

Vida ŽILINSKIENĖ

Matematikos ir informatikos institutas, Vilnius

**DAŽNINIS DABARTINĖS RAŠOMOSIOS LIETUVIŲ
KALBOS ŽODYNAS
(pirmieji tyrimo rezultatai)**

Jau funkcionuoja kompiuterinis dažninis dabartinės rašomosios lietuvių kalbos žodynas, sudarytas iš keturių funkcinių stilių (beletristikos, publicistikos, mokslo ir dalykinio) medžiagos. Plačiau apie žodyno rengimo metodiką ir principus žr. straipsnyje (Grumadienė 1995).

Kurios nors kalbos, tam tikro funkcinio stiliaus, rašytojo ar kokio nors kūrinio leksikos įvairovė yra įvertinama vadinamuoju iteracijos koeficientu (žodžių pavartojimų ir žodžių santykiu), kuris tuo pačiu reiškia ir vidutinį žodžio dažnį tekste. Atlikti skaičiavimai rodo, kad minėtųjų stilių ir visos lietuvių kalbos iteracijos koeficientas yra:

beletristikos – 9,53,
publicistikos – 10,51,
mokslo – 12,63,
dalykinio – 20,78,
visos kalbos – 20,06.

Kuo šis koeficientas mažesnis, tuo tiek pat žodžių pavartojimų (mūsų atveju tirta po 300000 žodžių pavartojimų iš kiekvieno funkcinio stiliaus tekstų) turinčiuose tekstuose rasta daugiau skirtingų žodžių, tuo to stiliaus leksika įvairesnė. Iš pateiktų skaičių labai aiškiai matyti, kad: 1) beletristikos leksika dukart įvairesnė už dalykinio stiliaus, 2) beletristikos, publicistikos ir net mokslo stilių leksikos įvairovė gana artima.

Tirtuose tekstuose rastų žodžių formų įvairovę galime lyginti gretindami atskirų stilių žodžių pavartojimų ir žodžių formų santykį:

beletristikos – 4,34,
publicistikos – 4,48,
mokslo – 4,83,
dalykinio – 7,29,
visos kalbos – 7,16.

Šie skaičiai irgi akivaizdžiai rodo, kad: 1) įvairesnių gramatinių formų vartojama beletristikos, publicistikos ir mokslo stiliuose, 2) šiuo požiūriu nuo kitų stilių taip pat gerokai skiriasi dalykinis stilius.

Kadangi turime pagal tą pačią metodiką parengtą lietuvių publicistikos dažninį žodyną (Žilinskienė 1990), skaičiuotą iš 1973–74 m. laikraščių ir žurnalų tekstų, ir latvių publicistikos dažninį žodyną (Jakubaite 1969), skaičiuotą iš 1966–67 m. tekstų, tai galime palyginti jų iteracijos koeficientus:

1993–95 m. lietuvių publicistikos	–	10,51,
1973–74 m. lietuvių publicistikos	–	15,38,
1966–67 m. latvių publicistikos	–	18,60.

Šie koeficientai aiškiai rodo, kad dabartinės lietuvių publicistikos leksika yra įvairesnė negu prieš dvidešimt metų buvusios lietuvių publicistikos leksika arba beveik prieš trisdešimt metų – latvių publicistikos leksika.

Palyginkime atskirų stilių ir visos kalbos dažniausių žodžių dešimtukus (žr. 1 lentelę; joje ir kitose lentelėse šalia žodžio esantis skaičius reiškia to žodžio dažnį).

1 lentelė. Atskirų stilių dažniausių žodžių dešimtukai

beletristika	publicistika	mokslo	dalykinis	visos kalbos
ir* 11442	ir* 9743	ir* 11140	ir* 10990	ir* 43315
jis 9634	būti 5825	būti 6294	būti 5138	jis 24613
aš 4580	jis 5824	jis 4692	jis 4463	būti 21366
į 4216	tas 3487	tas 2783	tas 3286	tas 13528
būti 4109	kad* 3235	šis 2138	kuris 2273	į 10715
tas 3972	į 2719	kuris 2054	Lietuva 2195	kad* 9877
o* 2822	aš 2602	į 1941	šis 2135	aš 9025
kad* 2811	kuris 1985	kad* 1844	kad* 1992	kuris 7610
kaip 2219	o* 1904	su 1838	ar 1903	iš 7162
kas 2136	su 1897	iš 1647	į 1839	su 7034

Žodžių, 1 lentelėje pažymėtų žvaigždute (*), dažnis skiriasi nuo tų pačių žodžių dažnio, pateikto tezėse (Žilinskienė 1995), nes čia yra pateikiamas bendras konkrečios leksemos dažnis, o tezėse pateiktas atitinkamos leksemos, tirtuose tekstuose einančios tik jungtuku, dažnis.

Iš 1 lentelės aiškiai matyti, kad: 1) į pirmąjį dažniausių žodžių dešimtuką pateko tarnybinės kalbos dalys, įvardžiai, kurie statistiniu požiūriu paprastai elgiasi kaip tarnybinės kalbos dalys, ir veiksmazodis *būti*, kuris tekste dažnai vartojamas kaip pagalbinis veiksmazodis, 2) pirmajame dažniausių žodžių de-

šimtuose šeši žodžiai (*ir, jis, būti, tas, kad, į*) yra tie patys, tik skiriasi jų rangai (pozicijų numeriai dažnių sąrašuose), aišku, išskyrus *ir*, 3) yra tik vienas konkrečią reikšmę turintis žodis *Lietuva* dalykinio stiliaus dažniausių žodžių dešimtuose.

Palyginkime dabartinės, senesnės lietuvių kalbos publicistikos ir latvių kalbos publicistikos dažniausių žodžių dvidešimtukus (žr. 2 lentelę).

2 lentelė. Lietuvių ir latvių publicistikos dažniausi žodžiai

1993–95 m. publicistika	1973–74 m. publicistika	latvių publicistika
ir 9743	ir 11890	un 10401
būti 5825	jis 5733	būt 6395
jis 5824	būti 4167	tas 5244
tas 3487	tas 3939	ar 3631
kad 3235	į 2614	par 3546
į 2719	aš 2461	kas 3119
aš 2602	kad 2455	viņš 2610
kuris 1985	kuris 2394	no 2342
o 1904	su 2046	ka 2325
su 1897	šis 2024	ari 2304
iš 1890	visas 1825	šis 2275
Lietuva 1839	darbas 1759	darbs 2059
šis 1715	iš 1602	mēs 2030
visas 1517	kaip 1594	gads 1789
savas 1463	metai 1563	bet 1643
bet 1273	o 1538	uz 1630
žmogus 1271	žmogus 1421	kā 1609
ar 1213	ne 1389	viss 1601
metai 1302	tik 1277	padome 1386
apie 1165	savo 1263	savs 1349

Lygindami abiejų periodų lietuvių publicistiką, matome, kad pačių dažniausių žodžių vartojimo dažnis labai panašus, tik trys žodžiai (*bet, ar, apie*), esantys

dabartinės publicistikos pirmajame dvidešimtuje, yra pasistūmę žemiau ankstesnėje publicistikoje (*bet* yra 22-oje, *ar* – 39-oje, *apie* – 26-oje pozicijoje) (Žilinskienė 1990), o žodis *Lietuva* ankstesnėje publicistikoje buvo užšifruotas kaip valstybės pavadinimas ir ten buvo skaičiuotas tik atskirų tikrinių žodžių grupių (pvz., vardų, pavardžių, valstybių pavadinimų, vandenvardžių ir t.t.) bendras dažnis, todėl žodžių žodyne mažėjančio dažnio tvarka nėra įrašytas. Tvarkant dabartinę publicistiką *savo* yra laikytas įvardžio *savas* kilmininku ir, aišku, kad tada fiksuota leksema *savas*, o taip perskaičius ankstesnę publicistiką *savas* būtų turėjęs dažnį 1461 ir būtų atsidūręs 17-oje pozicijoje. Be to, reika pastebėti, kad antrajame dažniausių žodžių dešimtuje jau atsirado ir daiktavardžių (*metai*, *žmogus*), tik *darbas* dabartinėje publicistikoje pasistūmė į trečiąjį dešimtuką.

Lygindami su latvių publicistika, galime matyti, kad į pirmąjį dešimtuką pateko dauguma tokių pačių tarnybines funkcijas atliekančių žodžių, tie patys ir du dažniausi daiktavardžiai, tik latvių *padome* dar labai dažnas (ankstesnėje lietuvių publicistikoje *taryba* pasikartojo 455, o dabartinėje – 137 kartus).

Palyginkime visų keturių stilių ir visos kalbos pirmajame dažniausių žodžių šimte esančius daiktavardžius, būdvardžius, skaitvardžius, veiksmažodžius irrieveksmius (žr. 3 lentelę).

3 lentelė. Atskirų stilių dažniausi vardažodžiai, veiksmažodžiai irrieveksmiai

beletristika	publicistika	mokslo	dalykinis	visos kalbos
žmogus	Lietuva	darbas	Lietuva	Lietuva
akis	žmogus	sistema	darbas	darbas
ranka	metai	Lietuva	dokumentas	žmogus
galva	darbas	žmogus	straipsnis	metai
diena	šalis	tyrimas	respublika	teisė
laikas	diena	dalis	teisė	dokumentas
namas	metas	metas	asmuo	laikas
žodis	valstybė	procesas	įmonė	diena
vyras	žemė	elementas	įstatymas	valstybė
kartas	Vilnius	medžiaga	pirmininkas	respublika
darbas	laikas	vanduo	valstybė	dalis
vaikas	gyvenimas	laikas	įstaiga	įstatymas
tėvas	mokykla	metai	organizacija	straipsnis
koja	Rusija	kultūra	kalba	įmonė

beletristika	publicistika	mokslo	dalykinis	visos kalbos
metas	vaikas	duomuo	dalis	žodis
moteris	lietuvis	lietuvis	žodis	kalba
gyvenimas	įmonė	sąlyga	deputatas	šalis
langas	kartas	būdas	teismas	asmuo
	bankas	skaičius	komisija	sistema
	įstatymas	reikšmė	klausimas	kartas
	pasaulis	atvejis	byla	žemė
	namas	kalba	nutarimas	metas
	ūkis	mokslas	tvarka	vieta
	Europa	forma	prekė	gyvenimas
	dalys	grupė	taryba	
	miestas	paviršius	šalis	
		metodas	darbuotojas	
		rezultatas	narys	
		veikla	žmogus	
		problema	atvejis	
		vieta	vyriausybė	
		savybė	pažeidimas	
			priemonė	
			diena	
			projektas	
			bankas	
			posėdis	
			forma	
			sutartis	
			laikas	
			kaina	
didelis	didelis	didelis	valstybinis	didelis
geras	naujas	įvairus	aukštas	naujas
	geras	pagrindinis	administracinis	aukštas
	svarbus	naujas	pagrindinis	
		svarbus	svarbus	

beletristika	publicistika	mokslo	dalykinis	visos kalbos
		bendras socialinis		
vienas du	vienas pirmas du trys	vienas pirmas du	vienas pirmas du	vienas pirmas du
būti galėti nebūti turėti sakyti žinoti eiti norėti imti reikėti pasakyti žiūrėti negalėti matyti atrodyti	būti turėti galėti nebūti sakyti reikėti	būti galėti turėti nebūti sudaryti naudoti nustatyti	būti turėti galėti reikėti nebūti numatyti norėti	būti galėti turėti nebūti reikėti žinoti nustatyti norėti sudaryti sakyti gauti negalėti
taip labai daug kur čia ten gerai tada	labai dabar taip todėl čia dar jau	labai daug todėl dažnai	taip pat daug taip dabar todėl dažnai	labai daug dar taip jau dabar čia taip pat kur todėl tik

Kaip ir reikėjo tikėtis, iš pirmojo šimto dažniausių žodžių aiškėja, kad pirmiausia skiriasi beletristika nuo negrožinių stilių: 1) beletristikoje mažiausia daiktavardžių (jų yra 18) ir būdvardžių (jų yra 2) bei dukart daugiau veiksmažodžių (yra 15), 2) beletristikai artimesnė publicistika tiek kalbos dalių santykiu, tiek konkrečių leksemų atžvilgiu (pvz., *namas, kartas, vaikas, gyvenimas, geras, sakyti, čia* yra tik šių stilių dažniausių žodžių pirmajame šimte). Kelios leksemos yra tos pačios visų stilių dažniausių žodžių pirmajame šimte: *žmogus, darbas, laikas, būti, galėti, nebūti, turėti*. Iš pirmojo šimto dažniausių daiktavardžių jau nesunkiai būtų galima pasakyti, iš kokio stiliaus tekstų šie žodžiai išrinkti. Pvz., niekaip neįmanoma įsivaizduoti, kad beletristikoje būtų dažni žodžiai *tyrimas, procesas, elementas, duomuo, sąlyga, reikšmė, forma*, lygiai kaip ir neįmanoma įsivaizduoti, kad *akis, ranka, galva, vaikas, tėvas* ir pan. dažnai pasikartotų mokslo stiliaus kalboje. Tokie žodžiai gali būti dažni tik nebent vienoje kokioje nors konkrečioje imtyje, kurioje būtent tik apie tai rašoma. Be to, labai akivaizdu, kad pirmojo dažniausių žodžių šimto veiksmažodžiai yra labai bendros reikšmės ir tekstuose dažniausiai funkcionuoja kaip pagalbiniai veiksmažodžiai.

Palyginkime dabartinės, ankstesnės lietuvių kalbos publicistikos ir latvių kalbos publicistikos pirmajame dažniausių žodžių šimte esančius daiktavardžius, būdvardžius, skaitvardžius ir veiksmažodžius (žr. 4 lentelę).

4 lentelė. Lietuvių ir latvių publicistikos dažniausi vardažodžiai ir veiksmažodžiai

1993–95 m. publicistika	1973–74 m. publicistika	latvių publicistika
Lietuva	darbas	darbs
žmogus	metai	gads
metai	žmogus	padome
darbas	mokykla	cilvēks
šalis	šalis	laiks
diena	kartas	partija
metas	respublika	tauta
valstybė	ūkis	valsts
žemė	diena	diena
Vilnius	gyvenimas	respublika
laikas	filmas	zeme
gyvenimas	rajonas	vieta
mokykla	laikas	skola
Rusija	miestas	rūpnīca
vaikas	mokslas	rajons
lietuvis	žemė	dzīve

1993–95 m. publicistika	1973–74 m. publicistika	latvių publicistika
įmonė kartas bankas įstatymas pasaulis namas ūkis Europa dalis miestas	vaikas liaudis taryba klausimas pasaulis vieta metas mokytojas gamyba organizacija sportas studentas moteris	pilsėta jautājums organizacija reize komiteja uzņēmums pasaule vārds ciņa uzdevums mācība spēks strādnieks
didelis naujas geras svarbus	naujas didelis geras tarybinis svarbus jaunas aukštas socialistinis tikras	jauns liels labs augsts
vienas pirmas du trys	vienas pirmas	pirmais viens divi otrs
būti turėti galėti nebūti sakyti reikėti	būti turėti galėti reikėti nebūti žinoti dirbti kalbėti norėti sakyti	būt nebūt varėt klūt notikt strādāt sākt dot

Šis lyginimas kaip tik ir rodo, kokios lekšemos būdingiausios publicistikai, t.y. kokios lekšemos yra dažnos ir abiejų periodų lietuvių publicistikoje, ir latvių

publicistikoje. Pagaliau tam tikros dažnos leksemos (pvz., tik dabartinėje publicistikoje tarp dažniausių žodžių jau atsидūrė *bankas, įstatymas*, o ankstesnėje publicistikoje buvo *tarybinis, socialistinis*) rodo atitinkamo laikotarpio politinius ir ekonominius aspektus.

Šiame tyrime naudotasi kompiuterinio Dažninio dabartinės rašomosios lietuvių kalbos žodyno, kuris buvo sudarytas remiant Lietuvos mokslo ir studijų fondui, medžiaga.

Gauta 1998 06 26

LITERATŪRA

- Grumadienė L. 1995: Dabartinės rašomosios lietuvių kalbos dažnumų žodynas: sandara ir darbo eiga. – *Lietuvių kalba: tyrėjai ir tyrimai*. Konferencijos pranešimų tezės, Vilnius: Lietuvių kalbos institutas, 18–19.
- Jakubaite T. 1969: Jakubaite T., Guļevska D., Ozola V., Prūse R., Rubina A., Sika N. *Latviešu valodas biežuma vārdnīca*. 2, Rīgā: Zinātne.
- Žilinskienė V. 1990: *Lietuvių kalbos dažninis žodynas*, Vilnius: Mokslo, 176.
- Žilinskienė V. 1995: Rašomosios lietuvių kalbos dažninis žodynas. – *Lietuvių kalba: tyrėjai ir tyrimai*. Konferencijos pranešimų tezės, Vilnius: Lietuvių kalbos institutas, 33–34.

HÄUFIGKEITSWÖRTERBUCH DER LITAUISCHEN SCHRIFTSPRACHE (erste Forschungsergebnisse)

Zusammenfassung

Dieser Beitrag bietet die Einschätzung der Häufigkeit der Wörter in der Lexik der verschiedenen funktionalen Stile der gegenwärtigen litauischen Schriftsprache nach Iterationskoeffizienten, wobei auf das Verhältnis der Wortformen in den Stilgattungen hingewiesen wird. Verglichen werden, die gebräuchlichsten Wörter in den Stilgattungen und die gebräuchlichsten in der litauischen Publizistik 1993–1995 bzw. 1973–1974 sowie in der lettischen Publizistik. Der Vergleich zeigt, welche Lexeme für die Publizistik besonders kennzeichnend sind und wie sie die politischen bzw. wirtschaftlichen Aspekte einer gewissen Zeitperiode widerspiegeln.