

VARDŲ DISTRIBUCIJOS TYRIMAS STATISTIKOS METODAIS

Ina Kliūnaitė, Aurimas Nausėda

Šiaulių universitetas, Humanitarinis fakultetas

Lietuvos ir užsienio šalių lingvistams pradėjus naudoti kompiuterines programas tekstams analizuoti, atsirado kalbų tekstynai, galimybės gilintis į kalbos savitumus. Atsiradusi kalbotyros rūšis – tekstynų lingvistika – leido lingvistams (Korsakas, 1998; Marcinkevičienė, 2001; Mikelonienė, 2000; Korsakas, Nausėda, 2007) gilintis į kalbos reiškinių savitumus, būdingus kalbos dalims, sakiniams.

Lietuviškų ir lotyniškų vardų tyrinėjimai atspindi mokslo literatūroje (Grumadienė, Žilinskienė, 1998; Korsakas, 1998; *LKE*, 1993; Mikelonienė, 2000; Marcinkevičienė, 2000; Sinclair, 2000; Tognini Bonelli, 2000; Utkā, 2005; Rimkutė, 2006; Kuzavinis, Savukynas, 2007 ir kt.). Lingvistai yra sudarę lietuvių ir kitų kalbų dažninius žodynus, tekstynus, parengę lietuvių vardų kilmės žodyną, pakankamai daug nagrinėjo kolokacijas ir pan.. Tyrinėjant lietuvių vardų kilmės žodyną panaudojama statistinė analizė siekiant išsamiau paaiškinti surinktus duomenis. Pasirenkami aprašomosios statistikos metodai – dažnių lentelė, grafinis duomenų vaizdavimas (dažnių lentelės pateiktos linijiniu arba stulpelinu grafikai). Dažnių lentelė laikoma svarbia aprašomosios statistinės analizės pirmąja dalimi, kuri svarbi tiriant, kaip kiekvieno tiriamojo dalys yra pasiskirsčiusios duomenų dalyje, atrinktoje tyrimui (Rudzkienė, 1995). Lingvistikos kontekste dažnio skaičiavimas, dažnio paplitimas naudojami nustatyti lingvistiniams vienetams, funkcionuojantiems tam tikrais kalbos konteksto naudojimo atvejais, t. y. dažniausiai žodžių skaičiaus tekstuose sudarant žodžių sąrašus, svarbius mokantis užsienio kalbos ar parenkant žodžius iš tekstyno (McArthur, 1998).

Straipsnio naujumas sietinas su lietuviškų ir lotyniškų vardų funkcionavimo intensyvumu distribucijose, kurių leksinė sudėtis apibūdinama matematinės kalbotyros (lingvostatistinės analizės) rezultatais.

Straipsnio tikslas – ištirti kai kurių lietuviškų ir lotyniškų vardų vartojimo intensyvumą distribucijose, nustatyti su vardais susijusios leksikos įvairovę, apimtį ir dažninę sudėtį, taip pat bendrybes ir skirtības centro bei periferijos atžvilgiu.

Uždaviniai:

1. Tekstyno pagrindu sudaryti tiriamųjų vardų distribucijų žodynėlius.
2. Nustatyti vardų ir su jais susijusių žodžių vartojimo dažnumus.

3. Apibūdinti pasirinktų vardų distribucijose intensyviai vartojamos leksikos statistikos parametrus.
4. Apibendrinti rezultatus ir parengti išvadas.

Tyrimo metodai. Naudojami aprašomuoju analitiniu būdu ir matematinės kalbotyros, arba statistinės analizės (lingvostatistikos), metodais. Skaičiavimai buvo atlikti statistiniu „Microsoft Excel 2003“ paketu.

Tyrimo rezultatai. Tyrimui laisvai parinktos trumpos moterų ir vyrų vardų sekos: 1) lietuvių: *Audronė, Aušra, Birutė, Eglė, Jūratė; Algirdas, Gediminas, Giedrius, Jogaila, Mindaugas, Šarūnas, Vytautas*; 2) lotynų: *Akvilė, Diana, Kristina, Patricija, Viktorija; Dominykas, Donatas, Modestas, Paulius, Remigijus*. Vardų distribucijų leksika buvo imama iš labai didelės imties enciklopedinių tekstų ar tekstynų (*DLKT*). Vardai ir su jais susietos leksemos diferencijuotos pagal funkcionavimo intensyvumą – dažnių vartojimą (centrą ir periferiją) originalios (savarankiškai sumanytos bei realizuotos) kompiuterinės programos pagrindu.

Lietuviškų vardų distribucijų leksikos lingvostatistinės charakteristikos atspindi mūsų programiškai sudarytuose dažniniuose žodynėliuose. Vardas – tarytum raktinis teminis žodis: jis distribucijų dažniniuose žodynėliuose yra beveik visada pirmosiose pagal vartojimo intensyvumą pozicijose ir priskirtinas prie menamojo leksikos branduolio, arba centro.

Apžvelgsime kai kuriuos statistinės analizės pavyzdžius. Lietuviškas vardas *AUDRONĖ* yra priesagos *-onė* vedinys (iš žodžio „audra“). Tekstyno pagrindu gauta distribucijos apimtis – 58 puslapiai; pavartota 28991 žodis; vidutiniškas leksemos ilgis – 7–8 raidės (t. y. leksemų ženklų, arba raidžių, su tarpais skaičiaus ir tekstuose pavartotų žodžių skaičiaus santykis [$231564 : 28991 = 7,99$]). Vardas *AUDRONĖ*, arba raktinis žodis, distribucijų dažniniame žodynėlyje yra antrame range (dažnumas – 1275), po jungtuko *ir* – 1384. Prie šio vardo distribucijos centro priskirtinos leksemos, pažymėtos triženkliais ar dviženkliais dažnumų indeksais, pavyzdžiui: *į* – 742, *kad* – 205, *jo* – 188, *su* – 183, *tai* – 107, *kaip* – 100, *buvo* – 93, *jos* – 92, *jau* – 84, *savo* – 83, *yra* – 68, *žvilgsnis* – 59, *kas* – 49, *skyriaus* – 49, *mano* – 43, *labai* – 41, *gal* – 39, *gali* – 36, *vakar* – 36, *dabar* – 33,

tapyba – 32, darbo – 31, mokyklos – 31, seimo – 27, apsaugos – 26, centro – 25, sveikatos – 25, universiteto – 25, mokytoja – 23, knyga – 21, vyriausioji – 20, kartu – 19, vienas – 18, moteris – 17, ligoninė – 16, mokslo – 15, nebuvo – 14, gyvenimo – 13, dienos – 12, mama – 11, direktorius – 10 ir kt.; tikriniai vardai: Lietuvos – 135, Kauno – 93, Kanapickienė – 71, Vilniaus – 63, Danielius – 24, Europos – 24, Vytautas – 23, Sabina – 21, Jonas – 19, Danijos – 18, Girdzijauskaitė – 17, Antanas – 16, Juozas – 15, Petras – 15, Algirdas – 11, Arvydas – 10 ir kt.

Iš pateiktų pavyzdžių matyti, jog kalbos leksinė sistema iš tikrųjų nuolat kinta, bet turi ir tam tikrų pastovumo, dėsningumo požymių. Pavyzdžiui, netgi kompiuteriškai (be žmogaus valios ir iš esmės mechaniškai) sudarytame konkreta vardo, kaip raktinio žodžio, dažniniame žodynėlyje pirmose rangų pozicijose yra kai kurie jungtukai, prielinksniai, įvardžiai,rieveiksmiai, įvairiomis formomis pavartotas veiksmožodis *būti* ir pan. Dažniausiai pavartoti kiti tikriniai vardai, pavardės ar pavadinimai atspindi sociolingvistinius tekstų bruožus: geopolitinius bei kultūrinius valstybės gyvenimo atspindžius, laikmečio iškilusių asmenybių pėdsakus ar žmonių sąmonės šliejimąsi prie Lietuvos istorijos vardų, tapatinimąsi per juos su įžymiaisiais žmonėmis.

Vardas *AUDRONĖ* dažniausiai buvo pavartotas periodikoje (šalies – 497 kartus, vietinėje – 427, populiariojoje – 159, specializuotoje – 71); be to, grožinėse knygose – 41, negrožinėje literatūroje – 34, memuaruose – 22, LR valstybiniuose dokumentuose – 24. Tai sudarė: a) 0,001 proc. nuo pasinaudoto tekstyno visumos [1275 x 100 : 102857327]; b) 3,4 proc. nuo moterų lietuviškų vardų [1275 x 100 : 37505]; c) 3,72 proc. nuo vyrų lietuviškų vardų [1275 x 100 : 34260]; d) 4,9 proc. nuo moterų lotyniškų vardų [1275 x 100 : 26000]; e) 4,71 proc. nuo vyrų lotyniškų vardų [1275 x 100 : 27087]; f) 3,01 proc. nuo šio vardo distribucijų leksikos visumos [1275 x 100 : 42329].

AUDRONĖS vardo distribucijos leksinė sudėtis (kaip ir visų kitų vardų leksika) yra labai įvairi. Tuo pabrėžtinai teigiama, jog didelio dažnio žodžių (kai pasikartojimų indeksai ne mažesni kaip 10) sudarytoje imtyje yra labai mažai (iki 5proc.). Mažo dažnumo indeksais (nuo 1 iki 4 pasikartojimų imtinai) pažymėta daugiau kaip 90 proc. pavartotų žodžių, arba: a) su indeksu „1“ rasta 65,19 proc. skirtingų leksemų, atrinktų iš 19,16 proc. šio vardo distribucijų visumos; b) su indeksu „2“ rasta 15,11 proc. skirtingų leksemų, atrinktų iš 2986 pavartojimų distribucijose, arba 8,8 proc.; c) su indeksu „3“ rasta 6,5 proc. skirtingų leksemų, atrinktų iš 1926 pavartojimų, arba 5,7 proc.; d) su indeksu „4“ rasta 3,5 proc. skirtingų leksemų, atrinktų iš 1384 pavartojimų, arba 4,12 proc.

Šio vardo distribucijos leksinę įvairovę ypač gerai atspindi mažo dažnumo pavyzdžiai, kurie čia teikiami su gretinamuoju indeksu, paimtu iš „Dažnio lietuvių kalbos žodyno“ (Grumadienė, 1998; indeksas teikiamas po žodžio, t. y. antrasis skaičius): *vadybininkas* – 1(0), *vaitas* – 1(3). *Vaivorykštė* – 1(4), *posmelis* – 1(5), *spamas* – 3(0), *spyris* – 1(2), *stakta* – 1(7), *steigtis* – 1(2), *tautodailininkas* – 1(5), *teatrologas* – 1(0), *telekomas* – 1(0), *temperamentingai* – 1(1), *tipologija* – 1(5), *tolerantiškas* – 1(1), *treniruotis* – 1(5), *turgavietė* – 1(4), *tvarkymasis* – 1(1), *žaismingumas* – 1(1), *akistata* – 1(6), *aklasis* – 1(0), *akumuliuoti* – 1(2), *alegorija* – 1(4), *alerginis* – 1(5), *altistas* – 1(1), *apgaulingas* – 1(3), *apyvoka* – 1(1), *apsisaugoti* – 1(6), *baiminimasis* – 1(1), *baletmeisteris* – 1(2), *cenzas* – 1(1) ir kt. Matome, kad šie žodžiai ir DDRLKŽ pažymėti mažais dažnumų indeksais arba jų pastarajame žodyne apskritai nėra. Kitaip tariant, jie priskirtini prie leksinės sistemos periferijos.

Su vardu *AUDRONĖ* distribucijose buvo tokie stabilūs junginiai (skaičius žymi pasikartojimus): *Audronė Kanapickienė* – 199, *Audronė Filimanavičienė* – 36, *Audronė Eitmanavičiūtė* – 29, *Audronė Girdzijauskaitė* – 28, *Audronė Vaitkutė* – 46, *direktore Audronė* – 22, *parengė Audronė* – 20, *vedėja Audronė* – 15, *Socialinės paramos skyriaus inspektore Audronė* – 2 ir kt.

Lietuviškam vardui *AUŠRA* buvo sudaryta 88 puslapių tekstyno dalies distribucijų apimtis; pavartota 45601 žodis; vidutiniškas leksemos ilgis – 7,6 raidės. Vardas *AUŠRA* distribucijų dažniniame žodynėlyje yra pirmajame range (dažnumas – 2768). Prie šio vardo distribucijos centro pavartota nemažai leksemų, pažymėtų triženkliais ar dviženkliais dažnumų indeksais, pavyzdžiui: *ir* – 1289, *į* – 1160, *jo* – 317, *kad* – 265, *su* – 265, *o* – 218, *buvo* – 216, *tai* – 178, *tik* – 166, *kaip* – 154, *savo* – 145, *kai* – 144, *jau* – 140, *prie* – 139, *apie* – 136, *ryto* – 127, *jos* – 125, *bet* – 114, *nuo* – 112, *po* – 111, *kas* – 102, *taip* – 90, *jis* – 80, *nes* – 63, *tas* – 55, *dabar* – 54, *man* – 52, *paroda* – 52, *vis* – 52, *labai* – 51, *vakar* – 51, *net* – 50, *gal* – 47, *daug* – 46, *mano* – 46, *spausdino* – 46, *kuri* – 45, *gali* – 44, *bus* – 41, *galerija* – 39, *spaustuvė* – 37, *metų* – 34, *muzikos* – 34, *vienas* – 34, *leido* – 33, *mokyklos* – 33, *turi* – 32, *vaiškų* – 32, *viena* – 30, *aprengti* – 29, *fotografijos* – 29, *geriausias* – 29, *dizainas* – 28, *mes* – 28, *metus* – 28, *prizai* – 28, *reikia* – 28, *sapnai* – 28, *galima* – 27, *tada* – 26, *tapo* – 26, *pati* – 25, *mokslo* – 24, *universiteto* – 24, *daugiau* – 23, *dienos* – 23, *kartu* – 23, *leidykla* – 23, *moteris* – 23, *nebuvo* – 23, *pats* – 23, *sako* – 23, *kaina* – 22, *kiek* – 22, *kovo* – 22, *tautos* – 22, *gyvenimo* – 22, *vasario* – 21, *darbo* – 21, *grafika* – 20, *kartais* – 20, *atgimimo* – 19, *atrodo* – 19, *jaunimo* – 19, *diena* – 19, *skyriaus* – 19, *varpas* – 19, *istorijos* – 18, *klubo* – 18, *tapyba* – 18, *kalba* – 17,

medicinos – 17, *naktis* – 17, *knyga* – 16, *namuose* – 16, *mokytoja* – 15 ir kt.; tikriniai vardai: *Lietuvos* – 285, *Klaipėda* – 132, *Kaunas* – 91, *Vilniaus* – 86, *Kauno* – 68, *Vytauto* – 59, *Europos* – 43, *Rasa* – 43, *Vilnius* – 40, *Lietuvoje* – 32, *Jonas* – 28, *Praninskienė* – 28, *Jono* – 22, *Kolbergas* – 20, *Panevėžys* – 17, *Birutė* – 16, *Vytautas* – 16, *Juozas* – 15, *Laima* – 15, *Marijampolė* – 15, *Valdas* – 15 ir kt.

Sudarytame vardo *AUŠRA* dažniniame žodynyje, kaip ir ankstesniu atveju, pirmose rangų pozicijose yra kai kurie jungtukai, prielinksniai, įvardžiai,rieveiksmiai, veiksmazodis *būti*.

Vardas *AUŠRA* dažniausiai buvo vartojamas periodikoje (šalies – 1127 kartus, vietinėje – 712, populiariojoje – 312, specializuotoje – 206); be to, grožinėse knygose – 60, negrožinėje literatūroje – 274, memuaruose – 28, LR valstybiniuose dokumentuose – 41, filosofinės literatūros vertimuose – 8. Ttai sudarė: a) 0,003 proc. VDU tekstyno visumos [2768 x 100 : 102857327]; b) 7,38 proc. nuo moterų lietuviškų vardų [2768 x 100 : 37505]; c) 8,08 proc. nuo vyrų lietuviškų vardų [2768 x 100 : 34260]; d) 10,65 proc. nuo moterų lotyniškų vardų [2768 x 100 : 26000]; e) 10,22 proc. nuo vyrų lotyniškų vardų [2768 x 100 : 27087]; f) 6,54 proc. nuo šio vardo distribucijų visumos [2768 x 100 : 42329].

Vardo *AUŠRA* didelio dažnio žodžių seka (su indeksais ne mažesniais kaip 10) yra labai maža ir sudaro 5,11 proc. Retieji žodžiai (su dažnumo indeksais nuo 1 iki 4 imtinai) sudaro 88,3 proc.; iš to skaičiaus: a) vieną kartą pavartotų skirtingų leksemų rasta 62,22 proc.; b) dukart (su indeksu „2“) gerokai mažiau – 16,25 proc.; c) triskart – 6,27 proc.; d) keturis kartus pasikartojusių skirtingų leksemų – 3,56 proc. Šio vardo likusi žodynėlio dalis – 6,59 proc. – priskirtina vidurio, arba pereinamosios zonos leksemoms (su indeksais nuo 5 iki 9 imtinai).

Vardo *AUŠRA* distribucijos leksinę įvairovę atspindi mažo dažnumo pavyzdžiai, kurie teikiami su gretinamuoju indeksu iš DDRLKŽ (Grumadienė, 1998 – po teikiamą pavyzdžio antrasis skaičius): *aerodinaminis* – 1(1), *anotavimas* – 1(0), *argumentuotai* – 1(1), *auksingas* – 1(0), *auditorius* – 1(2), *baisokas* – 1(1), *banalybė* – 1(5), *blynelis* – 1(1), *choralas* – 1(4), *dainynas* – 1(0), *desovietizacija* – 1(1), *dienovidis* – 1(1), *diskoteka* – 1(3), *dresuoti* – 1(0), *dulkinti* – 1(0), *egzaltacija* – 1(3), *emancipacija* – 1(2), *engti* – 3(4), *familiarumas* – 1(1), *feminizmas* – 1(1), *folija* – 1(1), *garbusis* – 1(0), *gausingas* – 1(1), *glamonė* – 1(5), *griovelis* – 1(0), *gvazdikas* – 1(2), *hierarchas* – 1(0), *įkaitas* – 1(1), *internetinis* – 1(0), *įpratimas* – 1(3), *jubilatas* – 1(4), *kapela* – 1(1), *kaukšmas* – 1(3), *keltas* – 1(1) ir kt.

Reto vartojimo žodžiai, susiję su vardo *AUŠRA* distribucijomis, sudaro didžiausią leksikos įvairovės koeficientą (žodynėlio dalį); jie priskirtini prie kalbos leksinės sistemos periferijos. Sudarytoje

leksikos sekoje daug tarptautinių žodžių, vedinių, arba derivatų. Kyla mintis, kodėl DDRLKŽ, kaip lietuvių kalbos leksinės sistemos modelyje, mažai arba visai neatsispindi, tarkim, deminutyvai, įvardžiuotiniai būdvardžiai, žargonas, kaip šnekamosios kalbos atmaina, ir pan. Pažymėtina, kad VDU Tekstyne tokie pavyzdžiai yra teikiami, ir tai kuria objektyvesnę vardų distribucijos leksinės sudėties vaizdą. DDRLKŽ nėra, pavyzdžiui, tokių žodžių: *griovelis*, *dainynas*, *dulkinti*, *atboginti*, *internetinis*, *kalendra*, *garbusis* ir pan.

Straipsnio apimtis neleidžia apžvelgti visų lietuviškos kilmės moterų bei vyrų vardų distribucijų leksikos matematinės analizės rezultatų. Todėl toliau teikiame kelių lotyniškų moterų ir vyrų vardų distribucijų leksikos gretinamojo tyrimo duomenis, kad būtų akivaizdesnės jų panašybės ir skirtybės. Lingvostatistikos pagrindu sudarytose distribucijų imtyse lotyniški, kaip ir anksčiau išnagrinėti lietuviški, vardai tarsi leksinė ašis dažnumų ranguose visad užimdavo pačias aukščiausias pozicijas, t. y. buvo žymimi didžiausiais indeksais, pavyzdžiui (vardo dažnumas nurodomas pirmuoju skaičiumi, o skliaustuose – dažnio rangas): *Akvilė* – 233(3), *Diana* – 759(4), *Kristina* – 1222(4), *Patricija* – 55(2), *Dominykas* – 206(2), *Donatas* – 916(2), *Modestas* – 318(3), *Paulius* – 1469(2) ir pan. Tačiau lotyniškų vardų distribucijų imtys bei leksinės charakteristikos buvo gerokai mažesnės negu lietuviškųjų. Tarkim, vardo *Akvilė* imtyje buvo iš viso 8887 leksemos, o skirtingų – 3159; mažiausieji dažnumų indeksai sudarė didžiausiąją distribucijos dalį („1“ pažymėta 2086 leksemos, arba 66,03 proc.; „2“ – 526, arba 16,65 proc.; „3“ – 166, arba 6,25 proc.; „4“ – 106, arba 3,36 proc.); vardo *Diana* imtyje buvo iš viso 32388 leksemos, o skirtingų – 9300; mažiausieji dažnumų indeksai sudarė didžiausiąją distribucijos dalį („1“ pažymėta 5987 leksemos, arba 64,38 proc.; „2“ – 1437, arba 16,45 proc.; „3“ – 591, arba 6,35 proc.; „4“ – 317, arba 3,41 proc.); vardo *Patricija* imtyje buvo iš viso 2572 leksemos, o skirtingų – 1451; mažiausieji dažnumų indeksai sudarė didžiausiąją distribucijos dalį („1“ pažymėta 1151 leksema, arba 79,32 proc.; „2“ – 160, arba 11,03 proc.; „3“ – 54, arba 3,72 proc.; „4“ – 19, arba 1,31 proc.); vardo *Dominykas* imtyje buvo iš viso 9044 leksemos, o skirtingų – 3640; mažiausieji dažnumų indeksai sudarė didžiausiąją distribucijos dalį („1“ pažymėta 2512 leksemų, arba 69,01 proc.; „2“ – 585, arba 16,07 proc.; „3“ – 198, arba 5,43 proc.; „4“ – 100, arba 2,75 proc.); vardo *Dominykas* imtyje buvo iš viso 9044 leksemos, o skirtingų – 3640; mažiausieji dažnumų indeksai sudarė didžiausiąją distribucijos dalį („1“ pažymėta 2512 leksemų, arba 69,01 proc.; „2“ – 585, arba 16,07 proc.; „3“ – 198, arba 5,43 proc.; „4“ – 100, arba 2,75 proc.) ir t. t.

Pažymėsime, jog lietuviški vardai buvo pa-

grindas sudaryti gerokai didesnes distribucijų imtis, negu tai galima buvo gauti lotyniškųjų vardų pagrindu, pavyzdžiui: 1) pagal vienaskaitos vardininko pavartojimų suminius dažnumus lietuvių vardų kilmės distribucijų imtys buvo 1,92 karto didesnės negu lotyniškųjų vardų; 2) pagal sudarytų distribucijų leksikos puslapių skaičių – 1,62 karto didesnės; 3) pagal iš viso distribucijose pavartotų žodžių skaičių – 1,55 karto didesnės; 4) pagal eilučių skaičių – 1,52 karto didesnės negu lotyniškųjų vardų. Taigi statistinės analizės rezultatai akivaizdžiai iliustruoja, jog beveik visi lietuviškieji vardai tekstuose vartojami gerokai intensyviau negu lotyniškieji.

1 lentelė. *Vardų distribucijų leksikos gretinamojo tyrimo rezultatai*

LIETUVIŠKI MOTERŲ IR VYRŲ VARDAI						
Vardas	Dažnis	Santykinis skaičius	Iš viso žodžių distribucijose	Santykinis skaičius	Skirtingų žodžių skaičius	Santykinis skaičius
Audronė	807	0,06046756	33628	0,07204636	9884	0,061537
Aušra	1394	0,10445077	12910	0,02765905	52918	0,329465
Birutė	1199	0,08983965	48955	0,10488372	12412	0,077277
Eglė	1429	0,10707328	55786	0,11951881	13477	0,083907
Jūratė	1238	0,09276188	52655	0,11281079	13656	0,085022
Algirdas	1746	0,13082572	69252	0,14836906	14548	0,090575
Gediminas	1459	0,10932114	59496	0,1274673	13497	0,084032
Giedrius	858	0,06428893	35070	0,07513578	9461	0,058904
Mindaugas	1330	0,09965533	27000	0,05784619	4921	0,030638
Vytautas	1886	0,14131575	72003	0,15426294	15844	0,098644
<i>Iš viso →</i>	13346	1	466755	1	160618	1
LOTYNIŠKI MOTERŲ IR VYRŲ VARDAI						
Akvilė	233	0,03358801	8887	0,03015312	3159	0,039675
Diana	759	0,10941329	32388	0,10989078	9300	0,116803
Kristina	1222	0,17615684	52925	0,17957174	12936	0,16247
Patricija	55	0,0079285	2572	0,00872666	1451	0,018224
Viktorija	1085	0,15640767	44910	0,15237727	11514	0,14461
Dominykas	206	0,02969583	9044	0,03068582	3640	0,045717
Donatas	916	0,13204555	38133	0,12938326	10051	0,126236
Modestas	318	0,04584114	13472	0,04570979	4499	0,056505
Paulius	1469	0,21176301	63298	0,21476679	14617	0,183582
Remigijus	674	0,09716016	29100	0,09873477	8454	0,106178
<i>Iš viso →</i>	6937	1	294729	1	79621	1

Iš lentelės duomenų matyti, jog lietuviški vardai vartojami daug intensyviau negu lotyniškai skolinėti vardai. Todėl pastarųjų distribucijos leksikos apimtys daugeliu atvejų reiškiamos mažesniais suminiais skaičiais. Intuityviai sunkoka nusakyti, koks konkretus vardas vartojamas dažniau arba rečiau. Remiantis statistine analize nustatyta, kad raštuose dažniausiai aptinkami lietuviškos kilmės vardai yra *Vytautas, Algirdas, Gediminas, Mindaugas, Eglė, Aušra, Jūratė*. Tačiau taip pat labai dažni kai kurie seniai vartojami skolininiai lotynų kalbos vardai: *Paulius, Kristina, Viktorija, Donatas, Diana*. Retai vartojami *Patricija, Dominykas, Akvilė, Modestas*.

Lentelėje teikiamų lietuviškųjų ir lotyniškųjų vardų sekos – tai tik labai trumpa jų apžvalgė.

Gali kilti klausimas, kodėl intuityviai akivaizdūs lietuviškų ir lotyniškų vardų vartojimo ar paplitimo lietuvių kalboje faktai turėtų būti aptariamai remiantis lingvistinės analizės rezultatais. Mūsų nuomone, tai daryti prasminga, kadangi kai kurie tariamai akivaizdūs kalbotyros faktai nebūtinai charakteringi kalbos leksinės sistemos visumai. Ir atvirkščiai, kai kurias kalbos faktų panašybes ir skirtis galima prognozuoti ir konstatuoti remiantis statistikos rezultatais.

Gretinamieji distribucijų leksikos statistinės analizės rezultatai pateikiami lentelėje (žr 1 lent.).

Kitaip tariant, čia bandyta parodyti, kaipremiantis lingvistine statistika galima apibūdinti įvairių vardų vartojimo konstatuojamąjį bei santykinį intensyvumą. Jeigu vardynas būtų praplėstas iki pasirinktinai nustatytų ribų, tai ir pavyktų gauti gana tikslų ne tik realų, arba faktinį, vaizdą, bet ir tų vardų santykinius dydžius, kurie atspindėtų vartojimo paplitimą.

Išvados

1. Lietuviškos bei lotyniškos kilmės vardų dažnis ar retesnis vartojimas lietuvių rašomojoje kalboje turi rezultatų naujumo bruožų, nes su vardu susietoje distribucijos leksemų visumoje tiesioginių tyrimų nerasta.

2. Distribucijų leksinės sudėties ašis yra pasirinktas vardas, kuris suponuoja žodžių atrankos nekategoriškas, kartais visiškai laisvas ribas. Pavyzdžiui, jeigu teikiamas vardas *Šarūnas*, tai pastarojo dešimtmečio tekstuose ir, suprantama, tekstynų kai kuriose dalyse (tarkim, periodikoje) labai tikėtina, jog bus pavartotos pavardės *Jasikevičius, Marčiulionis*; leksemos *krepšinio, lenta, komandinis, žaidimas* ir pan.
3. Vardų distribucijos tyrimas statistikos metodais padeda nustatyti leksemų vietą kalbos leksinėje sistemoje (tarkim, santykinį centrą ir periferiją). Tokie duomenys gali turėti teorinę ir taikomąją reikšmę, pavyzdžiui, struktūrinėje kalbotyroje, lingvodidaktikoje ir kt.
4. Tyrimo rezultatai rodo, kokie būna dažniausi ir rečiausi pastovūs junginiai, arba kolokacijos, vardų distribucijose.
5. Lingvostatistinė vardų distribucijų analizė parodė, jog šaltinių medžiaga turėtų būti nuolatos atnaujinama (ne rečiau kaip kas 5–10 metų): mat leksikos dinamika nuolat kinta; atsinaujina ir / arba gausėja profesinė terminija, ypač šnekamosios ir kompiuterinės kalbos elementai.
3. Korsakas, J., Nausėda, A., 2007. *Žodis mokymo procese: lingvostatistikos aspektas*, Verbum: pogranicza kultur: Rocznik naukowy. Białystok: Studium Hotelarstwa i Obsługi Turystyki w Białymstoku. P. 165–166.
4. Kuzavinis K., Savukynas B., 1998 *Lietuvių vardų kilmės žodynas*, 5-asis pataisytas ir papildytas leidimas. Vilnius: Mokslo ir enciklopedijų leidybos institutas.
5. *LKE 1993 – Lietuvių kalbos enciklopedija = Lithuanian Language Encyclopedia*. Vilnius: Mokslo ir enciklopedijų leidybos institutas.
6. Marcinkevičienė R., 2001, *The Importance of Syntagmatic Dimension in the Multilingual Lexical Database*. *International Journal of Corpus Linguistics*. 2001. Special Issue. P. 55–65.
7. McArthur, T., McArthur, R., 1998, *The Concise Oxford Companion to the English language*. Oxford University Press. P. 242.
8. Mikelionienė, J., 2000, Šiuolaikiniai metodai kalbos naujovėms tirti. *Darbai ir dienos*. Nr. 24. P. 65–73.
9. Rimkutė, E. 2006, Lietuvių kalbos kolokacijų žodynas: sandara ir paskirtis. *Prace Baltystyczne*. 2006. Nr. 3. P. 249–258.
10. Rudzkienė, V., 2005, *Socialinė statistika*. Vilnius: Mykolo Riomerio universitetas.
11. Sinclair, J., 2000, Lexical Grammar. *Darbai ir dienos*. Nr. 24, P. 191–203.
12. Tognini Bonelli, E., 2000, Corpus Classroom Currency. *Darbai ir dienos*. Nr. 24, P. 205–243.
13. Utka, A., 2005, Labai dažnų lietuvių kalbos žodžių ir žodžių formų ypatybės. *Lituanistica* 1(61), P. 48–55.
14. *DLKT – Dabartinės lietuvių kalbos tekstynas*. Elektroninis išteklius. Prieiga internete: <http://donelaitis.vdu.lt/main.php?id=4&nr=1_1_2> [žiūrėta 2010-04-10].

Literatūra

1. Grumadienė J., Žilinskienė V., 1998, *Dažninis dabartinės rašomosios lietuvių kalbos žodynas (abėcėlės tvarka)*, Vilnius: Lietuvių kalbos institutas, Matematikos ir informatikos institutas.
2. Korsakas J., 1998, *Definicijų leksika*. Šiauliai : Šiaulių universitetas.

RESEARCH ON DISTRIBUTION OF PROPER NAMES USING STATISTICAL METHODS

Ina Kližūnaitė, Aurimas Nausėda

Summary

In texts and corpora names as the linguistic units have distributions – the aggregate of related lexemes. Frequent and rare proper names of Latin and Lithuanian origin used in a written language are reviewed in the article. Distribution was limited by length of sentences (up to 300 characters, spaces included) and their aggregate parameters estimated statistically according to sampling (frequency of usage of proper names). The names of Latin origin were made Lithuanian. The lexical composition of all names was compared to distributions of more frequent and rare Lithuanian proper names. The results of statistical analysis describe usage of proper names: 1) on the basis of more widespread Lithuanian and Latin origin proper names, reasonably large distribution samplings were formed; 2) samplings of distributions of rarely used proper names were considerably smaller. Here names have been as conditionally thematic words, which reflect notional links to Lithuanian realia. Statistical analysis of distributions has shown that frequency of usage of Lithuanian names is often related to regular use of some lexemes. For example, with proper names *Audronė, Birutė, Eglė, Jūratė; Algirdas, Giedrius, Mindaugas, Vytautas* regularly and frequently words *buvo, yra, Lietuva, Vilnius, lietuvių, gyvenimas, galėti*, etc. are related. However, many accidental and rare use examples conditioned by time, fashion and other factors exist. Suppose, based on computer software not large distributions' samplings have been formed with some borrowed proper names (*Akvilė, Kristina, Patricija; Dominykas, Modestas, Remigijus*). The research data are given in the form of a table and a line chart.

Keywords: anthroponymics, frequency, vocabulary, distribution, lexis, lexicography, statistical analysis, corpus, proper names, list of proper names, Lithuanian and borrowed proper names.

VARDŲ DISTRIBUCIJOS TYRIMAS STATISTIKOS METODAIS

Ina Kliūnaitė, Aurimas Nausėda

Santrauka

Tekstuose ir tekstynuose vardai, kaip kalbiniai vienetai, turi distribucijas – susijusių leksemų visumą. Straipsnyje apžvelgiami lotyniškos ir lietuviškos kilmės dažnesni bei retesni vardai, pavartoti rašomojoje kalboje. Distribucija buvo ribojama sakinių ilgiu (iki 300 ženklų su tarpais), o jų visumos parametrai nustatomi statistiškai – pagal imtį (vardų vartojimo dažnumą). Lotynų kilmės vardai sulietuvinti. Visų vardų aplinkos leksinė sudėtis gretinama su dažnesnių bei retesnių lietuviškų vardų distribucijomis. Statistinės analizės rezultatai apibūdina vardų vartojimo ypatumus: 1) labiau paplitusių lietuviškos bei lotyniškos kilmės vardų pamatu susidarydavo gerokai didesnės distribucijos imtys; 2) rečiau vartojamų vardų distribucijos imtys buvo daug mažesnės. Vardai čia buvo tarsi sąlyginiai teminiai žodžiai, kurie atspindi reikšmines sąsajas su Lietuvos realijomis. Distribucijų statistinė analizė parodė, jog lietuviškų vardų dažnis vartojimas neretai sietinas su dėsningu kai kurių leksemų vartojimu. Pavyzdžiui, su vardais Audronė, Birutė, Eglė, Jūratė, Algirdas, Giedrius, Mindaugas, Vytautas sistemingai ir dažnai vartojami tokie žodžiai: buvo, yra, Lietuva, Vilnius, lietuvių, gyvenimas, galėti ir pan. Tačiau esama nemažai atsitiktinai ir retai vartojamų pavyzdžių, kurie sąlygojami laikmečio, madų bei kitokių faktorių. Tarkim, kompiuterinės programos pagrindu ne itin didelės distribucijų imtys buvo sudarytos su kai kuriais skolintiniais vardais (Akvilė, Kristina, Patricija; Dominykas, Modestas, Remigijus). Tyrinėjimo duomenys pateikiami statistinėje lentelėje ir grafike.

Prasminiai žodžiai: antroponimika, dažninis žodynas, distribucija, leksika, leksikografija, statistinė analizė, tekstynas, tikriniai žodžiai, vardynas, lietuviški ir skolintiniai vardai.

Įteikta 2010 06 02