

SAKYTINĖS LIETUVIŲ KALBOS TEKSTYNAS: TARMINIŲ FORMŲ VARTOSENA

GENOVAITĖ KAČIUŠKIENĖ

Šiaulių universitetas

geno@splius.lt

Pagrindiniai žodžiai: *tekstynas, sakytinė kalba, tarminės formos.*

Sakytinės lietuvių kalbos tekstynas. Sakytinės kalbos tekstynai daugelyje pasaulio šalių egzistuoja jau senokai. Gerai žinomi anglų, vokiečių, rusų, ispanų, olandų, estų ir kt. kalbų sakytinės kalbos tekstynai, apimantys nuo kelių tūkstančių iki kelių milijonų žodžių (Dabašinskienė 2009a, 9). Senokai galvota ir apie analogišką lietuvių kalbos tekstyną. Tačiau tik 2006–2008 metais, gavus Valstybinio mokslo ir studijų fondo finansinę paramą, tekstyno rengimo darbai pajudėjo. Šio tekstyno iniciatoriai buvo VDU mokslininkai, vadovaujami doc. dr. Inetos Dabašinskienės. Į sakytinės lietuvių kalbos tekstyno kūrimo darbus buvo pakviesti ir įsitraukė LKI, KU, ŠU ir VU mokslininkai¹. Tekstyną² sudaro 230 val. trukmės sakytinės kalbos, apimančios spontanišią šnekamąją ir parengtą viešąją kalbą, įrašai, iš viso apie 680 000 morfologiškai anotuotų žodžių³.

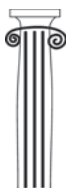
Sakytinės lietuvių kalbos tyrimai. Parengus „Sakytinės lietuvių kalbos tekstyną“, suaktyvėjo lietuvių šnekamosios kalbos tyrimai⁴. Remiantis tekstyno medžiaga parašyti ir atiduoti spaudai keli straipsniai, kuriuose aptariamos šnekamosios kalbos leksinės, morfologinės ypatybės (Dabašinskienė 2009d), gramatinių formų trumpinimo ir vartojimo dažnumo santykis (Dabašinskienė 2009c), deminutyvų vartoseną (Dabašinskienė 2009b), būdvardžio vartosenos skirtumai įvairiuose sakytinės kalbos registruose (Kamandulytė, Tuškevičiūtė 2009), klausiamieji pasakymai (Balčiūnienė, Kunigėlytė 2009) ir kt. Parengtų straipsnių pagrindu galima teigti, kad sakytinėje

¹ Projekte dalyvavo du Šiaulių universiteto Lietuvių kalbotyros ir komunikacijos katedros dėstytojai – Robertas Gedrimas ir Genovaitė Kačiuškienė.

² Internetinė prieiga sukurta projekto vykdytojo Vytauto Didžiojo universiteto Regionistikos katedros tinklalapyje <http://www.vdu.lt/LTcourses/> → *Mokslas*. Internetinė prieiga suteikia galimybę naudotis per 220 000 žodžių sakytinės lietuvių kalbos tekstynu, kurį sudaro spontanišios ir parengtos viešosios kalbos įvairių situacijų monologai, dialogai, polilogai.

³ Išsamiau apie tekstyno struktūrą, metodologiją bei jo rengimo eigą žr. Dabašinskienės darbus, pateiktus habilitacijos procedūrai teikiamų mokslo darbų apžvalgoje (2009a, 34–37).

⁴ Straipsnyje remiamasi I. Dabašinskienės habilitacijos procedūrai (VDU, 2009 03 06) pateiktų mokslinių darbų atspaudais ir projekto „Sakytinės kalbos tekstyno kūrimas“ rankraštines baigiamosios ataskaitos (2009) duomenimis.



kalboje, kitaip nei rašytinėje, labai dažnai vartojami trumpi ir sutrumpinti žodžiai, atliekantys jungimo, savęs identifikavimo, nurodymo funkcijas. Dažniausiai vartojami sakytinės kalbos žodžiai yra viensiemeniai: *aš, nu, ir, ne, bet, kad, tu, ten, ką, tai, mes, kas, ar*. Kalbos dalimis dažniausiai vartojami žodžiai pasiskirsto taip: daiktavardžiai (*kartas, diena, laikas, vieta*), būdvardžiai (*geras, gera, graži*), veiksmažodžiai (*nežinau, žinau, yra, nori, būt, nėra, pasakyt*), įvardžiai (*aš, tu, mes, tas*),rieveiksmiai (*ten, čia*). Šnekamosios kalbos morfologinė analizė parodė, kad kalbos dalys čia pasiskirsto kitaip nei rašytinėje kalboje. Dažniausia sakytinėje kalboje vartojama kalbos dalis yra veiksmažodis, o rašytinėje kalboje – daiktavardis. Skiriasi ir linksnių vartojimo dažnumas – dažniausias daiktavardžio linksnis sakytinėje kalboje yra vardininkas, o rašytinėje kalboje – kilmininkas. Taip pat pastebėta, kad vienos kurios nors kalbos dalies vartojimo dažnumą sakytinėje kalboje dažniausiai lemia įvairūs sakytinės kalbos registrai – pokalbio aplinka, forma, socialiniai dalyvių vaidmenys ir kt. Nuo kalbos registro priklauso net ir gramatinių kategorijų pasiskirstymas. Kiek skiriasi ir sakytinės kalbos klausiamųjų sakinių struktūra – čia gana dažni klausiamieji sakiniai su sakinio pabaigoje einančiomis klausiamosiomis dalelytėmis, kurios rašytinėje kalboje paprastai vartojamos žodžio pradžioje.

Tarminės formos „Sakytinės lietuvių kalbos tekстыne“. Transkribuojant sakytinės kalbos tekstus, pastebėta įvairių tarminių formų, gana dažnai vartojamų šnekamojoje kalboje⁵. Iki šiol sakytinės kalbos tekстыne vartojamos tarminės formos nėra tyrinėtos jokių aspektu. Todėl vienas iš šio straipsnio tikslų ir yra bent iš dalies užpildyti šią spragą, trumpai aptariant dažniausiai tekстыne pasitaikančias tarmines formas.

Dažniausios tarminės formos (po 5 iš kiekvienos dalies) išrinktos iš visų specialiai parengtos ar spontaniškos kalbos tekstų, pateiktų 6 pagrindiniuose tekстыno skyriuose (arba sakytinės kalbos variantuose):

- akademinė komunikacija
- žiniasklaidos komunikacija
- privati komunikacija telefonu
- privati tiesioginė komunikacija
- pusiau vieša tiesioginė komunikacija
- pusiau vieša komunikacija telefonu⁶.

Visame tekстыne pateikiama 226174 žodžiai, iš kurių 1227, arba 0,54% visų žodžių, yra tarminės formos, turinčios minėtą nuorodą @d. Viso tekстыno pirmasis dažniausių žodžių penketukas atrodo taip: *saka*81⁷, *buva*27, *reik*20, *būn*19, *viens*17.

Tarminės formos parengtos kalbos tekстыne. Parengtos kalbos tekстыne, kurį sudaro akademinė ir žiniasklaidos komunikacija, pateikiama 110674 žodžių, iš kurių 60, arba 0,05% visų žodžių, pavartoti tarminėmis formomis. Dažniausi iš jų: *reik*3, *a*3, *ka*3, *dunsta*2, *visks*2, pvz.: <...> *ir sugalvojo, kad reik@d* [: *reikia*] *pasisukti*;

⁵ Šios formos tekстыne pažymėtos nuoroda @d ir vadinamos dialektizmais.

⁶ Išsamiau apie sakytinės kalbos variantus žr. (Dabašinskienė 2009a, 10–11; Dabašinskienė, Kamandulytė, 2009).

⁷ Čia ir toliau šalia žodžio esantis skaičius rodo pavartojimo dažnumą.

<...> **a@d** [: ar] tai čia kiti padorūs piliečiai ? <...> pasirodė, **ka@d** [: kad] prieš pabaigą.

Akademinės komunikacijos tekstuose (30043 žodžių) rastos 29 tarminės formos, sudarančios 0,01% pavartojimo atvejų: *dunsta2, prisibudinka1, būtumb1, motula1, parka1*, pvz., <...> *ir tarkime, kaip pavyzdį galiu duoti, jeigu tos pačios šeimos atstovė vyresniosios kartos sako formą **dunsta@d** [: dūsta], <...> o ar tada kupiškėnų skirtingos formos **būtumb@d** ir **būtumai@d**?*

Žiniasklaidos komunikacijoje iš bendro 80631 žodžių skaičiaus 31 žodis, arba 0,04%, turėjo tarminę formą. Tai – *reik3, a3, ka2, kokia2, visks2*, pvz.: <...> **kokia@d** [: kokio] *amžiaus esi?* <...> tai **visks@d** [: viskas] *normaliai, pasitikrinau.*

Tarminės formos spontaninės kalbos tekстыne. Spontaninės kalbos tekстыne, kurį sudaro privati ir pusiau vieša komunikacija telefonu, privati ir pusiau vieša tiesioginė komunikacija, dažniausių tarminių formų penketukas yra toks pat, kaip ir viso sakytinės kalbos tekstyno – *saka81, buva27, reik19, būn17, viens16*. Beveik sutampa ir jų vartojimo dažnumas, tik paskutinių trijų žodžių pavartojimų skaičius nežymiai skiriasi. Iš 115500 spontaninės kalbos tekstyno žodžių 1167, arba 1,01%, turi tarmines formas. Lyginant su parengtos kalbos tekstynu, matyti, kad čia tarminės formos vartojamos kur kas dažniau.

Privačios komunikacijos telefonu metu užrašyta 8249 žodžiai, iš kurių 187, arba 2,27%, turėjo tam tikrų tarminių ypatybių. Dažniausios tarminės formos pasiskirsto taip: *nieka5, buva5, saki5, saka5, Egliki5*, pvz.: <...> *tu nesakyk jam **nieka@d** [: nieko], <...> **Eglikid@d** [: Eglikė] ant **jūku@d** [: juokais] **pasaki@d** [: pasakė], <...> **buva@d** [: buvo] *jau terbas **susikruovus@d** [: susikrovusi] Ingridžijanka.**

Iš pusiau privačios komunikacijos telefonu 3079 žodžių tarminių formų neužfiksuota.

Privačiai tiesioginei komunikacijai priskiriami 82714 žodžių, iš kurių 962, arba 1,16%, turi tarmines formas. Tai – *saka76, buva22, būn18, reik17, viens16*, pvz.: <...> *kai **būn@d** [: būna] gerai padaryta, <...> **viens@d** [: vienas] miręs.*

Pusiau privačios tiesioginės komunikacijos tekстыne užrašyta 21 458 žodžiai, iš kurių tik 18, arba 0,08 %, turi tarmiškas formas. Kaip dažniausi, išskiriami šie žodžiai: *kainuoj3, a2, ans2, mum 2, didelios1*, pvz., <...> *jis apie pusantrą tūkstančio **kainuoj@d** [: kainuoja] gal ?, <...> aš nežinau, gal **ans@d** [: anas] čia buvo?*

Preliminarios baigiamosios išvados.

1. Išanalizavus „Sakytinės lietuvių kalbos tekstyno“ tarmines formas, tekстыne pažymėtas **@d**, matyti, kad jos buvo pavartotos 1227 kartus, o tai sudaro 0,54% visų užfiksuotų žodžių.

2. Spontaninės kalbos tekстыne tarminės formos vartojamos daugiau kaip 20 kartų dažniau negu parengtos kalbos tekстыne (plg. 1,01% ir 0,05%).

3. Didžiausias tarminių formų procentas pavartotas privačiai kalbantis telefonu (2,27%) arba tiesioginiuose asmeniniuose pokalbiuose (1,16%), kada informantas nėra varžomas nei aplinkos, nei bendrinės kalbos taisyklių.



4. Palyginti mažai tarminių formų vartojama pusiau viešoje tiesioginėje komunikacijoje (0,08%), žiniasklaidos (0,04%) ir akademinėje komunikacijoje (0,04%), kai informantas privalo laikytis lietuvių bendrinės kalbos normų.

5. Pusiau viešai kalbantis telefonu tarminių formų nepavartota.

6. Dažniausiai vartojamos spontaninės kalbos tarminės formos *saka*, *buva*, *reik*, *būn*, *viens* sutampa su analogiškėmis viso „Sakytinės lietuvių kalbos tekstyno“ formomis.

7. Dažniausia „Sakytinės lietuvių kalbos tekstyno“ tarminė forma – 81 atvejis – yra *saka*.

8. Kaip ir lietuvių bendrinėje kalboje, į dažniausiai vartojamų sakytinės kalbos žodžių gretą patenka vienskiemeniai žodžiai *a* (ar), *ka* (kad), veiksmažodis *yr* (yra).

9. Straipsnyje aptarti dažniausiai vartojamų tarminių formų atvejai (išskyrus žemaitišką *ans* ar panevėžietišką *visks*), nesuteikia jokios informacijos nei apie informantų tarmes, nei apie konkrečias tarmių ypatybes, nes užfiksuotos tarminės formos – kokybinis bei kiekybinis žodžio galo trumpinimas (*buva*, *būn*, *viens*, *yr*) ir priebalsių nykimas jungtukuose ar vienskiemeniuose žodžiuose (*a*, *da*) – yra būdingas beveik visoms lietuvių kalbos tarmėms (Zinkevičius 1966, 111–121, 192–199). Norint gauti išsamesnį tarminių ypatybių vaizdą, reikia atlikti detalią visų tekстыne vartojamų tarminių formų analizę.

Literatūra

- Balčiūnienė I., Simonavičienė L. Klausiamieji pasakymai šnekamojoje kalboje. *Lietuvių kalba*. 2009 (atiduota spaudai).
- Dabašinskienė I. Sakytinė lietuvių kalba: sociolingvistiniai ir psicholingvistiniai tyrimai. Habilitacijos procedūrai teikiamų mokslo darbų apžvalga. Kaunas: Vytauto Didžiojo universiteto leidykla, 2009a.
- Dabašinskienė I. Intimacy, familiarity and formality: a case of diminutives in modern Lithuanian. *Lituanus*. 2009b (atiduota spaudai).
- Dabašinskienė I. Trumpinimas ir dažnumo poveikis šnekamojoje kalboje. *Darbai ir dienos*, 2009c (atiduota spaudai).
- Dabašinskienė I. Šnekamoji lietuvių kalba: leksinės ir morfologinės ypatybės, *Acta Linguistica Lituanica*. 2009d (atiduota spaudai).
- Dabašinskienė I., Kamandulytė L. Corpora of Spoken Lithuanian, *Estonian papers in applied linguistics*. Tallinn: Eesti Keele Sihtasutus, 2009 (atiduota spaudai).
- Kamandulytė L., Tuškevičiūtė M. Būdvardžio vartojimo skirtumai sakytinės kalbos registruose. *Darbai ir dienos*. 2009 (atiduota spaudai).
- Zinkevičius Z. *Lietuvių dialektologija*. Vilnius: Mintis, 1966.
- www.donelaitis.vdu.lt [žiūrėta 2009 08 19–20].
- <http://www.vdu.lt/LTcourses/> → *Mokslas* [žiūrėta 2009 04 19–20, 2009 08 19–20].



Genovaitė Kačiuškienė

**SAKYTINĖS LIETUVIŲ KALBOS TEKSTYNAS:
TARMINIŲ FORMŲ VARTOSENA**

Santrauka

Pagrindiniai žodžiai: *tekstynas, sakytinė kalba, tarminės formos.*

Straipsnyje aptariamas 5 mokslo institucijų (tarp jų ir Šiaulių universiteto) 2006–2008 metais parengtas „Sakytinės lietuvių kalbos tekstynas“ bei jo pagrindu parašyti moksliniai straipsniai.

Remiantis šio tekstyno medžiaga <http://www.vdu.lt/LTcourses/> → *Mokslas* analizuojamos dažniausiai pasitaikančios tarminės formos, tekстыne žymimos nuoroda @d. Aptariamos dažniausios tarminės formos, užfiksuotos akademinės komunikacijos, žiniasklaidos komunikacijos, privačios komunikacijos telefonu, privačios tiesioginės komunikacijos, pusiau viešos tiesioginės komunikacijos bei pusiau viešos komunikacijos telefonu skyriuose.

Analizuojamos medžiagos pagrindu daromos tokios išvados: 1) Iš 226174 tekstyne pavartotų žodžių tarmiškas formas turi 1227 žodžiai, kurie sudaro 0,54% visų užfiksuotų žodžių. 2) Spontantinės kalbos tekstyne tarminės formos vartojamos daugiau kaip 20 kartų dažniau negu specialiai parengtos kalbos tekstyne (plg. 1,01% ir 0,05%). 3) Didžiausias tarminių formų procentas pavartotas privačiai kalbantis telefonu (2,27%) arba tiesioginiuose asmeniniuose pokalbiuose (1,16%), kada informantas nėra varžomas nei aplinkos, nei lietuvių bendrinės kalbos taisyklių. 4) Dažniausia „Sakytinės lietuvių kalbos tekstyno“ tarminė forma – 81 atvejis – yra *saka*.

Genovaitė Kačiuškienė

**THE CORPUS OF THE SPOKEN LITHUANIAN LANGUAGE:
USAGE OF DIALECTAL FORMS**

Summary

Keywords: *corpus, spoken language, dialectal forms.*

The article presents “Corpus of the Spoken Lithuanian Language” prepared in the years 2006–2008 by five education institutions (including Šiauliai University) and scientific articles based on this corpus.

According to the material of the corpus (<http://www.vdu.lt/LTcourses/> → *Mokslas*) dialectal forms marked as @d in the corpus are analysed. The most frequent dialectal forms recorded in the departments of academic communication, news media communication, private communication by telephone, private direct communication, semi-public direct communication and semi-public communication by telephone are discussed.

Filologija 2009 (14)



The following conclusions based on the material under analysis are drawn: 1) 1,227 words of all 226, 174 words used in the corpus have dialectal forms. It makes up 0.54% of all recorded words; 2) dialectal forms used in the corpus of the spontaneous language are 20 times more frequent than in the specially prepared language corpus (compare 1.01% and 0.05%); 3) the peak per cent of dialectal forms is used privately speaking on the telephone (2.27%) or during direct private conversations (1.16%) when an informant is limited neither by the environment nor by the rules of standard Lithuanian; 4) the most frequent dialectal form of “*Corpus of the Spoken Lithuanian Language*” is *saka* (81 cases).

