

LIETUVIŲ KALBOS RAIDŽIŲ DAŽNUMAS PUBLICISTIKOS TEKSTUOSE

VIDA ŽILINSKIENĖ

Ligi šiol lietuvių kalbos raidžių dažnumas nėra išsamiai tyrinėtas. Todėl, sudarinėjant lietuvių kalbos publicistikos dažnumų žodyną, buvo skaičiuotas raidžių dažnumas.

Buvo nagrinėti visi respublikiniai laikraščiai ir bendrojo pobūdžio žurnalai, išspausdinti 1972–1973 metais. Žodyno apimtis – 300 000 žodžių formų. Visa tyrinėjama medžiaga suskirstyta penkiomis vienodo ilgio (po 60 000 žodžių formų) teminėmis grupėmis: 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 kiti straipsniai, nepatekę į pirmąsias keturias grupes. Iš kiekvieno laikraščio ar žurnalo atsitiktinai imta viena 1000 žodžių formų praba. Po kiekvieno žodžio buvo įrašytas atitinkamas indeksas, kurį sudarė raidė, nurodanti, kuriai kalbos daliai priskiriamas žodis, ir skaitmuo, nurodantis, ar žodis yra lietuviškos kilmės, ar skolinys (įskaitant vadinamuosius internacionalizmus)¹. Vietoj tikrinių daiktavardžių buvo užrašytos atitinkamos raidės, kurių nėra lietuviškoje abėcėlėje, ir aukščiau minėtas indeksas, pvz., visi vardai žymėti raide X, pavardės – XX, valstybių pavadinimai – W, vietovardžiai – Q ir t. t. Taip pasielgta todėl, kad rūpejo sužinoti tik bendrą atskirų tikrinių daiktavardžių (vardų, pavardžių, vietovardžių ir t. t.) skaičių. Taigi raidės, iš kurių sudaryti tikriniai daiktavardžiai, nebuvo skaičiuojamos.

Šitai paruoštas kiekvienas 1000 žodžių formų tekstas buvo perforuojamas. Kadangi perforatoriaus raidynas neatitinka lietuvių abėcėlės, vietoj kai kurių raidžių teko perforuoti du ženklus: 1) vietoj nosinių balsių perforuotos paprastos balsės ir šalia kabelis (pvz., *e* – *e*.)^{1/2} vietoj *ė*, *ū*, *š*, *ž* perforuota *e*, *u*, *s*, *z* ir šalia apostrofai (pvz., *ė* – *e'*, *š* – *s'*).

Žodynas skaičiuotas Vilniaus V. Kapsuko universiteto Skaičiavimo centre elektronine skaičiavimo mašina „Minsk-22“.

¹ Skoliniais laikomi žodžiai, nurodyti žodynuose: Fraenkel E. Litauisches etymologisches Wörterbuch. Heidelberg—Göttingen, 1955–1962, ir Tarptautinių žodžių žodynas. V., 1969.

Pirmiausia buvo skaičiuoti kiekvienos prabos dažnumų žodynai, o iš jų – vienos temos. Kiekvienos temos žodžiams skaičiuotas raidžių dažnumas, tam tikra raide prasidedančių žodžių ilgis ir tam tikroms kalbos dalims priskiriamų žodžių ilgis. Abiem atvejais domėtasi, ar šie parametrai priklauso nuo tekstų tematikos. Iš atskirų teminių grupių raidžių dažnumų nustatytas visos žodyno apimties raidžių dažnumas. Taip pat suskaičiuota, kiek yra viensiemenių, dviskiemenių, triskiemenių ir t. t. žodžių visame žodyne, atsižvelgiant į tai, kurioms kalbos dalims tie žodžiai atstovauja.

Skaičiavimo duomenimis, 55057-uose A grupės žodžiuose buvo 381256 raidės, 56515-oje B grupės žodžių – 375658 raidės, 57747-uose C grupės žodžiuose – 376985 raidės, 55411-oje D grupės žodžių – 364205 raidės ir 56146-uose E grupės žodžiuose – 365212 raidžių. Taigi visame žodyne buvo 280876 žodžiai, o juose – 1863159 raidės: iš jų 894772 balsės (48,03% visų raidžių) ir 968387 priebalsės (51,97%). Lietuvių abėcėlėje balsės sudaro 37,05%, o priebalsės – 62,95% visų raidžių. Palyginus šiuos duomenis, matyti, kad publicistikos tekstuose balsių vidutinis dažnumas didesnis, negu priebalsių.

Raidžių išsidėstymas pagal mažėjantį procentinį dažnumą atskirų teminių grupių tekstuose pateikiamas 1 lentelėje.

Kad būtų galima objektyviau spręsti apie raidžių eilės patikimumą, kiekvienai raidei apskaičiuotas variacijos koeficientas pagal formulę²:

$$v = \frac{100}{\bar{x}} \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

kur v – variacijos koeficientas, išreikštas procentais, \bar{x} – raidės vidutinis dažnumas, x_i – raidės absoliutus dažnumas, n – teminių grupių skaičius (šiuo konkrečiu atveju). Skaičiavimo duomenys irgi pateikiami 1 lentelėje.

Daugelio raidžių variacijos koeficientas svyruoja tik keliais procentais, o tai rodo, kad tų raidžių dažnumo nukrypimai nuo vidutinio dažnumo reikšmės yra nedideli ir kad skaičiavimo rezultatai visiškai patikimi. Raidžių \acute{e} , \ddot{u} , h variacijos koeficientas didesnis už leistiną ($v_1 = 40\%$), o raidės f koeficientas artimas leistinam. Raidžių \acute{e} , \ddot{u} dažnumų nukrypimai nuo vidutinio dažnumo yra dideli todėl, kad, perforuojant tekstus, vienur kitur padaryta klaidų, t. y. vietoj \acute{e} vietomis atspausdinta e , o vietoj \ddot{u} – u . Šie rezultatai laikytini nepatikimais. Bet tos perforavimo klaidos e ir u variacijos koeficiento reikšmėms pastebimesnės įtakos neturėjo, nes e ir u dažnumai labai dideli. Didelis raidžių f ir h variacijos koeficientas rodo, kad dėl jų mažo dažnumo ir vartojimo epizodiškumo tyrinėtoms medžiagos apimtims buvo per maža³. Pastaroji išvada dalinai liečia ir raidę \ddot{u} .

² Головин В. Н. Язык и статистика. М., 1971, с. 35.

³ Šiaip jau net ir žymus medžiagos apimties padidinimas rezultatų labai nepakeistų, nes, pavyzdžiui, tų raidžių grožiniuose tekstuose būtų dar mažiau.

Iš lentelės matyti, kad tekstų tematika turi įtakos raidžių dažnumui net ir to paties funkcinio stiliaus ribose, išskyrus raides *i, a, s*, kurių dažnumai yra didžiausi, lyginant su kitų raidžių dažnumais (sudaro 32,92%), ir visiškai nepriklauso nuo tekstų tematikos. Šią išvadą patvirtina raidžių dažnumas grožinės literatūros tekstuose⁴. Čia raidžių išsidėstymo tvarka pagal mažėjantį dažnumą yra tokia: *i, a, s, e = o, u, r, t, k, n, l, p, m, v, d, é, j, g, š, b, ž, y, q, u, i, č = ū, é, z, c, f, h*. Taip pat iš lentelės duomenų matyti, kad visų grupių tekstuose raidžių *m, l, p, d, v, j* eilė yra vienoda, o *n, u, k* irgi artima. Labiausiai keičia vietą eilėse rečiausios raidės, nes jos ryškiausiai atspindi tekstų tematiką, pvz., raidė *f* D grupės (dailė, teatras, kinas, literatūros mokslas) raidžių eilėje persikėlė į 27 vietą, nes šioje grupėje žodis *filmas* buvo daug dažnesnis, negu kitose.

Be absoliučių raidžių dažnumų, skaičiuoti ir santykiniai jų dažnumai, kurios šiuo atveju galima laikyti raidžių pasirodymo tikimybėmis⁵. Raidžių pasirodymo tikimybės surašytos 2 lentelėje. Iš duomenų matyti, kad tarpo tikimybė yra didžiausia, kadangi vidutinis lietuviško žodžio ilgis – 6,63 raidės. Be to, yra žinoma, kad kiekvienos raidės pasirodymo tikimybė priklauso ir nuo prieš ją esančių raidžių, pvz., turint balsę, bus žymiai didesnė tikimybė, kad po jos eis priebalsė; po *ę* niekada nebus *į, o, a* ir t. t. Šie dėsningumai išryškėtų, skaičiuojant raidžių junginius po 2 raides, po 3 raides ir t. t.⁶

Palyginus lietuvių kalbos raidžių ir fonemų⁷ pasirodymo tikimybes, tam tikras panašumas matyti tik tarp pačių dažniausių raidžių ir fonemų. Dažniausia fonema – *a*, o dažniausia raidė – *i*, nes ji žymi ne tik fonemą *i*, bet ir minkštumą. Be to, matyti, kad vienų raidžių ir fonemų šios tikimybės yra artimos (pvz., *s, d, r, n*), o kitų (pvz., *a, e, k, t*) gerokai skiriasi. Tik turint duomenų apie fonemų pasirodymo tikimybes publicistikos tekstuose, būtų galima pasakyti, kiek čia turi įtakos tekstų tematika ir kiek atitinkamų raidžių ir fonemų neatitikimas tarpusavyje.

Raidžių ir fonemų pasirodymo tikimybių palyginimui 1 ir 2 grafikuose nubrėžtos priklausomybės $p_i = f(r_i)$, kur p_i – bet kurios raidės ar fonemos pasirodymo tikimybė, r_i – atitinkamos raidės ar fonemos rangas. Iš šių grafikų matyti, kad tiek raidėms, tiek fonemoms šių priklausomybių pobūdis yra toks pat.

⁴ Duomenys imti iš matematikos fakulteto studentų (vadovautų vyr. dėst. Jasiūno) darbų. Jie, deja, niekur neskelbti ir galutinai nesutvarkyti, todėl nepavyko išsiaiškinti iš kokios apimties tekstų ši eilė apskaičiuota.

⁵ Яглом А. М., Яглом И. М. Вероятность и информация. М., 1973, с. 237.

⁶ Nagrinėta daugiausia R. Merkytės darbuose (žr. dar 4 išn.).

⁷ Сведевичюс Б. И. К вопросу о частоте встречаемости фонем в литовской письменной речи. – В кн.: Материалы коллоквиума лаборатории экспериментальной фонетики и психологии речи, т. II. Вильнюс, 1966, с. 19–21. Šis lyginimas labai sąlyginis, nes fonemos skaičiuotos tikrai iš dramų ir beletristikos dialogų. Be to, labai neaiškiai nurodyta tyrinėtoms medžiagos apimtis (4–6 puslapiai) ir ji, aišku, yra per maža.

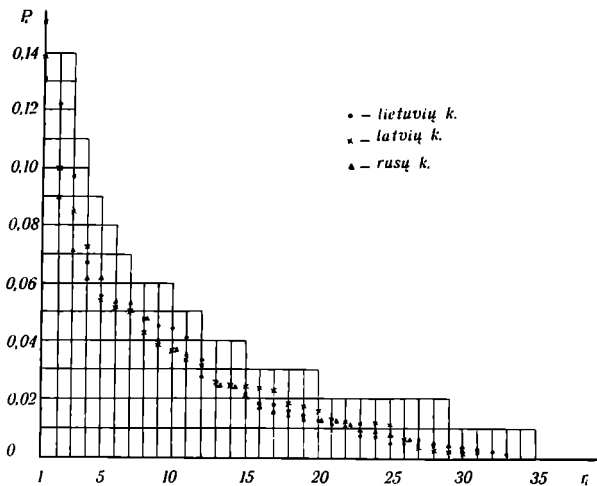
Atskirų teminių grupių ir viso žodyno raidžių dažnumų eilė.
Viso žodyno raidžių variacijos koeficientas

Eil.	Atskirų teminių grupių					Viso žodyno	Variacijos koef.
	A	B	C	D	E		
1	i 14,29	i 13,92	i 13,96	i 14,25	i 13,63	i 14,01	i 2,89
2	a 11,12	a 11,26	a 11,37	a 11,11	a 11,13	a 11,20	a 2,36
3	s 7,98	s 7,54	s 7,69	s 7,70	s 7,66	s 7,72	s 3,22
4	o 5,86	e 6,62	e 6,69	e 6,82	e 7,09	e 6,51	e 8,02
5	t 5,71	o 5,98	o 6,07	o 5,93	o 6,10	o 5,99	o 1,83
6	r 5,37	t 5,72	t 5,89	t 5,73	t 5,82	t 5,78	t 2,10
7	e 5,35	r 5,43	k 5,15	r 5,59	r 5,46	r 5,37	r 3,18
8	n 5,23	n 5,29	u 5,09	u 5,22	u 5,28	n 5,15	n 3,24
9	u 4,64	u 4,98	n 4,99	n 5,06	n 5,16	u 5,04	u 3,11
10	k 4,18	k 4,71	r 4,98	k 4,66	k 4,61	k 4,66	k 6,68
11	m 3,76	m 3,86	m 4,10	m 3,77	m 3,39	m 3,78	m 7,24
12	l 2,97	l 2,88	l 2,88	l 2,84	p 3,01	l 2,88	l 3,36
13	p 2,89	p 2,86	p 2,65	p 2,83	l 2,84	p 2,85	p 3,81
14	d 2,65	d 2,63	d 2,60	d 2,41	d 2,45	d 2,55	d 5,56
15	v 2,37	v 2,09	v 2,21	v 2,11	v 2,42	v 2,24	v 6,31
16	j 1,98	j 1,94	j 1,98	j 2,14	j 2,05	j 2,02	j 2,21

Eil.	Atskirų teminių grupių					Viso žodyno	Variacijos koef.
	A	B	C	D	E		
17	ų 1,66	g 1,93	g 1,91	g 1,78	g 1,83	g 1,82	g 6,15
18	b 1,65	b 1,63	y 1,56	b 1,42	b 1,53	b 1,55	b 6,82
19	g 1,62	ų 1,58	b 1,53	ų 1,37	ų 1,49	ų 1,51	ų 8,18
20	ė 1,55	y 1,39	ų 1,44	š 1,34	š 1,41	y 1,43	y 9,00
21	y 1,54	š 1,17	š 1,16	y 1,29	y 1,35	š 1,27	š 6,56
22	š 1,28	ž 0,91	ž 0,92	ž 1,00	ž 1,00	ž 0,94	ž 3,95
23	ž 0,87	ą 0,69	ą 0,77	ą 0,84	ą 0,75	ą 0,76	ą 5,56
24	ą 0,73	į 0,58	į 0,60	į 0,65	į 0,66	į 0,62	į 3,71
25	į 0,59	c 0,52	c 0,35	č 0,38	č 0,52	ė 0,57	ė 91,34
26	c 0,53	ė 0,49	č 0,34	c 0,37	c 0,27	c 0,41	c 25,94
27	ū 0,41	č 0,39	ė 0,31	f 0,36	z 0,25	č 0,40	č 14,54
28	č 0,36	z 0,26	z 0,24	z 0,34	ė 0,23	z 0,28	z 13,81
29	z 0,31	ū 0,20	e 0,21	e 0,23	e 0,23	e 0,21	e 6,93
30	e 0,21	e 0,19	f 0,16	ė 0,22	f 0,20	f 0,21	f 33,51
31	f 0,18	h 0,17	ū 0,11	h 0,13	ū 0,09	ū 0,18	ū 71,41
32	h 0,03	f 0,16	h 0,09	ū 0,08	h 0,07	h 0,01	h 44,84

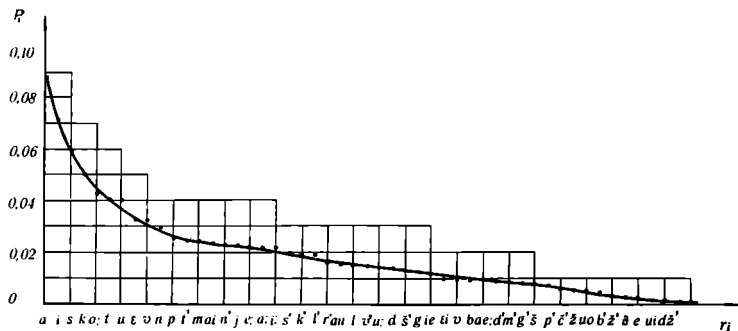
Raidžių pasirodymo tikimybės lietuvių, latvių ir rusų kalbose

Eil. Nr.	Lietuvių k.		Latvių k.		Rusų k.	
	raidė	pasirodymo tikimybė	raidė	pasirodymo tikimybė	raidė	pasirodymo tikimybė
1	tarpas	0,131	tarpas	0,139	tarpas	0,175
2	i	0,122	a	0,100	o	0,090
3	a	0,097		0,085	e, ė	0,072
4	s	0,067		0,073	a	0,062
5	e	0,056		0,054	и	0,062
6	o	0,052		0,051	т	0,053
7	t	0,050		0,050	н	0,053
8	r	0,047		0,043	с	0,045
9	n	0,045		0,039	р	0,040
10	u	0,044	ā	0,037	в	0,038
11	k	0,041	k	0,034	л	0,035
12	m	0,033	m	0,032	к	0,028
13	l	0,025	l	0,026	м	0,026
14	p	0,025	p	0,025	д	0,025
15	d	0,022	d	0,025	п	0,023
16	v	0,019	o	0,024	у	0,021
17	j	0,018	v	0,023	я	0,018
18	g	0,016	z	0,019	ы	0,016
19	b	0,014	i	0,017	э	0,016
20	ц	0,013	j	0,016	ь, ъ	0,014
21	y	0,012	g	0,013	б	0,014
22	š	0,011	ē	0,012	г	0,013
23	ž	0,008	b	0,012	ч	0,012
24	ą	0,007	c	0,012	й	0,010
25	į	0,005	š	0,011	х	0,009
26	ė	0,005	ņ	0,006	ж	0,007
27	c	0,004	ļ	0,003	ю	0,006
28	č	0,003	ū	0,002	ш	0,006
29	z	0,002	ž	0,002	ц	0,004
30	ę	0,002	ķ	0,001	щ	0,003
31	f	0,002	ģ	0,000	э	0,003
32	ū	0,002	f	0,000	ф	0,002
33	h	0,001	h	0,000		
34			č	0,000		



1 grafikas

Lietuvių, latvių ir rusų kalbų bet kurios raidės pasirodymo tikimybės p_i priklausomybė nuo raidės rango r_i



2 grafikas

Lietuvių kalbos bet kurios fonemos priklausomybė nuo fonemos rango

Palyginimui pateikiamos raidžių eilės ir pasirodymo tikimybės latvių kalbos publicistikos tekstuose⁸ ir rusų kalboje⁹ (2 lentelė). Peržiūrėjus šiuos duomenis, matyti, kad atskirų kalbų tarpo tarp žodžių pasirodymo tikimybės skiriasi. Kadangi rusų kalboje tarpo pasirodymo tikimybė yra didžiausia, galima daryti išvadą, kad rusų žodžiai yra trumpesni, negu lietuvių ar latvių. Lietuvių žodžiai ilgiausi, nors jų ilgis artimas latvių žodžių ilgiui. Paskaičiavus šių kalbų balsių pasirodymo tikimybes, gauta: lietuvių – 0,414, latvių – 0,378, rusų – 0,35, bet ir balsių skaičius abėcėlėje skirtingas (atitinkamai – 12, 9, 9). Vadinasi, balsės vidutinė pasirodymo tikimybė latvių kalboje yra didžiausia – 0,042, o lietuvių kalboje mažiausia – 0,035 (rusų kalboje – 0,039). Tiek lietuvių, tiek latvių kalbose žymiai didesnes pasirodymo tikimybes turi trumposios balsės, negu ilgosios. Palyginus tarpusavy šių kalbų priebalsių eiles, matyti, kad lietuvių ir latvių kalbų priebalsių, turinčių didesnes pasirodymo tikimybes, eilė visiškai sutampa, o tuo tarpu panašumas su rusų kalbos priebalsėmis menkas, nors visose trijose kalbose mažas pasirodymo tikimybės turi tos pačios priebalsės – *š, ž, č, c, f, h*.

Rusų ir latvių kalbų raidėms anksčiau minėta priklausomybė $p_i = f(r_i)$ irgi atidėta 1 grafike. Vienodas visų trijų kalbų priklausomybių pobūdis rodo, kad ši priklausomybė yra tokia pati visų kalbų raidėms¹⁰. Iš priklausomybių taškų tarpusavio padėties matyti, kad dažniausių raidžių pasirodymo tikimybės lietuvių kalboje yra didesnės, negu latvių ir rusų kalbose, vidutinio dažnumo raidžių pasirodymo tikimybės lietuvių kalboje yra didesnės, negu lietuvių ir rusų kalbose, o rečiausių raidžių pasirodymo tikimybės yra didesnės rusų kalboje, negu kitose.

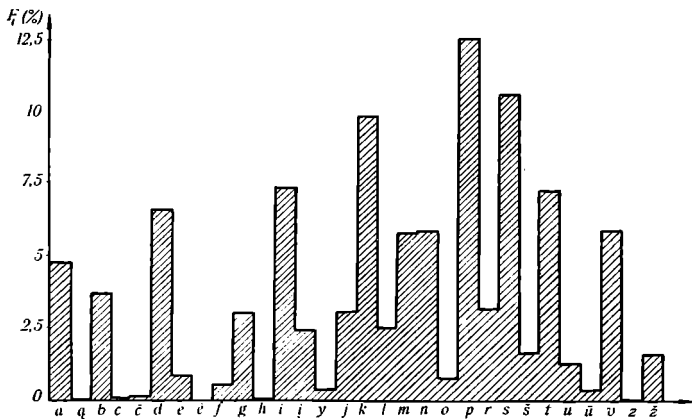
Skaičiuojant raidžių absoliutinius dažnumus, pasirodė, kad labai skiriasi pirmosios žodžio raidės dažnumas (3 grafikas). Net 40% pirmųjų raidžių sudaro tik keturios raidės – *i, k, p, s*, pirmiausia, todėl, kad šiomis raidėmis prasideda patys produktyviausi priešdėliai: *iš-, pa-, pri-, su-*. Dideliam pirmųjų raidžių *d, k, t* dažnumui, be abejonės, turėjo įtakos tekstų tematika, nes kaip tik šiomis raidėmis prasideda labai didelio dažnumo žodžiai *darbas, dalyvauti, karas, taika, taryba*.

Žodžiai, prasidedantys balsėmis, sudaro 18,58% visų žodžių, o priebalsėmis – 81,42% (kaip minėta anksčiau, abėcėlėje balsių yra 37,05%, o priebalsių – 62,95%). Skirtingi žodžiai, prasidedantys balsėmis, tesudaro 10,83% visų skirtingų žodyno žodžių. Taip yra todėl, kad balsėmis prasideda labai didelio dažnumo žodžiai *ir, i, o, iš*.

⁸ Андреев Н. Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л., 1967, с. 230.

⁹ Зг. Яглом А. М., Яглом И. М. Указ. соч., с. 238. Čia raidžių pasirodymo tikimybės skaičiuotos sumaišant ir traukiant po vieną lapelius, ant kurių užrašytos visos rusų abėcėlėje esančios raidės.

¹⁰ Šios priklausomybės buvo nubraižytos anglų kalbos grožinių ir radioelektronikos, vokiečių publicistikos, ispanų publicistikos, prancūzų radioelektronikos tekstų raidėms. Jų visų pobūdis vienodas. Raidžių pasirodymo tikimybės imtos iš Андреев Н. Д. Указ. соч., с. 231–234.



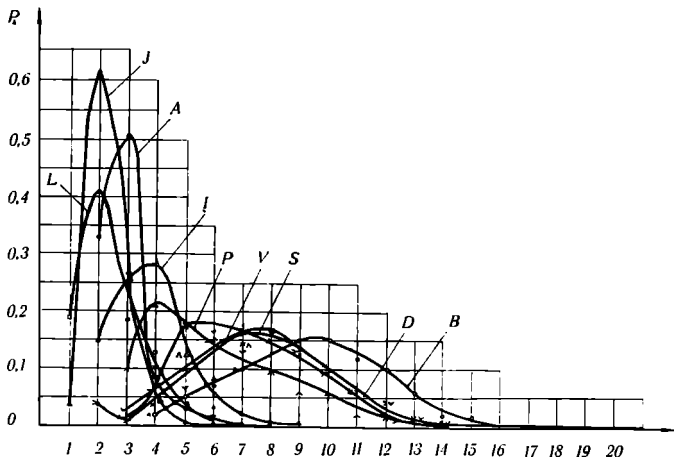
3 grafikas

Pirmosios žodžio raidės dažnumas.

Ordinačių ašyje atidėtas pirmosios žodžio raidės dažnumas — F_1 , išreikštas procentais

Raidžių skaičius žodžiuose, priklausančiuose skirtingoms kalbos dalims, labai nevienodas. Seniai įrodyta, kad trumpiausi žodžiai yra patys dažniausi ir jie priklauso tarnybinėms kalbos dalims, o ilgi žodžiai yra retesni ir priklauso pagrindinėms kalbos dalims. Žodžių, priklausančių tam tikroms kalbos dalims, ilgiai pavaizduoti 4 grafike. Kaip matyti iš grafiko, trumpiausi yra prielinksniai, jungtukai ir dalelytės. Net 40% visų prielinksnių sudaryti iš dviejų raidžių, 62% visų jungtukų irgi sudaryti iš dviejų raidžių, 33,6% dalelyčių sudarytos iš dviejų raidžių, 51% — iš trijų raidžių.

Įvardžiai, skaitvardžiai irrieveksmiai savo ilgiu išsiskiria į atskirą grupę. Tai kalbos dalys, gerokai ilgesnės už tarnybinės kalbos dalis, bet pastebimai trumpesnės už daiktavardžius, būdvardžius, veiksmažodžius. Iš šios grupės trumpiausi yra įvardžiai. 26,3% visų įvardžių sudaryti iš trijų raidžių, o 27,8% — iš keturių raidžių. Daugiausiarieveksmių yra sudarytų iš keturių raidžių (21,8%), orieveksmių, kurių ilgis penkios—aštuonios raidės, kiekis mažai skiriasi, ir tik ilgesniųrieveksmių skaičius žymiai sumažėja. Skaitvardžių, sudarytų iš penkių—devynių raidžių, skaičius mažai svyruoja, o trumpesnių arba ilgesnių — žymiai sumažėja. Įdomu pažymėti, kad iš dviejų raidžių sudarytas tik skaitvardis *du*, o iš trijų — *dvi*, tačiau jų santykinis dažnumas nemažas.



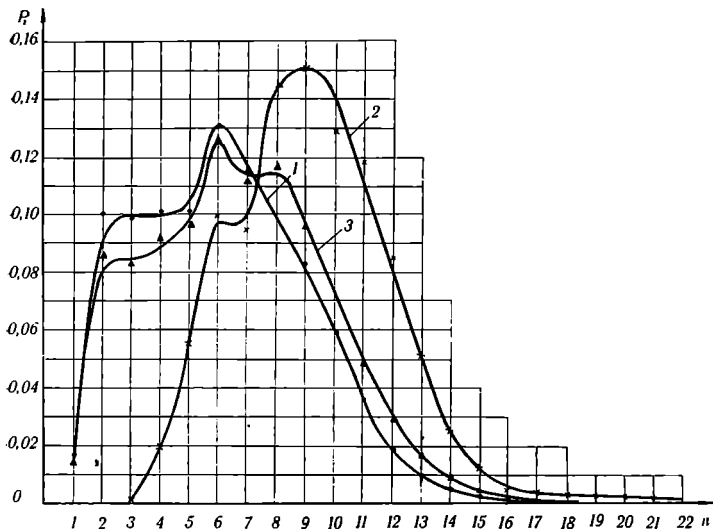
4 grafikas

Raidžių skaičius žodžiuose, priklausančiuose tam tikroms kalbos dalims.

Abscisių ašyje atidėtas raidžių skaičius žodyje – n , ordinačių ašyje – žodžių, priklausančių tam tikrai kalbos daliai, santykinis dažnumas – P_k . Kreives žymintys raidės nurodo: A – dalelytes, B – būdvardžius, D – daiktavardžius, I – įvardžius, J – jungtukus, L – prielinksnius, P –rieveiksmius, S – skaitvardžius, V – veiksmažodžius

Nagrinėjant pagrindinėms kalbos dalims priskiriamų žodžių ilgį, pastebima, kad nedidelė dalis daiktavardžių ir veiksmažodžių yra sudaryta iš trijų raidžių, o būdvardžių – iš keturių raidžių. Veiksmažodžių ir daiktavardžių ilgis labai artimas, tik trumpų veiksmažodžių yra daugiau, negu daiktavardžių. Daugiausia daiktavardžių yra penkių–dešimties raidžių ilgio (beveik 90%), o trumpesnių ir ilgesnių – nedaug. 67% visų veiksmažodžių sudaro žodžiai, kurių ilgis – šešios–devynios raidės. Daug veiksmažodžių keturių raidžių ilgio. Tai beveik vien didelio dažnumo veiksmažodžiai *buvo*, *būti*, *būtų*. Ilgiausi žodžiai – būdvardžiai. Jų daugiausia yra sudaryta iš aštuonių–dvylikos raidžių (64,4%).

Apskaičiavus lietuviškos kilmės ir skolintų žodžių ilgius, pasirodė, kad vidutinis lietuviškos kilmės žodžio ilgis – 6,41 raidės, o skolinto – 7,67 raidės. Skirtumas susidarė todėl, kad skoliniai yra daiktavardžiai, veiksmažodžiai, būdvardžiai (labai nedaug skaitvardžių irrieveiksmių), o šioms kalbos dalims priklausančių žodžių vidutinis ilgis irgi yra didesnis už bendrąjį žodžio ilgį. Lietuviškos kilmės ir



5 grafikas

Lietuvių kalbos žodžių ilgiai.

Abscisių ašyje atidėtas raidžių skaičius žodyje – n , ordinačių ašyje – žodžių santykinis dažnumas – P_2 . 1 kreivė – lietuviškos kilmės žodžių, 2 kreivė – skolintų žodžių, 3 kreivė – visų lietuvių kalbos žodžių

skolintų žodžių ilgių kitimą labai aiškiai rodo 1 ir 2 kreivės 5 grafike. Čia skolintų žodžių kreivė žymiai pasistūmusi į ilgesnių žodžių pusę. Iš šių kreivių matyti, kad daugiausia lietuviškos kilmės žodžių yra šešių raidžių ilgio, o skolintų – 8–9 raidžių. 5 grafike nubrėžta ir bendra lietuvių kalbos žodžio (nežiūrint jo kilmės) ilgio kitimo kreivė. Jos pobūdis artimesnis 1 kreivės pobūdžiui, nes šiame žodyne lietuviškos kilmės žodžiai sudaro 81,9%, o skolinti – 18,1% visų lietuvių kalbos žodžių.

Apskaičiavus skiemenis, pasirodė, kad vienskiemeniai žodžiai sudaro 20,50% visų žodyno žodžių (išskyrus tikrinius), dviskiemeniai – 28,39%, triskiemeniai – 27,28%, keturskiemeniai – 16,42%, penkiaskiemeniai – 5,73%, šešiaskiemeniai – 1,41%, visi likę – 0,27%. 3 lentelėje surašyta skiemenų skaičius žodžiuose, priklausomai nuo kalbos dalies.

Pagrindinėms kalbos dalims priskiriamų žodžių skiemenų skaičius

Skiemenų skaičius	Daiktavardžiai %	Būdvardžiai %	Veiksmazodžiai %
1	0,11	—	1,32
2	32,82	19,61	24,86
3	36,37	33,99	38,82
4	20,87	30,09	25,30
5	7,12	12,70	8,18
6	2,26	2,84	1,36
7	0,37	0,69	0,14
8	0,06	0,09	0,01

Iš lentelės duomenų matyti, kad pagrindinėms kalbos dalims priklausantys žodžiai dažniausiai sudaryti iš trijų skiemenų. Daugumas įvardžių – vienskiemeniai ir dviskiemeniai (91,62%). Tarnybinėms kalbos dalims priskiriami žodžiai, aišku, dažniausiai yra vienskiemeniai, pvz., vienskiemeniai jungtukai sudaro 93,31% visų jungtukų, vienskiemeniai prielinksniai – 96,98% visų prielinksnių.

Lietuvių kalbos raidžių dažnumo skaičiavimo rezultatus būtų galima pritaikyti praktikoje. Pagal raidžių dažnumų duomenis reikėtų pertvarkyti lietuviškų rašomųjų mašinelių klaviatūrą. Spausdinant akluoju būdu, abiejų rankų keturi pirštai (išskyrus nykščius) laikomi ant klavišų tokiu būdu:

a	s	d	š	g	h	j	k	l	ų	ė
maža-	bevar-	didy-	rodo-				rodo-	didy-	bevar-	maža-
sis	dis	sis	masis				masis	sis	dis	sis
kairės rankos						dešinės rankos				

Būtų patogiau ir lengviau, kad judriausi pirštai (rodomieji ir didieji) būtų laikomi ant dažniausių raidžių klavišų, t. y. ant *i*, *a*, *s*, *e*. O *i*, *e* klavišai yra kitoje klaviatūros eilėje, ir jų spaudimui perkeliama rodomieji pirštai, o tai jau sumažina spausdinimo greitį. Klavišus *a* ir *s* spausti kairės rankos mažuoju ir bevardžiu pirštais irgi nepatogu, nes šie pirštai mažiausiai paslankūs. Racionaliausia būtų vietoj klavišo *k* įtaisyti klavišą *i*, vietoj *š* – *a*, vietoj *l* – *s*, vietoj *d* – *e*. Klavišuose *g* ir *j* turėtų būti tolesnės pagal dažnumą raidės, t. y. *o* ir *t*, kadangi šiuos klavišus lengviausia pasiekti, perkeliama rodomuosius pirštus. Retesnės *g* ir *j* raidės galėtų stovėti nepatogesnėse pozicijose. Patogi ir klavišo *h* padėtis, o ši raidė dažnumų eilėje stovi paskutinė. Gal tik raidžių *z*, *ū*, *f*, *ę* klavišų padėtis pateisinama, nes tai rečiausios raidės ir jų klavišai yra nepatogiausiose vietose.

Baigdamą autorė dėkoja universiteto Skaičiavimo centro kolektyvui, o ypač skyriaus viršininkui Stasiui Liberiui už didžiulę paramą, skaičiuojant lietuvių kalbos publicistikos žodžių dažnumą.

ЧАСТОТА БУКВ ЛИТОВСКОГО ЯЗЫКА В ТЕКСТАХ ПУБЛИЦИСТИКИ

Резюме

Описана методика вычисления частоты букв по материалу частотного словаря литовской публицистики. Массив исследованных слов – 280 876, в котором было 1 863 159 букв. Расчеты проводились на ЭВМ „Минск-22“ в вычислительном центре Вильнюсского государственного университета. Составлен ряд букв по убывающей частоте и ряды букв по отдельным тематическим группам словаря. По этим данным установлено, что тематика текстов не оказывает влияния только на частоту трех самых частых букв. Указаны относительные частоты каждой буквы. О достоверности результатов судится по коэффициенту вариации. Вероятности появления букв литовского языка сопоставлены с вероятностями появления фонем литовского языка и букв латышского и русского языков. Даны графики зависимостей вероятности появления от ранга для букв и фонем литовского языка и для букв латышского и русского языков. Установлено, что характер этих зависимостей является общим для букв разных языков. Рассчитаны частоты первых букв литовских слов и дана гистограмма. Исследовано число букв литовских слов, принадлежащих разным частям речи и даны соответствующие кривые. Показано, что число букв в заимствованных словах заметно превышает число букв литовских слов. Рассчитано число слогов в литовских словах, принадлежащих разным частям речи. Обсуждено возможное практическое применение частоты букв литовского языка.