

## KALBOS DALIŲ STATISTINĖ CHARAKTERISTIKA LIETUVIŲ PUBLICISTIKOJE

VIDA ŽILINSKIENĖ

Pastaraisiais dešimtmečiais kalbotyroje imta plačiai taikyti statistinio tyrimo metodiką, kuri padėjo atskleisti naujų kalbos funkcionavimo aspektų. Pavyzdžiui, paaiškėjo, kad įvairiuose latvių kalbos funkciniuose stiliuose gramatiniai vienetai (kalbos dalys) yra pasiskirstę pagal Gauso (normalųjį) dėsnį<sup>1</sup>, o publicistikos stiliaus leksiniams vienetais (atskiroms žodžių formoms, paimitoms iš įvairių dažnumų žodyno zonų) ir darybiniams vienetais (daiktavardžių ir būdvardžių įvairioms priesagoms) normaliojo pasiskirstymo hipotezė turi būti atmetama<sup>2</sup>. Patyrinėjus daiktavardžių ir būdvardžių pasiskirstymo dėsningumus Puškino prozoje ir veiksmažodžių – ispanų kalbos matematiniuose tekstuose<sup>3</sup>, daromos tik preliminarinės išvados (dėl mažos nagrinėtų tekstų apimties), kad daiktavardžiai ir veiksmažodžiai yra išsidėstę pagal normalųjį pasiskirstymo dėsnį, o būdvardžių pasiskirstymo tikrinimas davė beveik vienodus rezultatus tiek normaliajam dėsniai, tiek Puasono dėsniai. Anglų kalbos moksliniuose techniniuose tekstuose apie laivų mechanizmus žodžių formos, išrinktos iš įvairių dažnumų žodyno zonų, pasiskirstę pagal skirtingus dėsnius: žodyno pradžioje esančių žodžių formos pasiskirsčiusios pagal normalųjį dėsnį, toliau eina mišri zona, kurios pradžioje vyrauja normalusis dėsnis, gale – Puasono, o rečiausios žodžių formos pasiskirsčiusios pagal Puasono dėsnį. Tarnybiniai žodžiai (artikeliai, prielinksniai, dalelytės, jungtukai) šiuose tekstuose pasiskirstę tik pagal normalųjį dėsnį<sup>4</sup>. Yra ir nuomonė, kad leksinių vienetų pasiskirstymo aprašymui Gauso ir Puasono dėsniai esą per daug grubūs, juos naudojant daugelio žodžių formų pasiskirstymas pasilieka neaiškus, ir siūloma remtis Markovo—Kolmogorovo pasiskirstymu (priartėjančiu prie Pirsono ribinių kreivių), leidžiančiu vienareikšmiškai aprašyti bet kokių leksemų pasiskirstymą<sup>5</sup>.

<sup>1</sup> Якубайтис Т. А. О математико-статистической модели распределения классов слов. — В кн.: Известия АН ЛатвССР 1968, № 7, с. 46; Ее же. Сравнение вида распределения языковых единиц разных уровней. — В кн.: Математические методы в языкознании. Рига, 1969, с. 59—66; Ее же. Статистическая характеристика частей речи в латышском языке. — В кн.: Статистика и функциональные стили языка. Рига, 1977, с. 7—28.

<sup>2</sup> Токарев В. П., Якубайтис Т. А. Математико-статистическая модель распределения лексем. — В кн.: Математические методы в языкознании, с. 7—21. Сика Н. Я., Якубайтис Т. А. Математико-статистическая модель распределения словообразовательных элементов. — В кн.: Математические методы в языкознании, с. 47—51.

<sup>3</sup> Фрумкина Р. М. О законах распределения слов и классов слов. — В кн.: Структурно-типологические исследования. М., 1962, с. 124—133.

<sup>4</sup> Бектаев К. Б., Лукьяненко К. Ф. О законах распределения единиц письменной речи. — В кн.: Статистика речи и автоматический анализ текста. Л., 1971, с. 47—112.

<sup>5</sup> Каширина М. Е. О типах распределения лексических единиц в тексте. — В кн.: Статистика речи и автоматический анализ текста. Л., 1974, с. 335—360.

Taigi įvairių lingvistinių vienetų pasiskirstymas tekste galutinai neišaiškintas, ir jis tebėra viena iš pagrindinių problemų tiek teorinėje, tiek taikomojoje matematinėje lingvistikoje.

Minėtieji dalykai visiškai netyrinėti lietuvių kalboje, todėl nežinoma lietuvių kalbos gramatinių, leksinių, darybinių ir t. t. vienetų statistinės savybės, kurios atspindėtų jų realią vartoseną ne tik literatūrinėje kalboje apskritai, bet ir atskiruose funkcinuose stiliuose.

Šiame straipsnyje ir norima parodyti atskirų kalbos dalių vartoseną lietuvių publicistikos, t. y. išaiškinti, pagal kurį statistinį pasiskirstymo dėsnį jos yra išsidėsčiusios bet kokiame to stiliaus tekste ir kokie yra pasiskirstymo dėsnio parametrai, būdingi kiekvienai konkrečiai kalbos daliai.

Kalbos dalių statistinė charakteristika buvo nustatinėjama pagal lietuvių kalbos publicistikos dažnumų žodyno medžiagą. Kaip jau minėta<sup>6</sup>, sudarinėjant šį žodyną, dėl publicistikos kalbos įvairumo ir daugiatemiškumo visa tyrinėjama medžiaga buvo suskirstyta penkiomis vienodo ilgio teminėms grupėms (po 60 000 pavartotų žodžių): A – vidaus ir užsienio politika, B – mokslas, technika, pramonė, žemės ūkis, C – švietimas, pedagogika, moralinės-ėtinės problemos, D – dailė, teatras, kinas, literatūros mokslas, E – sportas, skelbimai ir straipsniai, nepatekę į pirmąsias keturias grupes. Iš kiekvieno laikraščio ar žurnalo atsitiktinai imta viena 1000 žodžių praba ir suskaičiuota, kiek ir kokių kalbos dalių yra kiekvienoje praboje<sup>7</sup>. Kadangi prabų parinkimas buvo atsitiktinis, tai kiekvienoje praboje esančių atskirų kalbos dalių dažnumas irgi yra atsitiktinis dydis.

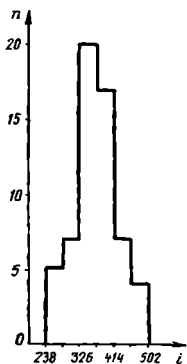
Iš atskirų prabų duomenų kiekvienai kalbos daliai<sup>8</sup> buvo sudarytos penkios aibės (penkioms teminėms grupėms) po 60 reikšmių kiekvienoje (vieną temą sudarė 60 prabų). Šios atsitiktinių dydžių aibės kaip pradiniai duomenys įvesti į elektroninę skaičiavimo mašiną „Minsk-22“. ESM kiekvienai aibei sudarė variacinę eilutę, suskirstė ją į intervalus, pagal kurių reikšmės nubrėžė histogramą (iš histogramos pobūdžio galima apytikriai nustatyti atsitiktinių dydžių empirinį pasiskirstymą) ir apskaičiavo pagrindinius eilutės parametrus: aritmetinį vidurkį, standartinį nukrypimą, variacijos koeficientą, asimetrijos koeficientą ir ekscesą.

Gautų histogramų pobūdis leidžia daryti prielaidą, kad kalbos dalys tekstuose yra pasiskirsčiusios pagal normalųjį (Gauso) dėsnį. Teorinė Gauso kreivė yra griežtai simetriška, tuo tarpu praktikoje, kur susiduriama su atsitiktinėmis prabomis, kreivėms būdingas tam tikras asimetriškumas, aiškiai pastebimas nagrinėjamose histogramose. Iš keturiasdešimties histogramų trisdešimt trims būdinga teigiamoji (dešininė) asimetrija ir tik septynioms – neigiamoji (kairinė), būtent: C temos daiktavardžiams, A – temos veiksmožodžiams, A ir B temų prievėksmams, B temos dalelytėms ir A temos prielinksniams. 1–4 brėžiniuose pateikti histogramų pavyzdžiai: B temos daiktavardžių, D temos būdvardžių, E temos veiksmožodžių

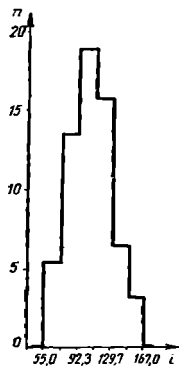
<sup>6</sup> Žilinskienė V. Lietuvių kalbos raidžių dažnumas publicistikos tekstuose. – Kalbotyra, 1977, t. XXIX (1), p. 83–95.

<sup>7</sup> Į prabos daiktavardžių kiekį neįeina tikriniai daiktavardžiai.

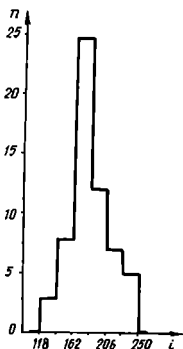
<sup>8</sup> Iš visų kalbos dalių, klasifikuotų pagal akademinę „Lietuvių kalbos gramatiką“, V., 1965, t. I; 1971, t. 2, statistškai netyrinėjami skaitvardžiai, jaustukai ir išiktukai. Pastarieji du netyrinėjami todėl, kad dėl returno jiems būtų nepakankama išnagrinėtos medžiagos apimtis (praboje jų arba visiškai nebuvo, arba pasitaikė tik vienas kitas). Skaitvardžiai netyrinėjami todėl, kad jie tekstuose užrašomi kartais skaičiais, kurie sudarinėjant dažnumų žodyną praleisti, kartais – žodžiais, kurie fiksuoti žodyne. Taigi skaitvardžių kiekis žodyne (kaip ir kitų kalbų dažnumų žodynuose) nepakankamai tiksliai parodo realią skaitvardžių vartoseną kalboje.



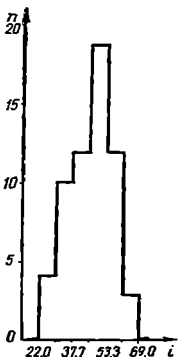
1 brėž. B temos daiktavardžių pasiskirstymo histograma.  
x-sų ašyje atidėtas daiktavardžių dažnumas,  
y-kų ašyje – reikšmių skaičius intervale.



2 brėž. D temos būdvardžių pasiskirstymo histograma.  
x-sų ašyje atidėtas būdvardžių dažnumas,  
y-kų ašyje – reikšmių skaičius intervale.



3 brėž. E temos veiksmažodžių pasiskirstymo histograma.  
x-sų ašyje atidėtas veiksmažodžių dažnumas,  
y-kų ašyje – reikšmių skaičius intervale.



4 brėž. A temos prielinksnių pasiskirstymo histograma.  
x-sų ašyje atidėtas prielinksnių dažnumas,  
y-kų ašyje – reikšmių skaičius intervale.

ir A temos prielinksnių. Tačiau esminiam pasiskirstymo pobūdžiui (jo normališkumui) parodyti vien histogramos grafinio vaizdo nepakanka, kadangi jos pobūdis priklauso nuo parenkamų intervalų pločio bei ribų. Todėl buvo apskaičiuoti minėtieji empiriniai momentai<sup>9</sup>, t. y.:

<sup>9</sup> Смирнов Н. В., Душин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. М., 1968, с. 82, 90, 91, 93.

aritmetinis vidurkis:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i,$$

standartinis nukrypimas:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2},$$

asimetrijos koeficientas:

$$A = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{N\sigma^3},$$

variacijos koeficientas:

$$V = \frac{\sigma}{\bar{x}},$$

ekscesas:

$$E = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{N\sigma^4} - 3,$$

kur  $n$  – variacinės eilutės reikšmių skaičius,  
 $x_i$  – tiriamojo parametro reikšmės.

Normaliojo pasiskirstymo atveju asimetrijos koeficiento skaitinė reikšmė turi būti artima nuliui, o koeficiento ženklas turi parodyti, kuriam kreivės šlaitui (kairiajam ar dešiniajam) būdinga asimetrija. Kaip matyti iš 1 lentelės, tik septyni asimetrijos koeficientai yra neigiami ir jie priklauso toms kalbos dalims, kurių histogramose pastebima kairinė asimetrija. Iš visų skaitinių asimetrijos koeficiento reikšmių tik kelios yra tokios, kurias būtų galima laikyti artimomis nuliui, o visos kitos reikšmės yra gerokai didesnės. Todėl reikėtų dar nustatyti ribas, kuriose lingvistinei variacinei eilutei aprašyti dar tikėtų normalusis pasiskirstymo dėsnis<sup>10</sup>. Tuo tikslu buvo apskaičiuotas asimetrijos vidutinis kvadratinis nukrypimas –  $\sigma(r_3)$ :

$$\sigma(r_3) = \sqrt{\frac{6(N-1)}{(N+1)(N+3)}},$$

parodantis empirinio pasiskirstymo kreivės sumarinį pasislinkimą teorinės kreivės centro atžvilgiu – gauta  $\sigma(r_3) = 0,30$ . Tada, norint vertinti empirinio pasiskirstymo ir normaliojo dėsnio artimumą, apskaičiuojamas koeficientas  $\alpha$ , lygus asimetrijos koeficiento ir asimetrijos vidutinio kvadratinio nukrypimo santykiui. Jeigu šis koeficientas gaunamas mažesnis už 3, tai empirinės kreivės pasislinkimą normaliojo dėsnio centro atžvilgiu galima laikyti neesminiu ir teigti, kad nagrinėjamos eilutės variantės yra pasiskirsčiusios pagal normalųjį dėsnį. Koeficiento  $\alpha$  reikšmės

<sup>10</sup> Пиотровский Р. Г., Бектаев К. Б., Пиотровская А. А. Математическая лингвистика. М., 1977, с. 260–261.

## Pasiskirstymo dėsnio empiriniai momentai

Tema	Vidurkis	Standartinis nukrypimas	Variacijos koeficientas	Asimetrijos koeficientas	Koeficientas $\alpha$	Ekscesas	Koeficientas $\beta$
<b>Daiktavardžiai</b>							
A	382,6	41,5	10,9	0,39	1,30	-0,23	-0,39
B	368,5	54,9	14,9	0,08	0,07	0,01	0,01
C	347,9	68,3	19,6	-0,10	-0,33	-0,17	-0,29
D	344,8	50,7	14,7	0,44	1,47	-0,24	-0,42
E	355,4	43,2	12,2	0,18	0,60	0,45	0,78
<b>Būdvardžiai</b>							
A	97,8	32,8	33,5	0,94	3,13	0,96	1,66
B	84,5	23,9	28,3	1,28	4,60	1,69	2,91
C	86,8	29,6	34,1	0,17	0,57	-0,91	-1,56
D	105,5	24,1	22,9	0,25	0,83	-0,20	-0,34
E	77,4	18,1	23,4	0,38	1,27	0,01	0,01
<b>Veiksmažodžiai</b>							
A	173,9	26,2	15,0	-0,44	-1,47	-0,78	-1,34
B	194,6	28,4	14,6	0,04	0,13	-0,09	-0,15
C	205,9	35,7	17,3	0,27	0,90	-0,39	-0,57
D	173,2	23,1	13,4	0,09	0,30	-0,49	-0,85
E	183,6	27,0	14,7	0,31	1,03	-0,02	-0,01
<b>Prieveiksmiai</b>							
A	52,2	12,1	23,2	-0,13	-0,43	-0,58	-1,00
B	72,0	18,4	25,5	-0,06	-0,20	0,39	0,67
C	71,8	17,7	24,6	0,51	1,70	0,19	0,32
D	62,8	18,3	29,1	0,48	1,60	-0,61	-1,05
E	59,9	12,2	20,4	0,82	2,73	0,58	1,00
<b>Įvardžiai</b>							
A	80,2	21,6	26,9	0,94	3,13	1,68	2,89
B	81,5	21,0	25,8	-0,32	-1,07	-0,02	-0,04
C	98,2	30,0	30,6	0,48	1,60	0,38	0,66
D	94,9	22,4	23,5	0,62	2,07	-0,24	-0,41
E	85,1	20,5	24,1	0,52	1,73	-0,14	-0,24

1 lentelė (tęsinys)

Tema	Vidurkis	Standartinis nukrypimas	Variacijos koeficientas	Asimetrijos koeficientas	Koeficientas $\alpha$	Ekscesas	Koeficientas $\beta$
<b>Dalelytės</b>							
A	16,6	8,0	48,3	1,12	3,73	1,11	1,91
B	29,5	14,1	47,9	0,12	0,40	-0,49	-0,84
C	34,2	12,1	35,5	-0,03	-0,10	-0,09	-0,16
D	32,9	11,7	35,5	0,72	2,40	0,52	0,89
E	29,4	11,1	37,7	0,43	1,43	0,10	0,17
<b>Jungtukai</b>							
A	60,8	11,9	19,6	0,28	0,93	0,07	0,12
B	65,9	15,5	23,5	0,20	0,67	-0,73	-1,26
C	72,8	16,5	22,6	0,56	1,87	0,07	0,12
D	69,9	15,4	22,1	0,37	1,23	-0,75	-1,29
E	66,4	16,3	24,5	0,82	2,73	0,54	0,93
<b>Prielinksniai</b>							
A	45,9	10,8	23,4	-0,33	-1,10	-0,32	-0,55
B	42,6	10,9	25,6	0,10	0,33	-0,42	0,72
C	43,3	12,4	28,7	0,40	1,33	-0,49	-0,84
D	41,5	8,9	21,5	0,16	0,53	-0,63	-1,09
E	48,9	12,0	24,6	0,14	0,47	0,37	0,64

pateikiamos 1 lentelėje. Kaip rodo duomenys, pasiskirstymo asimetriškumas yra leistinose ribose visų temų visoms kalbos dalims, išskyrus A ir B temų būdvardžius, A temos įvardžius ir dalelytes.

Kitai ribai nustatyti buvo apskaičiuotas dydis  $\sigma(E)$  – eksceso vidutinis kvadratinis nukrypimas:

$$\sigma(E) = \sqrt{\frac{24 N (N-2) (N-3)}{(N-1)^2 (N+3) (N+5)}}$$

Šiuo atveju gauta  $\sigma(E) = 0,58$ . Toliau empirinio ir teorinio dėsnų artimumo įvertinimui apskaičiuojamas koeficientas  $\beta$ , lygus eksceso ir dydžio  $\sigma(E)$  santykiui. Šio koeficiento reikšmės irgi surašytos 1 lentelėje. Normaliajam dėsnii keliami sąlyga, kad  $\beta < 3$ . Ši sąlyga patenkinama visoms minėtų temų kalbos dalims.

Iš to darytina išvada, kad galima formuoti vadinamąją nulinę hipotezę, teigiančią, kad tiriamuosiuose tekstuose kalbos dalys yra pasiskirsčiusios pagal normalųjį dėsnį. Ši hipotezė patvirtinama ar paneigiama  $\chi^2$  (Pirsono),  $\omega^2$  ir Kolmogoro-

## Pasiskirstymo kriterijų reikšmės

Tema	Laisvės laipsnių skaičius	Kriterijai			
		$\chi^2$ empirinis	$\chi^2$ ribinis	$\omega^2$	Kolmogorovo
<b>Daiktavardžiai</b>					
A	9	14,05	16,92	0,06	0,93
B	8	5,14	15,51	0,08	0,61
C	9	13,26	16,92	0,09	0,58
D	8	6,04	15,51	0,05	0,91
E	8	5,36	15,51	0,03	0,96
<b>Būdvardžiai</b>					
A	9	10,51	16,92	0,14	0,57
B	8	11,92	15,51	0,27	0,19
C	8	13,59	15,51	0,06	0,76
D	8	9,99	15,51	0,06	0,64
E	8	6,60	15,51	0,05	0,85
<b>Veiksmažodžiai</b>					
A	9	18,46	16,92	0,13	0,24
B	7	8,68	14,07	0,04	0,81
C	8	15,55	15,51	0,09	0,68
D	8	11,78	15,51	0,08	0,61
E	7	8,53	14,07	0,09	0,68
<b>Prieveiksmiai</b>					
A	8	22,18	15,51	0,03	0,94
B	8	8,72	15,51	0,06	0,89
C	8	5,23	15,51	0,08	0,79
D	8	15,04	15,51	0,13	0,28
E	8	8,45	15,51	0,11	0,64
<b>Įvardžiai</b>					
A	8	10,43	15,51	0,13	0,48
B	8	10,73	15,51	0,04	0,88
C	8	15,60	15,51	0,09	0,38
D	8	9,94	15,51	0,09	0,78
E	8	7,74	15,51	0,10	0,68

2 lentelė (tęsinys)

Tema	Laisvės laipsnių skaičius	Kriterijai			
		$\chi^2$ empirinis	$\chi^2$ ribinis	$\omega^2$	Kolmogorovo
<b>Dalelytės</b>					
A	7	14,55	14,07	0,28	0,16
B	8	7,18	15,51	0,03	0,98
C	7	6,03	14,07	0,03	0,99
D	7	6,68	14,07	0,09	0,66
E	7	7,39	14,07	0,11	0,56
<b>Jungtukai</b>					
A	7	8,30	14,07	0,07	0,68
B	8	6,36	15,51	0,04	0,96
C	9	12,58	16,92	0,05	0,89
D	8	5,64	15,51	0,08	0,71
E	8	7,54	15,51	0,12	0,39
<b>Prielinksniai</b>					
A	8	4,42	15,51	0,08	0,69
B	7	9,94	14,07	0,03	0,91
C	7	7,50	14,07	0,07	0,76
D	8	2,59	15,51	0,04	0,79
E	7	3,86	14,07	0,06	0,58

rovo kriterijais<sup>11</sup>, kurie palygina normaliojo pasiskirstymo kreivę su tyrinėjamoju empiriniu pasiskirstymu. Šie kriterijai skaičiuoti elektronine skaičiavimo mašina pagal universiteto Skaičiavimo centre paruoštas programas. Skaičiavimo rezultatai surašyti 2 lentelėje.

$\chi^2$  kriterijus leidžia priimti nulinę hipotezę, kai jo empirinė reikšmė neviršija ribinės, priklausančios nuo laisvės laipsnių skaičiaus ir priimtojo reikšmingumo lygmens (5%). Laisvės laipsniai ir ribinės reikšmės<sup>12</sup> irgi nurodomos 2 lentelėje. Palyginus empirinę ir ribinę  $\chi^2$  reikšmes, matyti, kad nulinę hipotezę galima priimti visais atvejais, išskyrus A temos veiksmožodžius, prieveiksnius, dalelytes ir C temos veiksmožodžius bei įvardžius.

$\omega^2$  kriterijus leidžia priimti nulinę hipotezę, kai jo empirinė reikšmė neviršija ribinės, kuri lygi 0,46. Skaičiavimo rezultatai rodo (žr. 2 lent.), kad visų temų atskirų kalbos dalių empirinės reikšmės yra mažesnės už ribinę.

<sup>11</sup> Смирнов Н. В., Дунин-Барковский И. В. Указ. соч., с. 268, 277; Митропольский А. К. Техника статистических вычислений. М., 1961, с. 241–244.

<sup>12</sup> Гмурман В. Е. Теория вероятностей и математическая статистика. М., 1972, с. 358.



Tema	Koeficientai $\alpha$ , $\beta$	$\chi^2$ kriter.	$\omega^2$ kriter.	Kolmogorovo kriterijus
<b>Daiktavardžiai</b>				
A	+, +	+	+	+
B	+, +	+	+	+
C	+, +	+	+	+
D	+, +	+	+	+
E	+, +	+	+	+
<b>Būdvardžiai</b>				
A	-, +	+	+	+
B	-, +	+	+	+
C	+, +	+	+	+
D	+, +	+	+	+
E	+, +	+	+	+
<b>Veiksmažodžiai</b>				
A	+, +	-	+	+
B	+, +	+	+	+
C	+, +	-	+	+
D	+, +	+	+	+
E	+, +	+	+	+
<b>Prieveiksmai</b>				
A	+, +	-	+	+
B	+, +	+	+	+
C	+, +	+	+	+
D	+, +	+	+	+
E	+, +	+	+	+
<b>Įvardžiai</b>				
A	-, +	+	+	+
B	+, +	+	+	+
C	+, +	-	+	+
D	+, +	+	+	+
E	+, +	+	+	+

Tema	Koeficientai $\alpha, \beta$	$\chi^2$ kriter.	$\omega^2$ kriter.	Kolmogorovo kriterijus
Dalelytės				
A	-, +	-	+	+
B	+, +	+	+	+
C	+, +	+	+	+
D	+, +	+	+	+
E	+, +	+	+	+
Jungtukai				
A	+, +	+	+	+
B	+, +	+	+	+
C	+, +	+	+	+
D	+, +	+	+	+
E	+, +	+	+	+
Prielinksniai				
A	+, +	+	+	+
B	+, +	+	+	+
C	+, +	+	+	+
D	+, +	+	+	+
E	+, +	+	+	+

Nulinės hipotezės patvirtinimui būtina, kad Kolmogorovo kriterijaus empirinė reikšmė būtų didesnė už 0,05, taigi ši sąlyga taip pat patenkinama visoms kalbos dalims (žr. 2 lent.).

Iš visų skaičiavimų rezultatų aiškiai matyti (žr. 3 lent.), kad jeigu vienas kuris iš taikytųjų metodų tam tikrai kalbos daliai ir rodo empirinio pasiskirstymo rezultatų neatitikimą normaliajam dėsniai, tai kiti metodai vis dėlto patvirtina gerą empirinio pasiskirstymo sutapimą su normaline kreive. Daiktavardžių, jungtukų ir prielinksnių pasiskirstymas ypač artimas normaliajam pasiskirstymui.

Taigi galima tvirtinti, kad, remiantis keliais metodais ( $\alpha$  ir  $\beta$  koeficientais,  $\chi^2$  [Pirsono],  $\omega^2$  ir Kolmogorovo kriterijais), nustatyta, kad lietuvių publicistikos įvairių temų tekstuose visoms kalbos dalims yra būdingas normalusis (Gauso) pasiskirstymas. Vadinas įvairių temų publicistikos tekstuose yra toks reguliarus atskirų kalbos dalių kiekis, kurį apibūdina žinomi normaliojo dėsnio parametrai: aritmetinis vidurkis, standartinis nukrypimas, asimetrijos koeficientas, ekscesas (žr. 1 lent.).

Įrodžius normaliojo pasiskirstymo dėsnio tikimą publicistikos tekstų gramatiniam vienetais, šios rūšies tekstų tyrinėjimams galima drąsiai vartoti Stjudento,

Fišerio, Romanovskio kriterijus, taip pat dispersinę analizę, t. y. tuos statistinius kriterijus, kurie remiasi normaliuoju dėsniu ir kurie leistų nustatyti, ar, pavyzdžiui, yra reikšmingas skirtumas tarp skirtingų tekstų, ar tuos skirtingų temų tekstus galima jungti į vieną visumą, ar kuris nors nagrinėjamas tekstas priklauso publicistikai, ar ne ir t. t. Kadangi publicistika yra vienas neutraliausių kalbos stilių, galima atsargiai spėti, kad panaši padėtis yra ir kitų stilių tekstuose. Todėl minėtuosius kriterijus ir dispersinę analizę (suprantama, su tam tikru rezervu) dabar bus galima taikyti kitiems statistiniams įvairių funkcinių stilių arba atskirų autorių kūrinių tyrinėjimams pagal kalbos dalių vartoseną (pavyzdžiui, nustatyti, ar kuris nors tekstas priklauso tam ar kitam funkciniam stiliui, išaiškinti tam tikrų kūrinių autorystę ir t. t.). Ypač tai svarbu stilistiniams literatūros kūrinių tyrinėjimams.

Iki šiol panašūs tyrinėjimai buvo atliekami, aprioriškai priimant normaliojo pasiskirstymo hipotezę, ir čia išdėstyti statistiniai skaičiavimai, atrodo, turėtų suteikti tvirtesni pagrindą šiai esminei prielaidai ir sustiprinti ja grindžiamų tyrinėjimų patikimumą.

## СТАТИСТИЧЕСКАЯ ХАРАКТЕРИСТИКА ЧАСТЕЙ РЕЧИ В ЛИТОВСКОЙ ПУБЛИЦИСТИКЕ

### *Резюме*

Цель данной работы — определение употребительности частей речи в литовской публицистике, т. е. выявление закона статистического распределения частей речи и его параметров.

Статистическая характеристика частей речи исследовалась по материалу частотного словаря литовской публицистики. Описана методика определения закона эмпирического распределения частей речи. Получены гистограммы распределения для каждой части речи, по виду которых делалось допущение, что части речи в текстах распределены по закону Гаусса (нормальному). На рис. 1—4 показаны некоторые гистограммы. Рассчитаны параметры эмпирического распределения: арифметическое среднее, стандартное отклонение, коэффициент вариации, коэффициент асимметрии и эксцесс. После анализа этих параметров была выдвинута гипотеза, утверждающая, что в исследованных текстах части речи распределены по нормальному закону. Проверка этой гипотезы осуществлена при помощи следующих критериев:  $\chi^2$  (Пирсона),  $\omega^2$  и Колмогорова. Установлено хорошее совпадение эмпирического и нормального законов распределения.

Полученные результаты показали, что в любом публицистическом тексте наблюдается строго определенное и стабильное количество отдельных частей речи, которое можно определить по параметрам нормального распределения: арифметической средней, стандартному отклонению, коэффициенту асимметрии и эксцессу.