ja *lysähdellä* päätyisivät uudismuodosteiksi, mitä ne eivät selvästikään ole.

Rakenteeltaan uudenlaisen muodosteen kielenmukaisuutta pystyy arvioimaan intuitiivisesti. Esimerkiksi seuraavat huomaa helposti satunnaismuodosteiksi: *joko ruohonleikkasit nurmikon, mulla äsken vähän päänsärkähti pää* (esimerkit ala-asteikäiseltä). Aineiston jäsennin tuottaa melkoisesti vastaavantapaisia kielenvastaisia uudismuodosteita, ja nämä virheelliset kantasanaistukset eli kielenvastaiset verbit olen poistanut aineistostani suoraan. Muiden osalta verbin ns. 'uutuus' perustuu yksin siihen, että (tarkistettu) verbi esiintyy HS:n aineistossa, vaikka sitä ei mainita ao. muodossa NS:ssa. Tämä on uudismuodosteen pääkriteeri, jonka avulla analysoin uudisverbien määriä ja vertaan niitä koko aineistoon. Lisäksi arvioin uudisaineistoa kahdella tiukemmalla lisäkriteerillä, jotka ovat seuraavat: (välitöntä) verbikantasanaa ei mainita NS:ssa tai sitä ei mainita CD-PS:ssa. Kieliyhteisön kannalta etenkin ensimmäisen kriteerin mukainen uudismuodoste voi olla uusi tai vanha, muilla kriteereillä määritellyt enimmäkseen uusia tai kielen vielä vakiintumattomia.

Tutkimuksen aineisto

Sanakirja-aineistonani ovat NS:n n. 16 000 (päähakusana)verbiartikkelia, jotka on koodattu tietokantamuotoon ja luokiteltu usean ominaisuuden suhteen (Pajunen, 1994–

ja verrattu tulosta aikaisempiin; näin on todennäköistä, että vaikka jotain voi edelleen puuttua, HL-lekseemien kattavuus on hyvä. Saatua aineistoa on verrattu myös SS2001-korpuksen aineistoon ja voi todeta, että ne vastaavat toisiaan hyvin. Työmäärä on ollut niin suuri, että jonkinasteinen virhemarginaali on edelleen mahdollinen, mutta toivottavasti se on siedettävä.

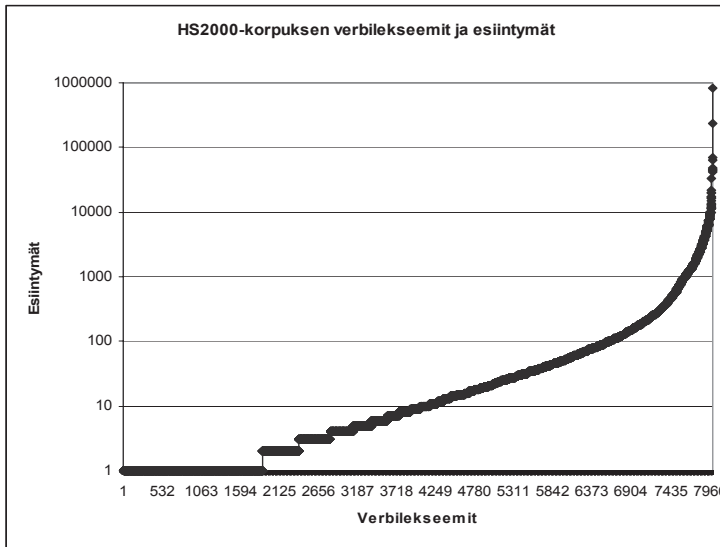
Aineiston uudismuodosteet on siten arvioitu muodollisin kriteerein. Verbi(johdokse)n on katsottu puuttuvan NS:sta, vaikka sen verbikantasana olisi NS:ssa. Näin muodostui tutkimuksen karsimaton uudisaineisto. Kirjainkohtaiset erot, joita aiheuttavat oikeinkirjoitusvirheet, murteellisuudet jne. on tasoitettu (*katella* ja *katsella* lasketaan siten yhdeksi lekseemiksi, samoin *akottua* ja *akoittua*, *kummaksua* ~ *kummeksua*). Kompleksisten *u*-johtimien korrelaatio-suhteiset mutta merkitykseltään toisiaan vastaavat tapaukset (vrt. *hautautua* : *hautaantua*) lasketaan niin ikään yhdeksi lekseemiksi. Näissä hakusanaksi on valittu frekventimpi tapaus (esim. *halvaantua*, pro *halvautua* mutta *harhautua* pro *harhaantua*), jonka esiintymisiin on laskettu myös korrelaatiin esiintymät. On tyyppillistä, että NS tuntee molemmat *u*-johdokset. Näin yhdistettyjä *u*-lekseemejä on yhteensä 78 ja niiden esiintymiä on yhteensä vajaa 10 000. Muut modifioivat ja muuntavat johdokset, kuten frekventatiivijohdos *huimailla* (< *huimata*) ja *ne*-johdos (*laihene-*, vrt. *laihtua*), sisältyvät karsimattomaan uudisaineistoon. Suurella osalla HS2000-korpuksen verbilekseemistä on joko nomini- tai verbikanta NS:ssä tai ainakin CD-PS:ssä. Jos verbilekseemillä on sanakirja-aineistossa ainoastaan nominikantasana, katson sen uudismuodosteeksi myös karsitussa aineistossa. Verbikannan osalta tilanne on karsitussa aineistossa toinen.

VERBIT HS2000-KORPUKSESSA

HS2000-korpuksessa on n. 3,1 milj. finiittisenä tai finiittisen muodon osana käytettyä verbitokenia eli -esiintymää. Eri verbityyppisiä, lekseemejä aineistossa on n. 8 000. NS:n verbilekseemistä yli 40 % (eli n. 6 500) on käytössä HS2000-korpuksessa; korpuksen lekseemistä runsas 80 % on peräisin NS:stä. Verbilekseemistä, jotka eivät esiinny NS:ssa mutta esiintyvät korpuksessa, vain muutama nousee yleisyydeltään yli 150 mutta yksikään ei yli 300 esiintymän. Jos uudismuodosteella on yli kymmenen esiintymää, se on käytössä myös Savon Sanomien korpuksessa; se on ts. vakiintunut (*established*). Yli puolet uudismuodosteista on HL-lekseemejä. Ne kuuluvat joko luokkaan jo olemassa olevat mutta vakiintumattomat (*existing*) tai sitten ne usein ovat tutun kannan merkitystä modifioivilla johtimilla muodostettuja johdoksia. – Jo pelkästään näistä luvuista voi päätellä, että HS:ssa käytetty verbisanasto on pääosin selvästi vanhaa. Tarkastelen kuvioiden 1–5 avulla aineistoni lekseemien ja esiintymien suhdetta eri näkökulmista ensin koko aineistossa ja sitten sen eri tavoin koostetuissa ja karsituissa osissa.

Verbilekseemien ja -esiintymien suhde koko aineistossa

Aineisto käsittää HS2000-korpuksen kaikki verbilekseemit, niin sanakirjalle ja korpukselle yhteiset kuin vain korpuksesta löytyvät. Verbilekseemejä on n. 8000 ja niillä esiintymiä n. 3.1 milj. Lekseemien ja niiden esiintymien suhde on hyvin vino, joten se on havainnollisuuden vuoksi kuvattu logaritmiselle asteikolle (ks. kuviot 1–3). Kuvion 1 tiedot voi tiivistää seuraavasti: Aineiston verbilekseemistä vajaalla puolella on alle 10 esiintymää, n. 6 000:lla (n. 75 %) on alle sata esiintymää, n. 7 000:lla on alle 1 000



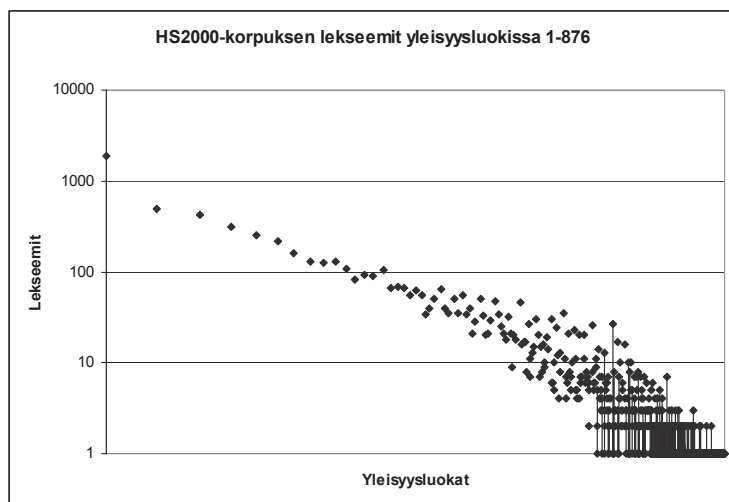
Kuvio 1. Verbilekseemejä 8011 ja esiintymiä n. 3.1 milj.

esiintymää (n. 90 %) ja vajaalla 8 000:lla eli jo lähes kaikilla on alle 10 000 esiintymää. Vain n. 30 lekseemillä on yli 10 000 esiintymää, näistä yleisimmät ovat kieltoverbi *ei* ja *olla*-verbi. HL-lekseemien osuus koko aineistossa on yli 20 %. Osaksi ne ilmaisevat sanaston uudistumista ja myös sanarakenteen mahdollisia kehitystendenssejä, sanakirjan ja korpuksen yhteisessä aineistossa myös sanaston rikkautta (vrt. Särkkä, 1987).

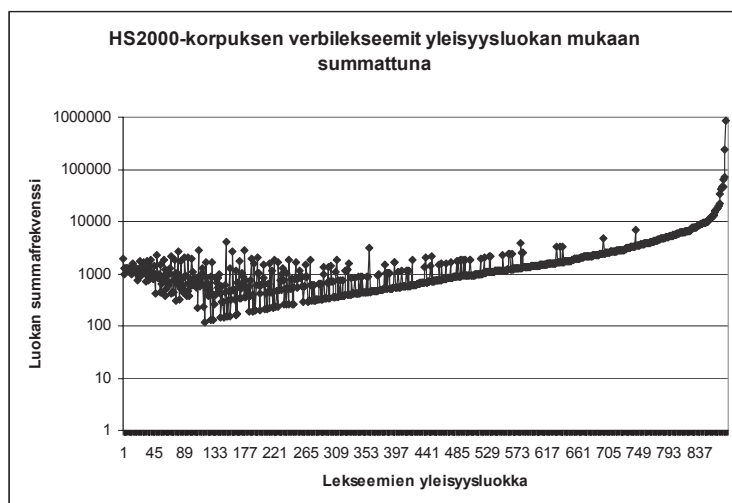
Lekseemien esiintymiä voidaan tarkastella myös yleisyysluokittain (kuvio 2). Aineiston n. 8000 verbilekseemiä jakaantuu n. 900 yleisyysluokkaan (= ne järjestysluvut, joihin esiintymiä sijoittuu). Luokissa lekseemien määrä vaihtelee asteikossa vasemmalta oikealle laskien n. kahdesta tuhannesta yhteen. Jakauma paljastaa ehkä kuviota 1 paremmin, miten paljon on harvoin esiintyviä lekseemejä ja miten vähän taajaan. Kuvion 2 perusteella yleisyysluokkien eroja voi ilmaista seuraavasti: yhden esiintymän luokka (HL-lekseemit) erottuu odotuksen mukaisesti suurena, luokissa 2–10 esiintymää lekseemejä on luokkaa kohden sadasta viiteen sataan, luokissa 11–100 esiintymää kymmenestä kahteenkymmeneen, luokissa 101–300 esiinty-

mää muutama ja kun lekseemin yleisyys nousee yli kolmen sadan, samalla yleisyystasolla on tyypillisesti enää vain yksi lekseemi.

Lekseemien ja esiintymien suhdetta voi edelleen tarkastella laskemalla kunkin aineistossa esiintyvän yleisyysluokan (1–876) lekseemien esiintymät yhteen (esim. luokassa 101 esiintymää on lekseemejä 11, joten luokan jäseniä on yhteensä 1111) (ks. kuvio 3). Tämä laskentatapa osoittaa, miten verbilekseemeillä ilmaistavat merkitykset keskittyvät asteikon yleisyysluokkiin 1–100, joissa summafrequenssi on keskimäärin n. tuhannessa; tosin sisäinen vaihtelu on melko suurta. Huomattakoon, että harvaan esiintyvät lekseemit ilmaisevat tyypillisesti spesifejä ja perustason leksikaalisia merkityksiä; yleiset lekseemit ovat paitsi monimerkityksisiä myös monifunktioisia. Ne ilmaisevat mm. tekstilajiin liittyviä merkityksiä (esim. *sanoa*, *kertoa*), propositionaalisia asenteita (*voida*, *pysytää*) ja syntaktisia funktioita (*olla*) sekä luonnollisesti suhteellisen geneerisiä merkityksiä (*mennä*, *saada*, *tehdä*, *tulla*). Edellinen ryhmä kattaa lekseemeistä valtaosan, jälkimmäinen vastaavasti valtaosan aineiston esiintymistä. Merkityksen spesifisyys- ja vastaavasti



Kuvio 2. Verbilekseemejä 8011, yleisyysluokkia (= ♦) 876.



Kuvio 3. Verbilekseemejä 8011, esiintymiä 3.1 milj., esiintymäluokkia 876.

geneerisyyseroa vastaa myös ero lekseemien merkkimäärien eroissa: yleisyysluokissa 1–500 lekseemien keskimääräinen merkkimäärä on 8, yleisyysluokissa 501–850 merkkimäärä on seitsemän; ääripäiden välinen ero on 4–5 merkkiä.

Verbilekseemien ja -esiintymien suhde aineiston eri osissa

Tarkastelen seuraavaksi aineistoa kiinnittämällä ensin huomiota sanakirjalle ja korpukseksi yhteiseen ainekseen ja sitten eri tavoin

kirjaimellista samuutta edellyttävää kriteeriä tiukemmin määriteltyihin uudismuodosteisiin. Tiukemmin määriteltyjen ryhmässä arvioin ensin niitä uudismuodosteita, joiden verbikantasana saattaa löytyä NS:sta, ja sitten niitä, joita ei tunne edes CD-PS.

Yhteinen aines

Nyky-suomen sanakirjalle ja HS2000-korpukseksi yhteinen aines kattaa koko aineistosta n. 80 %, jos yhteisyys määritellään kirjaimellisesti. Luku nousee muutamalla prosenttiyksiköllä, jos yhteiseen ainekseen las-

ketaan mukaan myös tutuista verbeistä johdetut modifioivat verbijohdokset. Esiintymien tasolla tämä yhteinen aines kattaa 99,7 % kaikista. Yhteisten lekseemien ja esiintymien jakauma vastaa tarkasti kuviota 1; mm. HL-lekseemien osuus on edelleen korkea eli n. 15 % kaikista. Yhteisten HL-lekseemeistä n. neljännes on frekventatiivijohdoksia (esim. *nostatella, pamahdella, puukotella*), edelleen neljännes on deskriptiivisiä (*hyreksiä, hyrskytää, hähättää, jolkutella, polottaa*); loput ovat joko tekstilajille epätyypillisiä tai passiivisnastoon kuuluvia yksittäisesiintymiä. Tyypillisesti nämä lekseemit ilmaisevat spesifiä merkitystä; se ilmenee mm. deskriptiivikanasta tai modifioivasta johtimesta. HL-lekseemeissä liikeverbit ja fysiologista tai psykologista tilaa ja toimintaa tai reaktiota sekä puheaktia ilmaisevat vallitsevat; juuri näillä verbiluokilla hierarkkinen rakenne on syvä (ks. Pajunen, 2001: 59–65).

Uudismuodosteen määrittely on suomen tapaisessa, johtamalla sanoja muodostavassa kielessä hankalaa, kuten edellä jo todettiin. Pohdin uudismuodosteen määrittelyä uudelleen, ja esitän sitten uudismuodosteiden ja niiden esiintymien suhdetta tiukemmin rajatussa aineistossa.

Uudismuodosteen läpinäkyvyys ja kompositionaalisuus

Aineistossa on edellä (kuviot 1–3) lekseemeiksi laskettu myös rakenteeltaan läpinäkyvät ja kompositionaalista merkitystä ilmaisevat johdokset sekä sanakirjoissa mainituista, tutuista osista koostuvat yhdysverbit. Pohdin näiden osalta uudismuodosteiden rajattavuutta uudelleen ja esitän tarkastelun tulokset uudisverbien lekseemi–esiintymä-suhdetta kuvaavissa tarkistetuissa kuvioissa 4 ja 5.

Rajaamattomaan uudisaineistoon lukeutui paljon mm. frekventatiivi- ja momentaanijohdoksia; kokonaislekseemimäärästä nii-

tä on 20 %. Näille kannan merkitystä modifioiville lekseemeille on tyypillistä, että niiden esiintymät painottuvat muutamisiin (1–10), joten esiintymistä nämä johdokset kattavat vain kolmisen prosenttia. Vain johdoksella *vaihdella* on yli tuhat esiintymää, muut yli tuhannen yltävät frekventatiivit ovat merkitykseltään ei-kompositionaalisia. Aika monella frekventiivi- ja momentaanijohdoksella on vastine NS:ssä, usein tosin vain kantanasana. Sanakirjoissa esiintymättömiä, merkitykseltään modifioivia johdoksia löytyy HS2000-korpuksesta n. 50, näistä suurin osa on puhekielisiä (esim. *jutskailla, sippaila, tsättäillä, meilaila*) ja myös niiden kantanasana puuttuu NS:sta. Frekventatiivijohtaminen on suomessa hyvin produktiivista, joten periaatteessa frekventatiivijohdos voidaan muodostaa lähes mistä tahansa kannasta.

Karsittuun uudisaineistoon en enää sisällyttänyt niitä frekventatiiveja, joiden kantanasana esiintyy NS:ssä ja jotka ovat merkitykseltään kompositionaalisia. Tämä on perusteltua kahdesta syystä: frekventatiiveja muodostetaan aika vapaasti ja toisaalta niitä otetaan sanakirjoihin melko sattumanvaraisesti. Siksi edellä sovellettu uudismuodosteen peruskriteeri – esiinny NS:ssä kirjaimellisesti samassa muodossa – erottelee erityisen huonosti.

Frekventatiivijohdokset ovat suurin kompositionaalisten johdosten ryhmä mutta muitakin löytyy, esim. reduplikatiiviset kausatiivijohdokset (ns. teettokausatiivit, *teetättää*). Kompleksisten *u*-johdosten muodostaminen on niin ikään produktiivista, lisäksi niiden joukossa – mutta myös muista – löytyy korrelaatiosuhteisia johdoksia (vrt. *kelottua* – *keloontua* – *keloutua*). Osa johdoksista on komplementaarissa distribuutiossa esim. siten, että vokaalivartalomuodossa käytetään yhtä, konsonanttivartalon yhteydessä toista johdinta (vrt. *jokin suurenee* : *on suurentunut*, ei *suurennut*). Kannan argumenttirakennetta

muuntavilla johtimilla muodostetut lekseemit sisällytän kuitenkin edelleen rajattuunkin uudisaineistoon mm. siksi, että produktiivisuus voi rajautua tiettyihin tyyppeihin tai olla yleensäkin näennäistä. Ainoastaan muuttaman korrelaatiosuhtaisen *u*-johdoksen esiintymät on aineistossa sisällytetty sarjan frekventempään, kuten edellä todettiin.

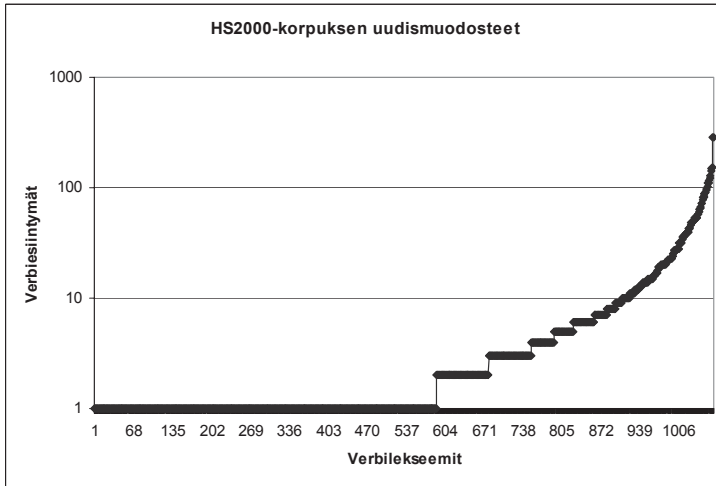
Rajaan nyt aineistosta uusia verbileksemejä arvioidessani kompositionaaliset frekventatiivit muutaman edellä ilmaistun tapausten lisäksi. Mm. erilaiset (muuntavat) johdosketjut ovat mukana, koska johtamissuunta voi niissä vaihdella. Ilman erillistutkimusta ei voikaan tietää, onko niiden kohdalla kyse sattumasta vai esim. takaperoisjohdosta, vrt. *annostaa* (NS, HS2000) > *annostua* (HS2000), mutta *autoistua* (NS, HS2000) < *autoistaa* (HS2000). Suomessa sananmuodostus tapahtuu pääasiassa joko johtamalla tai yhdyssanaistamalla, joten periaatteessa kaan kovin laajaan aineiston pelkistämiseen ei ole syytä mennä. Uudisaineistossa ovat myös nominikantaiset verbijohdokset.

Oman ongelmansa muodostavat yhdysverbit. Ne eivät ole vanhastaan suomessa produktiiveja, mutta jo NS tuntee niistä monia. Yhdysverbin perusosa (*ilotulittaa*, *arvokeskustella*) on usein vanha ja yleinen verbilekseemi ja samoin määriteosa on tuttu, joskaan ei verbaalinen. Toisaalta muutama yhdyssanan määriteosa (*jälleenrakentaa*, *esikäsittelä*, *tasalaatuistaa*, *ylireagoida*) tuottaa paljon yhdysverbejä. Kantasana on tyypillisimmin yhdysadjektiivi tai -substantiivi (*epämiellyttää*, *puhelinhaastatella*), inkorporoivat yhdysverbit ovat harvinaisia (*köyhyysspurkautua*). Siten aineistoa voisi pelkistää sekä yhdysverbin perusosan että määriteosan mukaan tai kannan tuttuuden mukaan. Osien tai kannan tuttuudesta huolimatta kaikki finiittisessä funktiossa käytetyt yhdysverbit kuuluvat karsittuunkin uudisaineistoon. Uudismuodosteiksi on laskettu

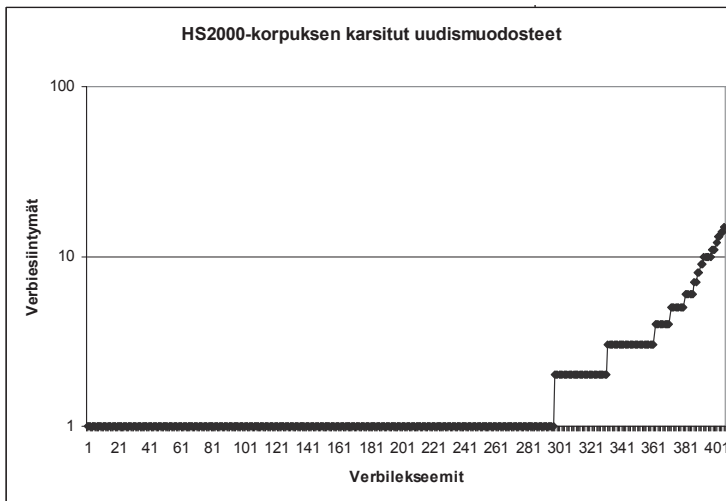
kaikki ne, joille ei sanakirjassa ole annettu omaa sana-artikkelia tai joita ei esimerkeissäkään mainita finiittisessä muodossa. Aineiston perusrajauksen vuoksi yhdysverbeihin ei ole laskettu niitä, joita on käytetty vain partitiippiattributteina.

Mainittujen rajoitusten tuloksena on, että uudismuodosteihin kuuluu nyt n. tuhat verbilekseemiä (kuvio 4). Näistä yli puolet on HL-lekseemejä (esim. *töllätä*, *jumia*, *shopata*, *tehoalvoa*, *brändäytyä*, *selättyä*, *epämiellyttää*, *uv-suojata*). Kolmasosa ryhmän lekseemeistä esiintyy myös SS2001-korpuksessa, yleisimmin ne, jotka eivät ole HL-lekseemejä. Yhteensä vain runsaat kymmenen karsitun aineiston uudismuodostetta nousee yleisyydeltään yli sadan; *kyseenalaistaa* esiintyy HS2000-korpuksessa vajaat kolmesataa kertaa. CD-PS tuntee korpuksen uudismuodosteista runsaan kolmanneksen. Yhdyslekseemejä uudismuodosteista on kolmannes, puhekielisiä n. neljännes. Johdoksista kausatiivi- ja passiivijohdokset vallitsevat.

Lekseemitasolla kuvion 4 uudismuodosteista puolet on teko- tai muutosverbejä, samoin esiintymistä. Sanakirjalle ja korpukselle yhteisessä aineistossa muutosverbejä on lekseemeistä vain viidennes ja esiintymistä neljännes. Ero on siis huomattava: uudismuodosteista suurin osa leksikaloi juuri tekoa tai muutosta. Tämä johtuu ensisijaisesti tarpeesta nimetä uusia valmistus- tai käsittelyprosesseja (esim. *painepestä*, *rakenteistaa*, *ruiskuvärjätä*, *viherrakentaa*). On oletettavaa, että verbisanaston tässä osassa lekseemit edelleen uudistuvat melko nopeasti. Vastaavasti liikeverbien luokassa uusien HL-verbilekseemien osuutta nostaa uusien liikuntalajien nimeäminen (*purjelautaila*, *riippuliitää*, *sauvailla*, *skeittailla*, *sormikoukkuilla*); uusia tapoja liikkua ei sen sijaan ole ollut tarvetta nimetä. Uusien käsitteiden nimeämisen tarpeeseen juuri näissä aihepiireissä on kiinnitetty toistuvasti huomiota mm. Kielikellos-



Kuvio 4. Lekseemejä 1066, esiintymiä 7872.



Kuvio 5. Verbilekseemejä 406, esiintymiä 788.

sa (ks. Grönroos, 1994, 2004; Jumppainen, 2003; Kolehmainen, 1991; Nissinen, 1988). Yhdyslekseemit ilmaisevat usein muutosta (*perhekeskeistää, syväjäädyttää, tasalautuistaa*), puhekieliset lekseemit tyypillisesti mentaalitiloja ja -tapahtumia ja puheakteja (*friikahtaa, pölöttää, päbkiä, vouhkata*).

Uudismuodosteella ei tuttua kantasanaa
Uudismuodosteita voidaan määrittellä myös niin tiukasti, että mukaan hyväksytään vain ne, joita ei mainita NS:ssä eikä CD-PS:ssä. Ryhmä on jo hyvin pieni (ks. kuvio 5).

Noin kolmasosa sanakirjoista puuttuvista lekseemeistä kuuluu slangisanoihin, puhekielisiin tai äänteellisesti suomen kieleen mukautumattomiin lainasanoihin (esim. *bonjata, klangata, klenuta, klobata, kälättää, natsata, skrabata, tsibata, tsiidata, tsimmata*). HL-lekseemejä ryhmästä on yli 70 % ja esiintymistä vajaa 40 %. Yleisin ryhmän verbeistä eli *päivittyä* ylittää vain 15 esiintymään, *digitoida, ulosliputtaa, lintsata* omaavat kukin 10–14 esiintymää. Ryhmän koostumuksesta voi päätellä, että tästä aineistoista monet kuuluvat alakulttuurikieleen; HS2000-

korpuksessakin niiden käyttö usein rajoittuu nuorten palstoihin tms.

Runsas kolmannes ryhmän lekseemeistä on yhdysverbejä. Ne ovat tyypillisesti yleiskielisiä (*eu-sopeuttaa, jälki-pastöroida, pakosiirtää, seinäkiipeillä*). Yhdysverbien määrä on suomessa tasaisesti kasvanut ja todennäköisesti kasvaa edelleen; yleisyys ei yhdysverbeillä nouse korkeaksi. Koko aineistossa yleisimmällä yhdysverbillä eli *allekirjoittaa*-verbillä on esiintymiä alle tuhat, yleisyysluokassa 100–900 on kolmetoista yhdysverbiä, yleisyysluokassa 10–99 on n. 50 yhdysverbiä ja peräti 58 % kuuluu HL-lekseemeihin.

LOPUKSI

Vanhan sanakirja-aineiston ja suuren korpuksen vertailu tuottaa tulokseksi, että verbisanaston uudistuminen 1900-luvulla on sanomalehtikielen aineiston perusteella ollut suomen kielessä vähäistä. Karsimattomassa aineistossa uudismuodosteita näyttää syntyneen muutamia kymmeniä mutta karsitussa aineistossa vain muutamia vuodessa. Vertailun vuoksi mainittakoon, että Uudissanasto 80, joka täydentää Nykysuomen sanakirjaa, sisältää n. 6 000 sana-artikkelia. Kielitoimiston sanakirja (2005) sisältää hakusanoja 4 000 enemmän kuin CD-PS 1997 ja 6 000 enemmän kuin Suomen kielen perussanakirja (ks. Grönroos, 2004). Jos lukuja suhteuttaa toisiinsa, voi olettaa, että uudismuodosteista n. 95 % olisi nomineja ja korkeintaan viitisen prosenttia verbejä.

Verbiaineistoni osoittaa ts. selvästi, että verbisanasto suomessa uudistuu hitaasti toisin kuin substantiivisanasto. Ts. avoimen sanaluokan uudistuminen ei verbisanaluokassa ole niin nopeata kuin kirjallisuudessa esitetään. Tulos on yllättävä ja alkuoletuksieni vastainen. Verbisanastoon vaikuttaa myös merkityksen muutos, mutta sitä olen käsitellyt vain enemmän tai vähemmän satunnai-

sesti. Tähän on kaksi syytä. Ensiksikin tärkeimpänä tutkimuskohteenani on ollut uusien sanojen synty; jos sana X saa uuden aineistosta poikkeavan merkityksen, niin normaalin ajattelutavan mukaan tämä ei merkitse, että olisi syntynyt uusi sana Y. (Aito homonymia on asia erikseen.) Toiseksi merkityksen muutoksen lainalaisuuksia on käsitelty hyvinkin laajasti viimeaikaisessa metafora- ja kieliopillistumistutkimuksessa, eikä minulla ole tässä yhteydessä lisättävää tähän keskusteluun. Jäljelle jäävissä sanaston uudistumiskeinoissa muutos käy selvästi ilmi sanan **muodosta**. Koska deskriptiivisanaston osuus on sanomalehtikielessä melko vähäinen, tutkimassani aineistossa ei ole juurikaan esimerkkejä *Urschöpfungista*. Sanaston uudistumista edustavat siis seuraavat tapaukset: johdos, yhdyssana ja lainasana.

Alussa mainitsin, että suomen tuloksista voi olla jotain hyötyä kielitypologisten verbisanastoa koskevien erojen arvioinnissa, erityisesti sen ymmärtämisessä, miten jossain kielessä verbit voivat muodostaa suljetun sanaluokan. Esimerkiksi kalamissa on Pawleyn (1987) mukaan yhteensä 90 verbilekseemiä, joista vain 30 on aktiivikäytössä; tämä tuntuu aluksi mahdottomalta suomen tapaista kieltä puhuvasta.

Korpusaineistosta voi nähdä, että suurin osa käytetyistä verbilekseemeistä ilmaisee melko spesifiä merkitystä: runsain variaatio oli yleisyysluokassa 2–300, johon kuuluu lähes 80 % kaikista aineiston verbilekseemeistä. On helppo kuvitella, että on kieliä, joissa teko, tapahtuma, liike jne. koodataan (yleismerkityksiseen) verbiin ja sitä modifioiva, spesifinen merkitys koodataan muilla keinoin (seriaalirakenteella, adverbiin, klassifikaattoriin jne.). Itse asiassa suomen murteista on ollut näistä keinoista käytössä seriaalirakenne (esim. *mennä lönkyttää*) ja määrän ja tavan adverbit (sekä tietysti johtimet) kuuluvat nykysuomeenkin. Suomen ja kala-

min ero käy ymmärrettäväksi.

Vastaavasti tuntuu mahdollomalta, että olisi kieliä, joissa vastaavalla tavalla pärjätisiin pienellä substantiivisanastolla. Yleisesti ottaen onkin totta, että juuri verbisanastot voivat olla suljettuja; tällainen tilanne vallitsee yleisesti papualais- ja australialaiskielissä. Se pakollinen vastaesimerkkikin tunnetaan: Mithunin (2000: 412) mukaan irokeesikielissä on substantiiveja suhteellinen vähän. Miksi näin voi olla, jää ilman vastausta tässä yhteydessä.

LÄHTEET

- Baayen, R.H. & Renout, A. (1996). Chronicling the Times: Productive lexical innovations in an English newspaper. *Language*, **72**, 69–96.
- Bauer, L. (2001). *Morphological productivity*. Cambridge: Cambridge University Press.
- CD-PS= Cd-Perussanakirja 1997. Helsinki Edita.
- Clark, E.V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Clark, E.V. & Clark, H. (1979). When nouns surface as verbs. *Language*, **55**, 767–811.
- Everett, D. P. <http://www.linguistlist.org>, viitattu 20.2.2005.
- FDG = Functional Dependence Grammar –jäsenin. Connexor.
- Grönros, E.-R. (1994). Yhdysverbit. *Kielikello*, 3/1994.
- Grönros, E.-R. (2004). Kielitoimiston sanakirja. *Kielikello*, 4/2004.
- Hudson, G. (2000). *Essential introductory linguistics*. Malden, MA: Blackwell.
- Itkonen, E. (1966). *Kieli ja sen tutkimus*. Helsinki: WSOY.
- Itkonen, E. (2001). *Maailman kielten erilaisuus ja samuus*. 2. uudistettu ja laajennettu painos. Yleisen kielitieteen julkaisuja 4. Turku.
- Jumppainen, H. (2003). Buuttaista ja renderöidä – atk-slangin verbejä. *Kielikello*, 2/2003.
- Kielitoimiston sanakirja (2005). Kotimaisten kielten tutkimuskeskus. Helsinki.
- Kolehmainen, T. (1991). Muuttuvan maailman muuttuvat sanat. *Kielikello*, 2/1991.
- Laine, M. & Virtanen, P. (1999). WordMill. Lexical Search Program 2.2. University of Turku, Center for Cognitive Neuroscience.
- Mithun, M. (2000). Noun and verb in Iroquoian languages. Teoksessa P. Vogel & B. Comrie (toim.), *Approaches to the typology of word-classes*. Berlin: Mouton de Gruyter.
- Nissinen, I. (1988). Mitä uutta kuuluu? *Kielikello*, 4/1988.
- NS = *Nykysuomen sanakirja 1951–1961*. WSOY.
- Nykysuomen sanakirja. <http://www.kotus.fi/julkaisut/sanakirjat/nssk>, viitattu 2.12.2003.
- Pajunen, A. (1994–2001). Verbit-tietokanta. Nykysuomen sanakirjan n. 16000 verbilekseemiä, luokiteltu n. 20 semanttis-syntaktisen ominaisuuden suhteen (Access-muotoinen).
- Pajunen, A. (2001). *Argumenttirakenne. Asiaintilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Suomi 187. Helsinki: SKS.
- Pajunen, A. (2002). SS2001-korpus. Savon Sanomat 1.1.2001–31.12.2001, 11,2 milj. sanaesiintymää. Analysoitu Connexorin FDG-jäsentimellä, Context Mill -muotoinen. Turun yliopisto: Yleinen kielitiede.
- Pajunen, A. (2003). HS2000-korpus. Helsingin Sanomat 1.1.2000–31.12.2001, 31,2 miljoonaa sanaesiintymää. Analysoitu Connexorin FDG-jäsentimellä, Context Mill -muotoinen. Tampereen yliopisto: Suomen kieli.
- Pajunen, A. & Palomäki, U. (1982). Tilastollisia tietoja Lauseopin arkiston aineksestä. [Julkaisemattomia frekvenssitietoja, tekijän hallussa].
- Paul, H. (1880 [1975]). *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemeyer Verlag.
- Pawley, A. (1987). Encoding events in Kalam and English. Teoksessa R.S. Tomlin (toim.), *Coherence and grounding in discourse*. Amsterdam: John Benjamins.
- Payne, T. E. (1997). *Describing morpho-syntax*. Cambridge: Cambridge University Press.
- Suomen kielen perussanakirja 1990–1994*. Kotimaisten kielten tutkimuskeskus. Helsinki: Edita.
- Schachter, P. (1985). Parts-of-speech systems. Teoksessa T. Shopen (toim.), *Language typology and syntactic description*. Volume I Clause structure, (s. 3–61). Cambridge: Cambridge University Press.
- Särkkä, T. (1987). Sanaston rikkaudesta ja sen mittaamisesta. *Virittäjä*, **91**: 129–137.
- Uudissanasto 80. *Nykysuomen sanakirja 5*. Porvoo: WSOY.
- Virtanen, P. (2003). *ContextMill. The easy-to-use-concordancer*. Version 1.80. 2000–2003 Aurit Software.

THE RENEWAL OF THE VERB LEXICON

Anneli Pajunen, School of Modern Languages and Translation Studies, University of Tampere

This article discusses the renewal of the verb lexicon in the Finnish language. Verbs belong to the open class words and are supposed to renew fast; this is, however, not true of the Finnish verb lexicon. The results are based on corpus studies on 31 million word tokens of newspaper text (Helsingin Sanomat volumes 2000 and 2001) on the one hand and on the dictionaries of modern standard Finnish (Nykysuomen sanakirja 1951–1961; Uudissanasto 80; Perussanakirja 1997) on the other. 80 per cent of the verb lexemes of the corpus and the oldest lexicon are the same, they cover over 99,5 % of all tokens in the corpus. New coinages are a) transparent modifying derivatives or b) hapax legomenon lexemes, existing but mostly not yet established lexemes. The new coinages often come from colloquial Finnish or then they name new industrial processes etc. The structure of these coinages is complex, they consist of both derivatives and compound words; words are longer than those in the basic vocabulary.

Key words: verb lexicon, lexical neologism, lexical frequency, hapax legomenon lexeme.