



Asta Kazlauskienė,
Erika Rimkutė, Andrius Utkā

Kiekybiniai tyrimai kalbotyroje (II)



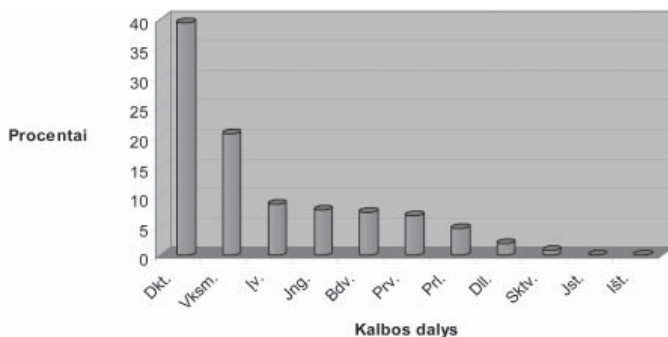
Kalbos dalių ir gramatinių formų dažnumas

Ankstesniame straipsnyje¹ rašėme apie statistinius garsinės ir morfeminės žodžių struktūros tyrimus. Ne mažiau įdomūs gali būti ir realios kalbos (vartosenos) gramatinės struktūros tyrimai. Lietuvių kalba pasižymi gramatinių formų įvairove, bet jų vartosenos kiekybiniai rodikliai labai skiriasi. Tą patvirtina nemažai tyrimų, kuriuose plačiai aprašyti gramatinių formų pasiskirstymo dabartinės kalbos vartosenoje dėsningumai (plačiau žr. straipsnio literatūros sąrašė pateiktus Vidos Žilinskienės, Inetos Savickienės, Erikos Rimkutės darbus ir juose cituojamus ankstesnius tyrimus).



Kad pailiuotume, kaip gali būti atliekami tokie tyrimai ir apibendrinami jų rezultatai, pasiremsime iš vieno milijono žodžių sudarytu morfologiškai anotuotu (t. y. nustatyti ir pažymėti kiekvieno žodžio morfologiniai požymiai) dabartinės lietuvių kalbos tekstynu (plačiau apie tekstyną galite paskaityti Rimkutės disertacijoje *Morfologinio daugiareikšmiškumo ribojimas kompiuteriniame tekстыne*, 2006). Kaip

matyti iš paveikslo, dažniausiai šiame tekстыne vartojamos kalbos dalys yra daiktavardžiai ir veiksmazodžiai (sudaro atitinkamai apie 40 ir 21 proc.). Kiek retesni įvardžiai (beveik 9 proc.), jungtukai (apie 8 proc.), būdvardžiai (7 proc.),rieveksmiai (beveik 7 proc.), prielinksniai (beveik 5 proc.). Kitos kalbos dalys kartu sudaro mažiau nei 2 proc. visų kalbos dalių.



¹ Kazlauskienė A., Rimkutė E., Utkā A. Kiekybiniai tyrimai kalbotyroje // *Gimtasīs žodis*, 2011, Nr. 10.

Ką iš minėto morfologiškai anotuoto tekstyno duomenų galima pasakyti apie kalbos dalių kaitybinių formų vartoseną?

Paaiškėjo, kad dažniausiai vartojamos vyriškosios **giminės** formos (maždaug 57 proc.). Tai lėmė turbūt ta priežastis, kad vyriškoji giminė yra nežymėtasis giminės kategorijos narys, t. y. vyriškosios giminės žodžiu galima įvardyti apibendrintus dalykus, daiktus, reiškinius. Moteriškosios giminės formos sudaro maždaug 37 proc. (kiti vardažodžiai gali būti bevardės giminės).

Kaip nežymėtasis **skaičiaus** kategorijos narys vienaskaitos formos yra dvigubai dažnesnės negu daugiskaitos (atitinkamai 65 ir 32 proc.; atkreipkite dėmesį, kad yra tokių žodžių, kuriems neaktuali skaičiaus kategorija, pvz., bevardės giminės formos, kiekiniai pagrindiniai skaitvardžiai).

Iš **linksnių** dažniausias yra vienaskaitos kilmininkas (beveik 23 proc.), šiek tiek rečiau vartojamos vienaskaitos vardininko formos (22 proc.). Panašaus dažnumo yra daugiskaitos kilmininko ir vienaskaitos galininko formos: atitinkamai 12 ir 11 proc. Iš šių duomenų matyti, kad visų linksniuojamųjų kalbos dalių linksniai aiškiai skirstosi į dvi grupes: pagrindinę (vardininkas, kilmininkas, galininkas) ir periferinę (įnagininkas (daugiau nei 6 proc.), vietininkas (beveik 5 proc.), naudininkas (beveik 4,5 proc.), šauksmininkas (mažiau nei 0,4 proc.).

Veiksmazodžių asmenuojamosios ir neasmenuojamosios formos yra pasiskirsčiusios gana tolygiai (atitinkamai 51 ir 49 proc.). Iš asmenuojamųjų formų dažniausiai vartojama tiesioginė **nuosaka** (91 proc.), liepiamoji nuosaka sudaro 3 proc., tariamoji – 6 proc. Iš neasmenuojamųjų formų dažniausi yra dalyviai (56 proc.) ir bendratys (beveik 34 proc.). Padalyviai sudaro maždaug 6 proc., pusedalyviai – apie 4 proc., būdiniai – vos 0,04 proc. Daugiausia morfologiškai anotuotame tekстыne yra esamojo (53 proc.) ir būtojo kartinio (31 proc.) **laiko**; trečiojo **asmens** formų (jos sudaro 81 proc.). **Dalyvių** rūšys pasiskirsčiusios taip: neveikiamosios rūšies dalyviai sudaro beveik 66 proc., veikiamosios – 32 proc., o reikiamybės – 2 proc.

Kitos rečiau vartojamos gramatinės formos pasiskirstė taip: daugiausia nelyginamojo **laipsnio** (jos sudaro 87 proc.) formų; aukštesniojo laipsnio formos sudaro 8 proc., aukščiausiojo – beveik 5 proc. Turbūt ir be išsamių tyrimų įsivaizduojate, kad daugiausia vartojama **neįvardžiutinių** formų (beveik 94 proc.), **įvardžiutinės** sudaro tik apie 6 proc.

Savaime suprantama, dažnesni yra **bendriniai** (89 proc.), o ne tikriniai daiktavardžiai (11 proc.). Vargu ar mokiniai gerai skiria skaitvardžius, tad būtų galima paanalizuoti, kiek ir kokios rūšies skaitvardžių vartojama pasirinktame tekste. Gautus rezultatus palyginkite su morfologiškai anotuoto tekstyno duomenimis. Jame kiekiniai pagrindiniai **skaitvardžiai** sudaro 58 proc., kelintiniai – beveik 38 proc., kiekiniai dauginiai – 4 proc., o kiekiniai kuopiniai – tik 0,3 proc. (kiekinių trupmeninių nerasta, nes jie dažniausiai užrašomi skaitmenimis, o morfologiškai anotuotame tekстыne analizuota tik tai, kas užrašyta žodžiais).

Statistinius duomenis reikia interpretuoti gana atsargiai, nes kartais juos gali iškreipti kelių neįprastų arba labai dažnų žodžių vartoseną. Tekstynų lingvistikos tyrimai atskleidė, kad bet kurioje kalboje yra labai mažai dažnai vartojamų žodžių

(dažniausiai tai nekaitomi žodžiai; populiariausi lietuvių kalbos nekaitomi žodžiai yra *ir, į, su, iš, o, ar*) ir labai daug retai vartojamų žodžių. Palyginkite ankstesnėse pastraipose aprašytus kalbos dalių vartosenos rezultatus su toliau pateiktais. Jei skaičiuosime ne visų žodžių pavartojimo atvejus, o analizuosime tik skirtingų žodžių vartojimo dažnumą, gausime kiek kitokius duomenis: skirtingi daiktavardžiai sudaro 45 proc., veiksmažodžiai – 39 proc., t. y. daugiau nei analizuojant visus tiriamojoje medžiagoje rastų daiktavardžių ir veiksmažodžių pavartojimo atvejus. Kitų kalbos dalių vartoseną lieka panaši. Šiuos duomenis, tarkim, veiksmažodžių vartoseną, reikia interpretuoti taip: kalboje paprastai vartojama gana daug skirtingos reikšmės veiksmažodžių, bet kiekvieno iš jų dažnumas skiriasi. Tai reiškia, kad didelė dalis veiksmažodžių tekstyne pavartota vos kelis kartus (pvz., morfologiškai anotuoto tekstinio publicistikos dalyje *atriboti, išdegti* pavartoti po 2 kartus, *motyvuoti* 5 kartus). Žinoma, yra ir tokių veiksmažodžių, kurie labai dažni. Tai dažniausiai savo pagrindinę reikšmę praradę, modaliniais virtę veiksmažodžiai *būti* (morfologiškai anotuoto tekstinio publicistikos dalyje pavartotas beveik 5000 kartų), *galėti* (pavartotas beveik 1200 kartų), *turėti* (pavartotas beveik 1000 kartų). Taigi apibendrinant galima pasakyti, kad skirtingų veiksmažodžių įvairovė didelė, bet jų pavartojimo atvejų kiekis kitų kalbos dalių atžvilgiu mažas (tik penktadalis).

Morfologiškai anotuoto lietuvių kalbos tekstinio duomenys paneigė mūsų įsivaizdavimą apie lietuvių kalbos gramatinių formų gausą vartosenoje. Juk visi žinome, kad, pavyzdžiui, daiktavardis, teoriškai gali turėti 14 formų: 7 vienas-kaitos, 7 daugiskaitos linksnius. O kaip yra realioje kalboje, ne gramatikoje? Jau nustatėme, kad tik vardininkas, kilmininkas ir galininkas vartojami dažnai. Taigi paaiškėjo, kad dabartinėje kalboje dažniau vartojamos tik 2–3 tam tikros kalbos dalies formos, tiksliau, vienai lemai (antraštiniam žodžiui, žodyninei formai) vidutiniškai tenka tik 2,3 formos. Reikia atsižvelgti į tai, kad, lyginant visą medžiagą, kaitybinių formų skaičių mažina nekaitomosios kalbos dalys, kurios vartojamos labai dažnai. Tačiau galima būtų atlikti tyrimą ir analizuoti tik kaitomųjų kalbos dalių formų įvairovę vartosenoje.

Vartojamos vidutiniškai tik 2,5 daiktavardžių, 2,6 veiksmažodžių kaitybinės formos. Didele formų įvairove pasižymi įvardžiai (vidutiniškai 8,2 formos, nes yra įvardžių, kurie turi giminę, skaičių, linksnį, netgi gali būti įvardžiuotiniai (pvz., *tasai*); yra tokių įvardžių, kurie neturi skaičiaus (pvz., *keli*); yra visiškai nekaitomų įvardžių (pvz., *keliadešimt*). Taigi įvardžių įvairovė didžiulė, labai skiriasi jų gramatinės kategorijos, galbūt dėl to kai kurie įvardžiai (jau minėti *keli, keliadešimt*, taip pat *keliolika, kelintas*) neretai supainiojami su skaitvardžiais). Nemažai vartojama ir skirtingų būdvardžių (4,72) ir skaitvardžių (3,55) formų. Prieveiksmių vartojama 1,52 (prisiminkite, kadrieveiksmiai gali būti laipsniuojami). Vadinasi, galima daryti išvadą, kad nors lietuvių kalba yra sudėtinga fleksinė kalba, pasižyminti gramatinių formų įvairove, tačiau realiai pavartojama tik labai nedidelė tų formų dalis. Galbūt tai ir yra viena iš priežasčių, kodėl kai kurias formas mokiniai sunkiai atpažįsta.

Straipsnyje pateikėme gana daug anksčiau atliktų tyrimų nuorodų. Tikimės, kad tiek mokytojai, tiek mokiniai susidomės kiekybiniais morfologijos tyrimais ir patys pabandys paanalizuoti. Galėtume pasiūlyti ištirti, tarkim, kurio nors rašytojo vartojamų gramatinių formų pasiskirstymą ir palyginti su kito stiliaus kūrinium, pavyzdžiui, moksliniu tekstu. Net ir suskaičiavus duomenis iš nedidelės ištraukos, paaiškės, kad grožiniame kūrinyje (ypač jei jame yra dialogų) bus daugiau veiksmažodžių, ypač liepiamosios nuosakos (kuri rašytinėje kalboje gana retai vartojama), nemažai šaukšmininkų, jaustukų, išiktukų, o moksliniame tekste būtų keista rasti emocinių ekspresinių kalbos dalių, čia turėtų dominuoti daiktavardžiai.

Įdomių dalykų galima rasti nagrinėjant ir kitokius tekstus. Tarkim, kokie tekstai išsiskiria aukštesniojo ir aukščiausiojo laipsnio formomis. Gal komercinės reklamos? Juk viskas ten geriausia, gražiausia, skaniausia...

Atlikti tokių tyrimų visai nesudėtinga: galima pasinaudoti Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centre sukurtu morfologiniu anotatoriumi, t. y. programa, kuri automatiškai gali nurodyti kalbos dalį ir gramatines formas (programą rasite Kompiuterinės lingvistikos centro internetinėje svetainėje <http://donelaitis.vdu.lt/> dalyje „Darbai“).

Jei sumanysite atlikti gramatinių formų analizę, tarkim, pasirinktame kūrinyje, palyginkite duomenis su rezultatais, gautais ištyrus didesnės apimties medžiagą, t. y. su tekstyno rezultatais, aprašytais šiame straipsnyje ar kituose literatūros sąrašė nurodytuose darbuose. Pamatysite, kokių skirtumų ir panašumų išryškėja.

Kiekybinių ir kokybinių morfologijos tyrimų sąsajos

Morfologija apima ne tik žodžių skirstymą į morfemas, kalbos dalis, morfologinių požymių nustatymą, bet ir kalbos dalių klasifikavimą į tam tikras grupes (asmenuotes, linksniuotes). Todėl galimas ir tokio pobūdžio kiekybinis tyrimas.

Ankstesniame straipsnyje buvo rašyta, kad kiekybiniai tyrimai gana dažnai padeda atskleisti kokybinius kalbos aspektus. Vienas iš tokių kiekybinių ir kokybinių tyrimo metodų sąsajos pavyzdžių yra daiktavardžių paradigmų analizė (Savickienė, Kazlauskienė, Kamandulytė, 2004). Šiame tyrime, remiantis natūraliosios morfologijos teorinėmis koncepcijomis, naujai aprašyta daiktavardžio klasių (mokyklinėse gramatikose įprastas *linksniuotės* terminas) sistema, atsižvelgiama ne tik į galūnes, bet ir į akcentinius (kirčiuotes) bei morfonologinius (kamieno paskutinio priebalsio minkštumą, kietumą) kitimus morfemose, remiamasi ir kiekybiniais duomenimis.

Toliau pateikiama lentelė iš šio straipsnio, kur gerai matyti sąsajos tarp kirčiuočių ir daiktavardžio galūnių (duomenys rinkti iš *Dabartinės lietuvių kalbos žodyno*, 4 leidimo). Pavyzdžiui, nustatyta, kad daiktavardžiams, turintiems galūnes *-as*, būdingiausias pirmoji (30 proc.) ir antroji kirčiuotės (61 proc.).

Daiktavardžių pasiskirstymas pagal klases ir kirčiuotes

Klasė	Vns. vard. galūnė		Kirčiuotė							
			1	2	3	3 ^a	3 ^b	3 ^{4a}	3 ^{4b}	4
-(i)a, -i 5385 (21 %)	-a	mot.	1571	635	19	58	369	16	1	405
		vyr.	1	11	0	0	0	0	0	0
		bendr.	207	32	3	1	6	0	0	16
	-ia, -ja	mot.	1821	114	0	0	22	0	0	43
		-i	mot.	0	0	0	0	0	0	0
-ė 6031 (23 %)	-ė	mot.	2668	2895	14	30	125	14	1	202
		vyr.	0	4	0	0	0	0	0	0
		bendr.	2	6	0	0	7	0	0	4
-is 290 (1 %)	-is	mot.	50	2	7	10	76	1	0	99
		vyr.	1	1	2	6	10	0	1	3
		bendr.	0	0	0	0	3	0	1	1
-(i)as, -is/ys 13893 (53 %)	-as	vyr.	1907	3499	90	35	393	48	5	340
	-ias, -jas	vyr.	921	16	1	0	0	0	0	6
	-is	vyr.	3085	2188	0	0	0	0	0	0
	-ys	vyr.	0	0	16	192	581	48	25	282
-(i)us 523 (2 %)	-ius, -jus	vyr.	186	314	0	0	1	0	0	2
	-us	vyr.	0	7	2	0	0	0	0	10
-uo	-uo	vyr.	0	0	0	23	38	1	0	0
Iš viso			12420	9724	154	355	1631	128	34	1415

Suprantama, atlikti tokį tyrimą mokiniams būtų gana sudėtinga. Tačiau su jais galima pabandyti nustatyti, kokių linksniuočių daiktavardžiai vyrauja vartosenoje. Taip mokiniai įtvirtintų ir linksniuočių nustatymo įgūdžius, ir atliktų savarankišką tyrimą.

Statistiniai kirčiuočių duomenys

Kitas su kirčiavimu susijęs tyrimas yra kirčiuočių produktyvumas. Statistiniai kirčiavimo paradigmų (arba kirčiuočių, akcentinių paradigmų) produktyvumo duomenys labai svarbūs norint išsiaiškinti kirčiavimo kitimo tendencijas. Lietuvių kalbos kirčiuočių dažnumas jau tyrinėtas: Bonifacas Stundžia (1984) nustatė tokius lietuvių kalboje esančių dviskiemenių kirčiavimo atžvilgiu nesvyruojančių (t. y. neturinčių kirčiavimo gretybių) daiktavardžių kirčiuočių santykius: 3 k. : 2 k. : 1 k. : 4 k. = 1 : 3 : 4 : 8 (mažiausiai yra 3 kirčiuotės žodžių, o 4 kirčiuotės žodžių yra aštuonis kartus daugiau nei 3, santykiai gaunami didesnę kiekį padalijus iš mažesnio). Tie santykiai labai pasikeičia, kai pridedami ir daugiaskiemeniai žodžiai. Vidos Žilinskienės duomenys iš *Atgalinio dabartinės lietuvių kalbos žodyno* (1995) rodo, kad mažiausiai yra 4 kirčiuotės, o daugiausiai 1 kirčiuotės žodžių: 4 k. : 3 k. : 2 k. : 1 k. = 1 : 2 : 11 : 15.

Tokia situacija yra žodyne. Tačiau vieni žodžiai vartojami dažniau, kiti rečiau, todėl vartosenos tyrimo rezultatai gali nesutapti su tais, kurie gauti remiantis žodynu. Astos Kazlauskienės (2000) tyrimas rodo, kad lietuvių kalboje dažniausiai vartojami 1 kirčiuotės žodžiai (42 %), kiek rečiau 2 kirčiuotės (38 %), o 3 ir 4 kirčiuočių žodžiai gana reti (sudaro po 10 %). Kirčiuočių santykiai tokie: 3 k. : 4 k. : 2 k. : 1 k. = 1 : 1 : 4 : 4. Tad kiekybinio tyrimo duomenys rodo, kad ir žodyne, ir vartosenoje vyrauja 1 ir 2 kirčiuočių žodžiai, t. y. patys lengviausi kirčiavimo požūriū žodžiai.

Tokius tyrimus siūlyti mokiniams gana sunku, nes pats kirčiavimas vidurinėje mokykloje yra vienas iš sudėtingesniū dalykū. Tačiau jie gali pabandyti nustatyti, kuriame skiemenyje dažniausiai kirčiuojami lietuviū kalbos žodžiai. Tam tereikia pasirinkti teksto fragmentą, sužymėti kirčiuotus skiemenis, juos suskaičiuoti ir apibendrinti rezultatus. Pasitikrinti, ar tinkamą skiemenį pažymėjo, mokiniai gali internetinėje kirčiuoklėje, kurią rasite Kompiuterinės lingvistikos centro internetinėje svetainėje (<http://donelaitis.vdu.lt/>, dalyje „Darbai“).

Baigiamosios pastabos

Čia pateikėme tik kelias gaires galimiems darbams ir nuorodas į ankstesnius tyrimus, kuriais mokiniai galėtų remtis pradėdami savo tyrimą. Vienus iš šiū darbū atlikti nesunku, kitiems reikės gilesniū lietuviū kalbos žinių ir tvirtesniū įgūdžių. Tačiau kiekybiniai tokiū tyrimū aspektai visada tie patys: reikia pasirinkti tiriamąją imtį, ją atitinkamai parengti, suskaičiuoti, kiek yra tiriamųjų vienetū, ir aptarti rezultatus. Kita vertus, rengiant tyrimui medžiagą ir apibendrinant duomenis galima ne tik paskatinti mokinius atidžiau pažvelgti į lietuviū kalbos faktus, bet ir įtvirtinti lietuviū kalbos gramatikos, leksikos ar kirčiavimo įgūdžius.

Literatūra

1. *Atgalinis dabartinės lietuviū kalbos žodynas* (sud. V. Žilinskienė). Vilnius: Matematikos ir informatikos institutas, 1995.
2. *Kazlauskienė A.* 2000: Daiktavardžio kirčiuočių dažnumas // *Darbai ir dienos*, 2000, t. 24, p. 83–88.
3. *Rimkutė E.* 2006: Morfologinio daugiareikšmiškumo ribojimas kompiuteriniame tekстыne. Daktaro disertacija, Vytauto Didžiojo universitetas.
4. *Rimkutė E.* 2006: Lietuviū kalbos gramatinių formū vartoseną morfologiškai anotuotame tekстыne // *Lituanistica*, 2 (66), p. 34–55.
5. *Savickienė I.* 2005: Linksnū vartojimo dažnumas ir daiktavardžių reikšmė // *Acta Linguistica Lituanica*, p. 52, 59–65.
6. *Savickienė I., Kazlauskienė A., Kamandulytė L.* 2004: Naujas požūris į lietuviū kalbos daiktavardžio linksnivimo tipus pagal natūraliosios morfologijos teoriją // *Acta Linguistica Lituanica*, t. 50, p. 79–98.
7. *Stundžia B.* 1984: Lietuviū kalbos dviskiemeniū daiktavardžių kirčiavimo variantai // *Kalbotyra*, t. 35 (1), p. 86–92.
8. *Žilinskienė V.* 2007: Veiksmažodžio formū vartojimas lietuviū ir latviū kalbū skirtingū funkcinių stiliū tekстыse // *Baltū filologija*, p. 139–151.