

Loetelukatse ja kognitiivse esiletuleku indeks¹

Urmas Sutrop

Eesti Keele Instituut

Artiklis analüüsitakse loetelukatset ja selle kaht parameetrit: sõna sagedust ja keskmist astakut nimekirjas. Pakutakse ka välja uus ja lihtne kognitiivse esiletuleku indeks $S=F/(N \cdot mP)$, mis kasutab mõlemat nimetatud parameetrit, ning selgitatakse selle arvutamise protseduuri. Indeks on normeeritud varieeruma ühest nullini. Kõige paremini tulevad esile iga valdkonna (domeeni) põhisõnad. Ideaalselt esiletuleva sõna esiletuleku indeks on 1; sõnal, mida üldse ei mainita 0. Kognitiivse esiletuleku indeks võimaldab võrrelda eri uuringute tulemusi, kuna indeks ei sõltu loetelu pikkusest. Artiklis võrreldakse kognitiivse esiletuleku indeksit ka mõne varasema indeksiga.

Käesolevas artiklis tutvustan kõigepealt uut kognitiivse esiletuleku indeksit, milles on ühendatud kaks parameetrit. Edasi analüüsin varasemaid samalaadseid koefitsente, autoriks Smith (1993; Smith jt. 1995; Smith ja Borgatti 1997) ja ma ise (Sutrop 1998, 2000), ning võrdlen neid oma uue indeksiga. Lühidalt peatun loetelukatsel ja selle kahel parameetril, milleks on sõna sagedus ja sõna keskmine astak (järjenumbrer analüüsitavates loeteludes). Lõpuks esitan lühidalt reeglid kognitiivse esiletuleku indeksi arvutamiseks ning paar järeldust.

¹ Artikkel "List Task and a Cognitive Saliency Index" on tõlgitud ajakirjast *Field Methods*, kd 13, nr 3, august 2001, lk 263–276. Täna tõlkijat Sirje Ainsaart. Artikli alusel on Antwerpenis (Belgia) 12.–14. aprillil 2000 toimunud Kognitiivse tüpologia konverentsil peetud ettekanne. Avaldan tänu prof Frans Plankile Konstanzi Ülikoolist (Saksamaa), prof J. Jerome Smithile Lõuna-Florida Ülikoolist (USA) ja prof Jüri Allikule Tartu Ülikoolist abi eest. Täna ka kolme anonüümset retsensenti tulusate märkuste eest.

1. Loetelukatse²

Loetelu all mõistan ma teatud kirjalikke või suulisi küsitlusi, mis on kasutusel antropoloogias, lingvistikas, psühholoogias jm ühiskonnateadustes. Katse tüüpkuju on: “Loetlege, palun, kõik X-id, mida teate” Kirjalik küsitlus algab sõnadega: “Pange, palun, kirja kõik X-id, mida teate” Viimatimainitud formaat sobib muidugi ainult kirjaoskajatele. Suulisel katsel (katsealuste kirjaoskus pole oluline) tuleb öelda “Nimetage, palun, kõik X-id, mida teate”

Võib näiteks küsida: “Nimetage, palun, kõik värvid, mida teate”, “Pange, palun, kirja kõik loomad, mida teate”, “Palun, loetlege kõike, mida võib tunda ninaga” Sääraseid kirjeldavaid fraase nagu viimatimainitu (“kõik, mida võib tunda ninaga”) läheb vaja väikeste domeenide korral või kui klassisõna ise kuulub samasse kultuuri- või kognitiivsesse valdkonda, mida ta tähistab.

Võrreldes näiteks värvi- ja lõhnasõnu näeme, et sõna “värv” on paljudes (kuigi mitte kõigis) keeltes klassisõna, mis tähistab konkreetset domeeni, ise sellesse kuulumata. Sõna “lõhn” seevastu on paljudes keeltes sama domeeni osa, mida ta tähistab.

Väljakujunenud domeenidele polegi nimi kohustuslik. On ju näiteks hästi teada, et paljudes turgi keeltes on hobuste nimetamiseks äärmiselt keerukas süsteem, mis arvestab hobuste sugu, vanust, värvust jne, üldsõna hobuse kui niisuguse jaoks aga puudub. Seades küsitlusele – olgu suulisele või kirjalikule – ajapiirangu, näiteks kaks või viis minutit, saame piiratud loetelukatse. Kui ajapiirang puudub (näiteks kui küsimus kõlab: “Palun, nimetage kõik X-d, mida teate”, ning kui katsealune loetlemise lõpetab, küsitakse temalt: “Kas teate veel mõnda X-i?”), on meil tegemist vaba pikkusega loeteluga. Et enamasti pole vahetegemine piiratud ja piiramata loetlemiskatse vahel oluline, võime rääkida lihtsalt loetelukatsest. Loetelukatse

² Sõna “loetelukatse” võimalikud ingliskeelsed vasted oleksid *list task* (Smith 1993), *free-list* (Smith jt 1995), *free-recall listing* (Weller 1998), *free listing* (Weller, Romney 1988; Trotter, Schensul 1998), *free listing task* (Robbins, Nolan 1997), *free-listing* (Robbins, Nolan 2000), *freelist* (Borgatti 1999), samuti *Approach A* Lennebergi värvisõnade uurimuses (Lenneberg 1967: 339–340, 345 *associative responses* (Bousfield, Sedgewick 1944) jt.

metoodikat on kirjeldanud sellised autorid nagu Weller ja Romney (1988), Weller (1998) ja Borgatti (1999).

2. Sõnasagedus

Loetelukatse puhul tuleb kindlasti arvestada uuringu eesmärki. Kui kõik, mida antropoloog, lingvist või psühholoog soovib hetkel teada, on teatud kultuuri- või tunnetusvaldkonna piirid, piisab selleks täiesti vaid ühest loetelukatse parameetrist, milleks on sõnasagedus. Tõsi, küsitletavate arv ei tohi olla liiga väike. Üldtunnustatud miinimumiks on 20–30 katsealust (Weller, Romney 1988: 14; Borgatti 1999: 122–30). Davies ja Corbett (1994) küsitlesid vene põhivärvide väljaselgitamiseks 77 katseisikut. Mu enda uuringud eesti keele värvi-, lõhna-, maitse- ja temperatuurisõnadest rajanevad 80 küsitlusel (Sutrop 1998, 2000, ilmumas).

Kui katsealuseid piisab, võib rahulikult rääkida sagedusest kui parameetrist. Sõnad, mida on mainitud vaid üksikutes loeteludes, tuleks arvata juhusõnade hulka. Võimalusi, kuidas domeenipiiri kindlaks määrata, on Borgatti (1999:125) välja pakkunud kolm:

- 1) arva hulka kõik sõnad, mis esinevad rohkem kui ühes loetelus;
- 2) püüa leida loomulikke rühmi;
- 3) aseta piir suvalisse kohta.

Õigupoolest on siin tegemist vaid kahe reegluga, kuna esimene on lihtsalt kolmanda erijuht: üks ole ju otsus asetada piir ühes ja kahes loetelus esinemise vahele samuti suvaline.

Rusikareegel oleks järgmine: kui katsealuseid on vähe (20), jätta välja sõnad, mis esinevad vaid ühes loetelus, ent kui katsealuste arv on suurem (nt 50–80), jätta välja sõnad, mida on mainitud kolmes loetelus või harvemini.

Põhjus, miks harvaesinevad sõnad domeeni arvamata tuleks jätta, on järgmine: loeteludes esinevad sageli vaid need sõnad, mis on (antud keeles) aktiivselt kasutuses. Harvaesinevad sõnad seevastu on kas passiivsed või kasutatakse neid vaid mõnes idiolektis. Muidugi satub hulk passiivseid termineid ikkagi domeeni ka sel juhul, kui ainult vähestes loeteludes esinevad terminid välja jätta, toimuigu see siis suvaliselt või loomulikku murdekohta ehk rühmitumist arvestades. Sestap tuleks domeenis aktiivselt kasutatavate sõnade kindlaksmääramiseks appi võtta ka teine loetelukatse parameeter, nimelt sõna keskmine astak (astakukeskmine).

3. Sõna keskmine astak

Sõna keskmine astak (astakukeskmine) arvutatakse ainult nende loetelude põhjal, kus vastav sõna tõepoolest esineb. Üldiselt usutakse, et sõna sageduse ja keskmise astaku (järjenumbr) vahel valitseb küllaltki tihe positiivne korrelatsioon (st kõige arvukamalt loeteludes esinevad sõnad tulevad tavaliselt meelde esimeste hulgas, neid sõnu aga, mis esinevad vaid mõnes loetelus, mainitakse harilikult loetelu lõpupoole) (Bousfield, Barclay 1950). Seetõttu võibki astakukeskmise arvestamata jätta, kui tegemist on vaid domeeniipiiride määramisega.

Reaalsed andmed näitavad ometi, et seos sõna sageduse ja astakukeskmise vahel pole kaugeltki ideaalne. Näiteks mu enda uurimuses eesti keele temperatuurisõnadest polnud sõnade “kõrge” ja “madal” esinemus sugugi suur – neid mainis vaid 9 katsealust 80st, kuid sõna “kõrge” keskmine astak oli 1 ja sõnal “madal” oli see 4. Täheandab, kui neid sõnu üldse mainiti, tehti seda kohe esimeste seas. Antud juhul annaks hälvet võib-olla seletada koolifüüsika mõjuga üldkeelele, sest just seal jagatakse temperatuure eranditult kõrgeteks ja madalateks (Sutrop 1998: 93).

Keeleantropoloogile harilikult ei piisa üksnes domeeniipiiride määramisest. Tema vajab täpsemat teavet ka domeeni siseehitusest. Teda võib näiteks huvitada domeeni jagunemine põhi- ja lisaõnadeks.³ Muid võimalikke loetelukatse rakendusi (sõnade koosesinemine, mitmemõõtmeline skaleerimine, kahelijaotus, kobaranalüüs jne.) on kirjeldanud Borgatti (1999: 126–31) ning Robbins ja Nolan (1997, 2000).

Domeeni põhisõnade väljaselgitamiseks (enamasti on need lihtsa morfoloogiaga lühikesed omasõnad) on kõigepealt vaja mingeid objektiivseid kriteeriume, mis võimaldaksid sõnu järjestada ning eristada põhisõnu mittepõhilistest. Omaette võetuna ei piisa selleks ei

³ Mõistan “põhisõna” samas vaimus, nagu Berlin ja Kay (1969: 5–7) oma põhivärvisõna määratluses: põhisõna on psühholoogiliselt esiletulev, enamasti morfoloogiliselt lihtne omasõna, mis kuulub oma valdkonna prototüüpse(te) liikme(te)ga samasse klassi ning millel on viimastega ühesugune grammatiline potentsiaal. Üldiselt tähistab põhisõna mingit objekti, omadust või nähtust põhitasemel ning on kasutatav kõigis asjakohastes situatsioonides (vt Sutrop 1998: 61, ilmumas).

sõna sagedusest ega ka astakukeskmisest, kuna kumbki annab erinevad põhisoanakandidaadid. Seetõttu näib mõistlik koondada sõna sagedus ja astakukeskmise üheks kognitiivse esiletuleku indeksiks.

4. Kognitiivse esiletuleku indeks

Borgatti (1999: 123) suhtub üpris skeptiliselt neisse, kes “mõistes, et nii sagedus kui ka keskmine astak peegeldavad tegelikult üht ja sama omadust (s.o esiletulekut) [...], ühendavad need meelsasti üheks näitajaks” Järgnevalt kavatsen näidata, et tema skeptitsismil pole alust.

Kui ma alles asusin Eestis tajusõnu koguma, kasutasin põhivärvisõnade väljaselgitamiseks Daviese ja Corbetti (1994, 1995) välimeetodit. Protseduur koosneb kahest etapist: kõigepealt sooritavad katsealused loetelukatse (“palun, nimetage nii palju värve kui teate”), seejärel nimeandmiskatse (igale katseisikule näidatakse 65 värviruudukest juhuslikus järjekorras, paludes tal nimetada iga ruudukese värvi). Katsete tulemused kattusid.

Et tajusõnadel ei ole stiimulistandardit, mille täpsus oleks võrreldav värvisõnade omaga (65 kindla tooniga värviruudukest), laien-dasin loetelukatse meetodikat uurimaks teisigi meeli peale värvitaju. Intervjueerinud 80 katsealust iga domeeni kohta, sain tulemuseks teatud korpused värvi-, lõhna-, maitse- ja temperatuurisõnu. Seejärel arvutasin igale sõnale kaks parameetrit: sageduse ja keskmise astaku (Davies ja Corbett 1994, 1995). Samas leidsin, et need parameetrid vastavad psühholoogilise esiletuleku kriteeriumidele põhivärvisõna definitsioonis. Nimelt kirjutavad Berlin ja Kay (1969: 6) oma põhivärvisõna määratluses, et psühholoogilist esiletulekut iseloomustavad muuhulgas järgmised näitajad:

- 1) kalduvus esineda loetelude alguses;
- 2) esinemisstabiilsus nii katseisikute kui ka kontekstis;
- 3) kuulumine kõigi katsealuste idiolekti.

Seepeale otsustasin esimese ja kolmanda näitaja ühendada, kuna kalduvus esineda loetelude alguses vastab sõna keskmisele astakule ning sõna esinemine kõigi katsealuste idiolektides vastab selle sõna sagedusele.

Kognitiivse esiletuleku indeks S (*saliency*) iseloomustab sõna psühholoogilist esiletulekut loetelukatses, arvestades nii sagedust F (*frequency*) kui ka keskmist astakut mP (*mean position*). Esiletuleku valem on

$$S = F/(N mP),$$

kus sõna esinemiskordade arv F kõigis katses moodustunud loeteludes kokku on jagatud sõna keskmise astakuga mP kõigis loeteludes, kus see sõna esineb, ning katsealuste arvuga N . Kui kõik katsealused on nimetanud teatud sõna ($F=N$) esimesena ($mP=1$), siis on selle sõna esiletuleku indeks $S=1$.

Kognitiivse esiletuleku indeks on normeeritud varieeruma 1-st 0-ni. Igas domeenis on see indeks kõrgeim põhisõnadel. Ideaalse esiletulekuga sõna indeks $S=1$, samas kui sõnadel, mida mainitakse viimases järjekorras ja harva, on indeksi S väärtus nullilähedane. Sõnal, mida ei mainita üheski loetelus, on esiletulek 0. On oluline, et kognitiivse esiletuleku indeks võimaldab võrrelda eri uuringute tulemusi, kuna indeks ei sõltu loetelude pikkusest.

Nii sagedust, astakukeskmist kui ka kognitiivse esiletuleku indeksit võib kõiki kasutada domeeni põhisõnade väljaselgitamiseks. Mõnikord tuleb siiski vahet teha ka enam ja vähem põhiliste sõnade vahel. Sel puhul on abi teatud lingvistilistest kriteeriumidest.

5. Arutelu

Lingvistilises ja antropoloogilises kirjanduses on esiletulekut ka varem mõõdetud. Sõltumatult on Smith leiutanud oma indeksi, mis samuti koondab sageduse ja keskmise astaku üheks parameetriks, vaba pikkusega loetelus esiletuleku indeksiks (VPLEI) (Smith 1993; Smith jt 1995; Smith ja Borgatti 1997).

Oma esimeses selleteemalises artiklis pakub Smith oma indeksi arvutamiseks järgmise valemi:

$$\text{sõna VPLEI} = \frac{\text{sõna protsentilastakute summa}}{\text{loetelude koguarv, protsentilastak} = ((\text{loend järjenumber})/\text{loend})100)/N}.$$

Kui tähistada loetelude koguarv N -iga, loend tähendaks loetelu pikkust L_i (*length*) ja järjenumber vastava sõna astakut R_j (*rank*) loetelus L_i , võiksime sõnad ülaltoodud valemis asendada, teisendades seega valemi (Smith 1993) kujule

$$S_j = (\sum (((L_i - R_j) / L_i) 100)) / N.$$

Smith annab esimesele sõnale loetelus väärtuse 0, teisele 1 jne. Ideaalne sõna, mis esineb kõigil katsealustel ja sealjuures alati esimesena, saab indeksi S_j väärtuseks 100%, kuna

$$S_1 = (\sum (((L_i - 0) / L_i) 100)) / N = 100\%.$$

Loetelu viimase sõna järjenumbri astak on $R_j = loend - 1 = L_i - 1$. Järelikult läheneb alati viimasena mainitud sõna esiletulek nullile, kuna

$$S_{j=L} = (\sum((L_i - (L_i - 1))/L_i)100)/N = (\sum((1/L_i)100))/N > 0\%.$$

Artiklis Smith jt (1995) on ülaltoodud (vaba pikkusega loetelus esiletuleku) indeksit rakendatud inglise värvisõnade domeenianalüüsil. Ka siin leidub arvutusprotseduuri sõnaline kirjeldus, mille näpuviga paraku mõistetamatuks muudab (autori vabandus on toodud artiklis Smith and Borgatti 1997: 208).

Tegelikult puudub tollest valemist üksainus sõna "miinus", mille lisamisel saaksime:

$$\text{sõna indeksi väärtus} = (\text{loetelu pikkus MIINUS sõna astak}) \times (1/(\text{loetelu pikkus} - 1)).$$

Kui loetelude koguarv oleks N , loetelu pikkus L_i ja sõna astak loetelus L_i oleks R_j , saaks sõna S_j esiletuleku indeksit vaba pikkusega loetelukatses väljendada järgmise valemiga:

$$\text{sõna indeksi väärtus} = (L_i - R_j)/(L_i - 1) = (L_i - R_j)/(L_i - 1).$$

Et sõna keskmine indeksiväärtus arvutatuna kõigi katses saadud loetelude põhjal (olenemata sellest, kas vastav sõna neis esineb või mitte) ongi selle sõna esiletuleku indeks vaba pikkusega loetelus, saame sõna indeksit S_j väljendada järgmise valemiga:

$$S_j = (\sum((L_i - R_j)/(L_i - 1)))/N.$$

Ideaalne sõna, s.o sõna, mida nimetavad kõik katseisikud ja sealjuures alati esimesena, saab indeksi $S_1 = 1$, kuna

$$S_1 = (\sum((L_i - 1)/(L_i - 1)))/N = 1.$$

Seevastu sõna, mida mainitakse alati viimasena, saab indeksiks $S_{j=L} = 0$, kuna

$$S_{j=L} = (\sum((L_i - L_i)/(L_i - 1)))/N = 0.$$

Oma järgmises artiklis püüavad Smith ja Borgatti (1997) korrigeerida ülalmainitud eksitust (Smith jt 1995). Põhjendusi toomata annavad nad indeksi väärtuse 0 neile sõnadele, mis üheski loetelus esile ei tulnud (Smith, Borgatti 1997: 209). See vastab varem ilmunud meetodikale (Smith 1993), kus nende sõnade indeks, mis alati loetelu lõppu jäid, küll lähenes nullile, kuid päris nulliks ei muu-

tunud, mistõttu 0 jäi n-ö “nähtamatule sõnale, mida ei mainita üheski loetelus” (Smith jt 1993: 1). Kummalgi juhul erineb protseduur varasemast meetodikast, kus indeksi väärtus 0 kuulus sõnale, mis esines igas loetelus viimasel kohal (Smith jt 1995).

Uues valemis (Smith ja Borgatti 1997) tähistab L loetelu pikkust (s.o sõnade arvu loetelus), R_j on positsioonil j asuva sõna astak (järjenumber) loetelus ($R_1=1$) ja N on analüüsitavate loetelude arv:

$$S = ((\sum(L - R_j + 1))/L)/N.$$

Et L on loetelu pikkus, võime seda täiendada indeksiga i (L_i), mis läbi valem teisendub kujule

$$S = ((\sum(L_i - R_j + 1))/L_i)/N.$$

Sellest kujust aga nähtub, et ka parandatud valem pole veatu. Nimelt tuleb selle järgi kõigepealt summeerida üksikud loetelud ning seejärel jagada saadud summa ühe konkreetse loetelu pikkusega, sellel jagamisel pole lihtsalt mõtet. Valem hakkab tööle alles siis, kui sulgusid pisut nihutada:

$$S = (\sum((L_i - R_j + 1)/L_i))/N.$$

Mu enda varasem kognitiivse esiletuleku indeks S kujutab endast kahe teguri korrutist (Sutrop 1998, 2000):

$$S = (F/N)[(L - mP)/(L - 1)].$$

Esimene tegur F/N saadakse sõna loetelukatses esiletuleku sageduse (s.o loetelude arvu, kus vastavat sõna on mainitud) jagamisel katseisikute arvuga. Kui mingi sõna esineb kõigil katseisikutel, siis selle sõna $F/N=1$. Teine tegur on $(L - mP)/(L - 1)$, kus mP on sõna astaku kaalutud keskmine ja L on loetelu keskmine pikkus. Kui sõna keskmine järjenumber on 1, on teise teguri väärtus samuti 1. Et ideaalsel põhisõnal, s.o sõnal, mille esiletulek on maksimaalne, on mõlemad tegurid võrdsed ühega, siis ka nende $S=1$. Kui leidub sõna, mille keskmine astak mP võrdub loetelu keskmise pikkusega L , st kui sõna $mP=L$, siis selle sõna esiletuleku indeks $S=0$. Kui aga sõna astaku keskmine on suurem kui loetelu keskmine pikkus, on sõna esiletuleku indeks negatiivne ($S<0$). Seega võib ülaltoodud indeksi positiivseid väärtusi tõlgendada kognitiivselt enam esiletulevate sõnade tunnuseks ja vähem esiletulevate sõnade tunnuseks. Nullpunktis võiks asuda sõna, mille järjenumber võrdub loetelude keskmise pikkusega. Järelikult võimaldab indeks

tõepoolest liigendada teatud kultuuri- või tunnetusdomeeni olenevalt selle valdkonna sõnade esiletulekule loetelukatses.

Siinkohal eelistaksin tollest indeksist siiski loobuda ja seda kahel põhjusel: esiteks saab indeksit kasutada vaid juhul, kui jätta arvestusest välja kõik sõnad, mis esinevad üksnes vähestes loeteludes. Muidu võiks mõni sõna, mida on loetelu lõpus maininud vaid üks katseisik, anda paremaid tulemusi kui paari-kolme isiku mainitud sõnad. Teiseks – ja see kehtib ka Smithi ja ta kolleegide väljapakutud indeksite kohta – tulemused sõltuksid loetelu pikkusest, st lühikesed ja pikad loetelud annaksid erinevaid tulemusi. Viimane asjaolu tekitab aga raskusi nii konkreetse uuringu tulemuste tõlgendamisel kui ka eri teadlaste ja eri uuringute tulemuste võrdlemisel.

Käesolevas artiklis kirjeldatud uus kognitiivse esiletuleku indeks (KEI) erineb äsja analüüsitud esiletuleku indeksist vaba pikkusega loetelus. Nimelt toimub viimase arvutamine üksikute loetelude põhjal, samas kui KEI arvutatakse sõna sagedusest ja astakukeskmisest järgmise valemi järgi:

$$S = F/(N mP).$$

KEI võimaldab liigendada analüüsitavat kultuuri- või tunnetusvaldkonda vastavalt sõnade esiletulekule selles domeenis. Alustuseks võime kasutada sõna esinemissagedust katses saadud loeteludes ja jätta arvestusest välja kõik äärmiselt madala sagedusega, st antud katses harva esinevad sõnad. Ideaalse (maksimaalse) esiletulekuga sõna S väärtus oleks 1. Võime ka eeldada, et suurema S väärtusega sõnad on aktiivses kasutuses, samas kui väga väikese S väärtusega sõnad (nagu ülimaldala sageduse tõttu väljajäetud sõnadki) kuuluvad passiivse sõnavara hulka. Vajadusel võib veel omistada väärtuse $S=0$ kõigile neile sõnadele, mis loeteludes mainimist ei leidnud. (Valemi järgi klapiib kõik ka algebraliselt, sest kui jagatav $F=0$, on jagatis samuti 0; siinjuures tuleb eeldada, et katses mainimata jäetud sõna astak pole mitte 0, vaid lõpmatus.)

Skeptitsismiks pole KEI puhul mingit põhjust. Kui ülesanne seisneb vaid kultuuri- või tunnetusvaldkonna ehk -domeeni piiritlemises, pole KEI arvutamine üldse vajalik. Loetelukatse puhul piisab sel juhul sagedusmõõdustki. KEI-d vajavad keeleantropoloogid, psühholoogid, tunnetusteoreetikud jt ühiskonnateadlased selleks, et järjestada teatud valdkonna sõnu nende esiletuleku alusel ning jagada need sõnad põhilisteks ja mittepõhilisteks (põhi- ja lisasõnadeks). Et näidata, kuidas KEI toimib, esitan järgnevalt mõningaid isiklikke

F – sagedus, *R* – astak (järjenumber), *mP* – keskmine astak, *eS* – varasem esiletuleku indeks, *S* – esiletuleku indeks, *f/S* – esiletuleku indeks vaba pikkusega loetus. Kursiiv tähistab domeeni põhisõna.

Mu enda poolt varem kasutatud indeks sõltus samuti loetelude pikkusest, kuna esiletulek sõltub loetelude keskmisest pikkusest. Sõna “parfüüm” indeksi väärtus sai seetõttu ootamatult madal. Samas sai sõna “nohu”, mida mainiti vaid korra ja seda neljandana ($mP=4$), esiletulek kõrgema väärtuse kui “vine”, mida mainiti kaks korda ($mP=4$).

Nagu ülaltoodust näha, tasub eelistada uut lihtsat KEI-d, mis rajaneb kahel tunnetuslikult olulisel parameetril (sõna sagedus ja keskmine astak) ning on vaba kõrvalmõjudest, mis on seotud loetelude pikkusega ning ilmnevad nii varasema kognitiivse indeksi kui ka VPLEI puhul. Uus KEI sobib ka lühikeste loetelude ja väikese arvu katsealuste korral, mida ei saa öelda vana valemi kohta, mis mäletatavasti sõltus loetelude pikkusest. Lisaks eeldas kognitiivse esiletuleku arvutamine loetelukatses vana indeksi järgi piisavat arvu katseisikuid. Pealegi andis too indeks korralikke tulemusi ainult sel tingimusel, kui harva (st üksnes mõne katsealuse poolt) mainitud sõnad jäeti arvestamata.

Värvisõnade valdkond on eesti keeles väga suur. Loetelukatses kogunes 80 katsealuselt 285 eri värvisõna (neist enamik liitsõnad ja tuletised). Loetelukatse ja nimeandmiskatse kokku andsid koguni 638 eri värvinime. Loetelu keskmine pikkus oli 18.94 sõna. Tervelt 71 värvisõna nimetati loetelukatses nelja või enama katseisiku poolt. Tabelis 2 näeme kõige esiletulevamaid värvisõnu (sealhulgas põhisõnad on toodud kursiivis). KEI näitab selget vahet põhisõnade ja suurema KEI väärtusega lisasõnade vahel. Põhisõna “hall” viib esiletuleku indeks 11. kohale, jättes esimese lisasõna “beež” 12. kohale.

Tabel 2. Eesti värvisõnad loetelukatses

Sõna	F	R	mP	R	S	R
<i>sinine</i>	71	2	3.66	1	0.242	1
<i>punane</i>	71	2	4.46	2	0.199	2
<i>roheline</i>	69	4.5	5.75	3	0.150	3
<i>kollane</i>	71	2	6.07	4	0.146	4
<i>must</i>	69	4.5	7.13	5	0.121	5

Sõna	F	R	mP	R	S	R
<i>valge</i>	66	7	7.74	6	0.107	6
<i>oranž</i>	68	6	9.79	13	0.087	7
<i>lilla</i>	58	9	8.82	9	0.082	8
<i>pruun</i>	52	11.5	9.71	12	0.067	9
<i>roosa</i>	53	10	11.24	19	0.059	10
<i>hall</i>	64	8	13.80	31	0.058	11
<i>beež</i>	52	11.5	11.44	21	0.057	12
<i>hele-sinine</i>	28	13	7.82	7	0.045	13
<i>tume-sinine</i>	22	14.5	9.90	14	0.028	14
<i>violetne</i>	15	19	8.73	8	0.021	15
[...]						

F – sagedus, R – astak, mP – keskmine astak, S – esiletulek. Kursiivis on toodud 11 põhivärvisõna.

VPLEI annab teistsuguse tulemuse. Uurimuses inglise värvisõnadest (Smith jt 1995), kus loetelukatset analüüsiti VPLEI abil, platseerus sõna *violet* ‘violetne’, mis inglise keeles ei ole põhivärvisõna, indeksi väärtuse alusel 11. kohale, samas kui põhivärvisõna *grey* ‘hall’ jäi kaheteistkümnendaks. Näeme, et põhiliste ja mittepõhiliste sõnade eristamisel osutub KEI edukamaks kui VPLEI.

6. Kognitiivse esiletuleku indeksi (KEI) arvutamine

1. Registreerige sõnad täpselt samas järjekorras, kui need loetelus esile tulid.

2. Loetelukatse tulemuste põhjal arvutage välja sõna keskmine astak (*mP*), kusjuures

F – sõna sagedus (s.o loetelude arv, kus see sõna esineb),

N – loetelude (katseisikute) koguarv,

R_j – sõna astak (järjenumber) konkreetses loetelus.

Sellise sõna sagedus, mis üheski katse tulemusel saadud loetelus esile ei tule, on 0 (*F₀* = 0) ning selle sõna keskmine astak on lõpmatus (*mP₀* = ∞).

2.1. Sõna keskmine astak on leitav järgmisest valemist:

$$mP = (\sum R_j) / F$$

3. Arvutage sõna KEI:

$$S = F / (N \cdot mP).$$

3.1. Vahel on praktilisem arvutada KEI järgmisest valemist:

$$S = F^2 / (N \sum R_j).$$

Katsealuseid peaks loetelukatsest KEI arvutamiseks olema vähemalt 20–30.

Kokkuvõte

Loetlemiskatse võib olla suuline või kirjalik, ajaliselt piiratud või vaba pikkusega. Kahest loetelukatse parameetrist (s.o sõna sagedusest ja keskmisest astakust neis loeteludes, kus vastavat sõna on mainitud) saab ühtekokku tuletada esiletuleku indeksi. Artiklis pakutakse välja kognitiivse esiletuleku indeks (KEI) S ja järgnev valem indeksi arvutamiseks:

$$S = F / (N mP).$$

KEI lähtub kahest tunnetuslikult olulisest parameetrist: need on sõna sagedus ja keskmine astak (järjenumber). Keskmine astak (astaku-keskmine) peegeldab sõna võimalikku kalduvust esile tulla loetelu alguses, sõna sagedus aga väljendab sõna esinemist võimalikult kõigi katsealuste idiolektides.

Kolmas parameeter on katseisikute arv. Antropoloogidele, keeleteadlastele jt ühiskonnateadlastele, kes end matemaatikas võib-olla vähem kodus tunnevad, peaks selline lahendus, kus kaks mõõdetavat suurust koondatakse ühteainsasse parameetrisse, sümpaatne tunduma.

Et loetelude pikkus KEI-d ei mõjuta, muutuvad kõik KEI-d kasutatavad uuringud omavahel võrreldavaks. KEI abil saab analüüsivat kultuuri- või tunnetusvaldkonda (domeeni) ka liigendada. Iga domeeni põhisõnad on ühtlasi kõige esiletulevamad. Absoluutselt esiletulevaim sõna, s.o sõna, mida kõik katseisikud nimetavad esimesena, saab KEI väärtuseks 1, kusjuures aktiivses kasutuses olevatel sõnadel tuleb indeksi väärtus suurem kui passiivsetel sõnadel. Kui sõna esiletulek on vähene, läheneb selle indeksi väärtus nullile. Sõnad, mida loetelus üldse ei mainita, saavad indeksi väärtuseks 0.

Kirjandus

- Berlin, B.; Kay, P. 1969 (1991). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press. (Kordus-trükk 1991 koos täiendava bibliograafiaga, mille koostas L. Maffi)
- Borgatti, S. P. 1999. Elicitation techniques for cultural domain analysis. – *Enhanced Ethnographic Methods: Audiovisual Techniques, Focused Group Interviews, and Elicitation Techniques*. Toim J. J. Schensul, M. D. LeCompte, B. K. Nastasi, S. P. Borgatti. *Ethnographer's Toolkit, Volume 3*. Walnut Creek, CA: AltaMira. 115–51.
- Bousfield, W. A.; Barclay, W. D. 1950. The relationship between order and frequency of occurrence of restricted associative responses. – *Journal of Experimental Psychology* 40 (5), 643–647.
- Bousfield, W. A.; Sedgewick, C.H.W. 1944. An analysis of sequences of restricted associative responses. – *The Journal of General Psychology* 30, 149–165.
- Davies, I.; Corbett, G. 1994. The basic color terms of Russian. – *Linguistics* 32, 65–89.
- Davies, I.; Corbett, G. 1995. A practical field method for identifying basic colour terms. – *Languages of the World* 9 (1), 25–36.
- Lenneberg, E. H. 1967. *Biological Foundation of Language*. New York: John Wiley.
- Robbins, M. C.; Nolan, J. M. 1997. A measure of dichotomous category bias in free listing tasks. – *Cultural Anthropology Methods Journal* 9 (3), 8–12.
- Robbins, M. C.; Nolan, J. M. 2000. A measure of semantic category clustering in free-listing tasks. – *Field Methods* 12 (1), 18–28.
- Smith, J. J. 1993. Using ANTHROPAC 3.5 and a spreadsheet to compute a free-list salience index. – *Cultural Anthropology Methods Journal* 5 (3), 1–3.
- Smith, J. J.; Borgatti, S. P. 1997. Salience counts – and so does accuracy: Correcting and updating a measure for free-list-item salience. – *Journal of Linguistic Anthropology* 7 (2), 208–209.
- Smith, J. J.; Furbee, L.; Maynard, K.; Quick, S.; Ross, L. 1995. Salience counts: A domain analysis of English color terms. – *Journal of Linguistic Anthropology* 5 (2), 203–16.
- Sutrop, U. 1998. Basic temperature terms and subjective temperature scale. – *Lexicology* 4 (1), 61–104.
- Sutrop, U. 2000. The basic colour terms of Estonian. – *Trames* 4 (1), 143–168.

- Sutrop, U. ilmumas. The vocabulary of sense perception in Estonian: Structure and history. (Opuscula Fenno-Ugrica Goettingensia.)
- Trotter, R. T., II; Schensul, J. J. 1998. Methods in applied anthropology. – Handbook of Methods in Cultural Anthropology. Toim. H. R. Bernard. Walnut Creek, CA: AltaMira. 691–735.
- Weller, S. C. 1998. Structured interviewing and questionnaire construction. – Handbook of Methods in Cultural Anthropology. Toim. H. R. Bernard. Walnut Creek, CA. AltaMira. 365–409.
- Weller, S. C.; Romney, A. K. 1988. Systematic Data Collection. (Qualitative Research Methods, 10.) Newbury Park, CA: Sage.