



# 11 klasių mokinių gimtosios lietuvių kalbos žodyno dydis

Rugilė Miškinytė<sup>1</sup>, Loreta Vilkienė<sup>2</sup>

<sup>1</sup> Vilniaus universitetas, Universiteto g. 5, 01131 Vilnius, [rugile.miskinytee@gmail.com](mailto:rugile.miskinytee@gmail.com)

<sup>2</sup> Vilniaus universitetas, Taikomosios kalbotyros institutas, Lituanistinių studijų katedra, Universiteto g. 5, 01131 Vilnius, [Loreta.Vilkienė@ff.vu.lt](mailto:Loreta.Vilkienė@ff.vu.lt)

**Anotacija.** Straipsnyje aprašomas tyrimas, kurio tikslas – išsiaiškinti, koks yra kelių mažų Lietuvos miestų 11 klasių mokinių gimtosios lietuvių kalbos žodyno dydis, be to, siekiama jį palyginti su Vilniaus mokinių rezultatais ieškant atsakymo į klausimą, ar gyvenamoji vieta daro įtaką mokinių žodyno dydžiui. Taip pat pateikiama įžvalgų, kurie sociolingvistiniai veiksniai daro įtaką mokinių žodyno dydžiui. Tyrimui pasirinkta patogioji imtis. Analizuojant duomenis atliktas T testas ir regresinė analizė.

**Esminiai žodžiai:** žodyno dydis, gimtakalbis, lietuvių kalba, 11 klasės mokinys, lyties veiksnys.

## Įvadas

Kas trejus metus EBPO PISA (angl. *Programme for International Student Assessment*) vykdo tarptautinį penkiolikmečių tyrimą, kuriame tikrinamos įvairios mokinių kompetencijos: skaitymo gebėjimai, matematinis ir gamtamokslinis raštingumas. Mokinių pasiekimai analizuojami keliais pjuviais. Šie tyrimai vykdomi ir Lietuvoje (plačiau žr. PISA 2018). Pavyzdžiui, tokia 2018 m. PISA tyrime iš viso dalyvavo 364 Lietuvos mokyklos, 6 885 mokiniai. Atlikus šį tyrimą, buvo pastebėta, kad ne mažiau kaip pusės didelių miestų mokinių skaitymo rezultatai yra geresni už Lietuvos vidurkį, o penkiolikmečių merginų skaitymo gebėjimai yra statistiškai reikšmingai aukštesni nei vaikinų (PISA 2018). Atsakymo, kodėl taip yra, PISA tyrėjai nepateikia. Tai paskatino šio straipsnio autorės iškelti probleminį klausimą, ar minėti skaitymo gebėjimų patikrinimo rezultatai gali būti susiję su konkrečios lyties mokinių žodyno dydžiu? Tokį klausimą suformuluoti leido įvairių mokslininkų, pavyzdžiui, Paul'o Nation'o (2001, 144), James'o Milton'o ir

Jeanine Treffers-Daller (2013) darbai, kurie rodo, kad skaitymo gebėjimai yra glaudžiai susiję su žodyno dydžiu. Kitaip tariant, skaitymas padeda plėsti žodyną, o kuo didesnis žodynas, tuo geresni žmogaus skaitymo gebėjimai. Paprastai analizuojant žodyno dydį lyties aspektas yra įtraukiamas kaip stebimas veiksnys, bet aiškių, vienareikšmių atsakymų į klausimą, ar jis veikia žodyno dydį, nepateikiama: vienų mokslininkų darbuose nėra išvelgiama lyties veiksnio įtaka žodyno dydžiui (Keuleers et al., 2015), o kai kuriuose – priešingai – randama įrodymų, kad žodyno dydis susijęs su respondentų lytimi (Vilkienė et al., 2019; Andriuškevičiūtė ir Vilkienė, 2019). Žodyno dydžio ir gyvenamosios vietos – didelių ar mažų miestų, miestelių – priklausomybė nėra tirta. Visa tai paskatino šio straipsnio autorės imtis tyrimo ir atsakyti į klausimą, koks žodyno dydis yra Lietuvos mokinių – merginų ir vaikinų, gyvenančių sostinėje ir mažuose miestuose ar miesteliuose, nuo ko jis priklauso.

Pasakytina, kad ieškoti atsakymo į minėtą klausimą buvo pradėta nuo jau Lietuvoje atlikto Vilniaus mokinių žodyno dydžio tyrimo. Loreta Vilkienė, Laura Vilkaitė-Lozdienė ir Justina Bružaitė-Liseckienė (2019) yra vykdžiusios mokslinį projektą, kurio tikslas buvo vienuoliktų klasių mokinių lietuvių kalbos kokybės ir nuostatų, mokymosi motyvacijos atskleidimas. Minėtos mokslininkės analizavo 6 Vilniaus gimnazijų (po dvi lietuvių, lenkų ir rusų ugdomosiomis kalbomis) duomenis. Tyrime dalyvavo 127 mokiniai iš mokyklų lietuvių ugdomąja kalba. Vienas iš tyrimo aspektų buvo žodyno dydžio matavimas. Atlikus šį tyrimą buvo nustatyta, kad vidutinis Vilniaus miesto gimtakalbio, kurio amžius maždaug 17 metų, lietuvių kalbos žodynas – maždaug 52 800 lemų. Daugiau mokinių žodyno dydžio tyrimų Lietuvoje nesama. Tačiau pabrėžtina, kad buvo tirti tik Vilniaus mokiniai. Tai paskatino šio straipsnio autorės imtis konceptualaus (angl. *constructive, conceptual*)<sup>1</sup> Vilkienės et al. (2019) tyrimo pakartojimo ne sostinės mokyklose. Toks konceptualus pakartojimas leidžia patikslinti duomenis ir juos analizuoti kitu pjūviu, o šiame straipsnyje aprašomo tyrimo atveju – ieškoti atsakymo, ar gyvenamoji vieta – sostinė ir nedidelis miestas – gali daryti įtaką žodyno dydžiui.

Tyrime pasirinkti analizuoti patogiosios imties duomenys, rinkti 4 mažuose Lietuvos miestuose. Šiame straipsnyje remiamasi Lietuvos Respublikos Teritorijos administracinių vienetų ir jų ribų įstatymu, kuriame miestu vadinama „kompaktiškai užstatyta gyvenamoji vietovė, kurioje gyvena 3000 ir daugiau gyventojų“, o miesteliu – „kompaktiškai

<sup>1</sup> Šiuo metu pasaulyje taikomosios kalbotyros mokslininkai plačiai kalba apie tai, kad kiekybiniai tyrimai turi būti pakartojami. Tokie pakartojimai gali būti tikslūs, apytiksliai ir konceptualūs. Šiame straipsnyje aprašomas tyrimas yra konceptualus pakartojimas, t. y. jis „tikrina pradinį tyrimą, bet taip pat gali tikrinti siūlomus rezultatų paaiškinimus, interpretacijas. <...> konceptualiame pakartojime galima keisti neesminius kintamuosius ir tai, kaip jie matuojami, norint įvertinti pateiktus paaiškinimus“ (Vilkienė et al., 2020, 35). Taigi konceptualiai pakartojamas Vilkienės et al. (2019) tyrimas, kuriame pakeista viena tiriamųjų charakteristika – gyvenamoji vieta. Plačiau apie tyrimų pakartojimus žr. Vilkienė, Vilkaitė-Lozdienė, Juknevičienė, Bružaitė-Liseckienė, Geben, Ryvitytė, 2020; taip pat *Replication studies in language learning and teaching*, 2008.

užstatyta gyvenamoji vietovė, kurioje gyvena nuo 500 iki 3000 gyventojų<sup>2</sup>. Pasakytina, kad straipsnyje aprašomame tyrime sutiko dalyvauti trys miestų (gyventojų skaičius 11 262, 5 690, 3 215) mokyklos ir viena mokykla, esanti miestelyje (gyventojų skaičius 2 046). Tačiau aprašant duomenis pasirinkta visą imtį vadinti mažų miestų imtimi, lyginant ją su didelio miesto – sostinės – tiriamaisiais.

Taigi šiame straipsnyje aprašomo tyrimo *tikslas* – išsiaiškinti, kokio dydžio yra ne sostinės, o mažų Lietuvos miestų 11 klasių mokinių gimtosios lietuvių kalbos žodynas, jį palyginti su Vilniaus gimnazijose atlikto tyrimo (Vilkienė et al., 2019) rezultatais ieškant atsakymo į klausimus, ar gyvenamoji vieta daro įtaką žodyno dydžiui. Be to, kelti tokie *uždaviniai*: 1) patikslinti, koks yra apie 17 m. amžiaus gimtakalbio lietuvių žodyno dydis; 2) kadangi PISA (2018) tyrimo rezultatai parodė, kad merginų skaitymo gebėjimai yra geresni nei vaikinų, patikrinti, ar ir kiek skiriasi mažų miestų merginų ir vaikinų žodyno dydis; 3) pateikti išvalgų, kurie sociolingvistiniai veiksniai iš esmės daro reikšmingą įtaką žodyno dydžiui.

## Žodynas: produkcija ir recepcija

Šiame straipsnyje žodynas suprantamas kaip visi kurios nors kalbos žmogaus mokami žodžiai (Murphy, 2010, 4). Mokslininkai žmogaus sukauptą žodyną linkę suvokti kaip dvejopą: produkcijos ir recepcijos. Kitaip tariant, yra teigiama, kad žmogus vienus žodžius labiau linkęs vartoti produkuodamas kalbą, o kitus – recepcinėje kalbinėje veikloje (Maskor, Baharudin, 2016). Vadinas, galima teigti, kad egzistuoja tarsi du žodynai: produkcijos ir recepcijos. Be abejo, abu jie yra vienodai svarbūs. Kalbos recepcijos veikla stimuliuojama tuomet, kada žmogus žodžius skaito ar supranta iš klausos. Kaip teigia Ingegerd Enström (2004) ir Inger Lindberg (2006), recepcijai reikalingame žodyne yra saugomi žodžiai, kurių reikšmę žmogus supranta ir atpažįsta tekste, tačiau pats tuos žodžius nebūtinai vartoja. Produkcinės veiklos žodynas apima tuos žodžius, kuriuos žmogus ir supranta, ir pats gali juos laisvai, spontaniškai pavartoti ir sakytinėje, ir rašytinėje kalboje (Enström, 2010). Taigi šis žodynas aktyvuojamas tuomet, kada žmogus pats kalba ar rašo (Lindberg, 2006). Be to, teigiama, kad žmogaus produkcinės veiklos žodynas yra mažesnis nei recepcijos, nes yra lengviau suprasti žodį ir jo reikšmę iš konteksto nei pačiam tinkamai jį pavartoti (Enström, 2010). Taigi abu žodynai – recepcijos ir produkcijos – yra svarbūs žmogui, kad jis gebėtų suprasti ir vartoti kalbą.

Pasakytina, kad mokslininkai tiria tiek recepcijai, tiek produkcijai reikalingą žodyną. Pavyzdžiui, yra atliktas tyrimas, kurio rezultatai rodo, kad recepcijos žodyno žinios įsisavinamos greičiau nei produkcijos ir kad besimokantieji antrosios kalbos

<sup>2</sup> Plačiau žr. <https://e-seimas.lrs.lt/portal/legalActPrint/lt?jfwid=5qig2spbl&documentId=82dd8bc0fd-4b11e5bf4ee4a6d3cdb874&category=TAP>

atskleidžia geresnius tos kalbos vartojimo gebėjimus, kai reikalaujama taikyti receptijos, o ne produkcijos žodyno žinias (Lindberg, 2007). Pasak Stuart'o Webb'o (2013), kai žmogus mokosi žodžių juos užrašydamas arba sakydamas, yra taikomos ir receptijos, ir produkcijos žinios. Teigiama, kad abi šios veiklos yra labai susijusios, veikia viena kitą žmogui plečiant žodyną. Tačiau įvairių kalbų tyrėjai rašo ir apie tai, kad receptijos ir produkcijos žodynai nėra vienodo dydžio. Pavyzdžiui, Francine Melka (1997) mano, kad skirtumas tarp receptijos ir produkcijos žodyno yra mažas, ir teigia, kad 92 proc. receptijos žodyno yra vartojama ir produktyviai. Kiti tyrimai rodo, kad 50–75 proc. receptijos žodyno yra mokama produktyviai (Laufer, Paribakht, 1998; Fan, 2000). Taigi galima suprasti, kad mokslininkų dar nėra pakankamai išsiaiškinta, koks produkcijos ir receptijos žodyno santykis. Vis dėlto abu šiuos žodynus išmatuoti nėra lengva ir nuolatosis ieškoma metodikos, kaip tai atlikti. Vis dėlto mokslininkų manoma, kad lengviau išmatuoti receptijos žodyną. Tai dažniausiai ir daroma, nes produkcijos žodyno dydžiui išmatuoti dar nėra sukurtų patikimų įrankių. Šiame straipsnyje taip pat dėmesys yra sutelkiamas į lietuvių kalbos gimtakalbio receptijos žodyną, būtent jis yra matuojamas specialiu testu.

## Žodyno tyrimai ir jų rezultatai

Kadangi šiame straipsnyje yra analizuojamas gimtakalbio žodyno dydis, aptartini ankstesni gimtakalbio žodyno dydžio tyrimai. Visų pirma, reikėtų paminėti, kad nemažai žodyno dydžio tyrimų buvo atlikta rezultatus skaičiuojant žodžių lizdais, į kuriuos įeina pagrindinė žodžio forma ir visos išvestinės formos (tos pačios šaknies žodžiai), kurios sudaromos su priešdėliais, priesagomis, galūnėmis (Jakaitienė, 2009, 75; Schmitt, 2010, 189). Tačiau ne visų mokslininkų toks žodyno dydžio skaičiavimas yra pripažįstamas. Teigiama, kad žmogus, besimokantis svetimosios kalbos, nebūtinai gali atpažinti kurią nors išvestinę žodžio formą, nors ir moka jo pagrindinę formą (Kremmel, 2016, 978; Vilkienė et al., 2019, 24). Be to, vienas žodžių lizdas gali būti didesnis, o kitas mažesnis, tad, jei žmogaus žodynas bus skaičiuojamas lizdais, gali būti neaišku, kiek žodžių iš tikrųjų moka asmuo. Todėl vėliau nuspręsta, kad patikimiau žodyno dydį skaičiuoti lemomis, t. y. pagrindinėmis žodžio formomis. Kadangi šiame darbe žodyno dydis taip pat yra skaičiuojamas lemomis<sup>3</sup>, toliau aptariami tik tie tyrimai, kuriuos atliekant buvo pasirinktas toks skaičiavimo vienetas.

---

<sup>3</sup> *Lema* ir *žodis* toliau šiame straipsnyje vartojami sinonimiškai.

1 lentelė  
*Žodyno dydžio tyrimų rezultatų apžvalga*

Tyrėjai, tyrimo metai	Tiriamųjų gimtoji kalba	Tiriamųjų amžiaus vidurkis	Žodyno vidurkis (lemomis)
Goulden, Nation, Read, 1990	anglų	daugiau nei 22 m.	17 200
Lindberg, 2006	švedų	5–6 m. (bepradedantys lankyti mokyklą)	9 000
Keuleers, Stevens, Mandera, Brysbaert, 2015	olandų	nuo 12 m. iki 72 m.	apie 13 m. – 29 000 apie 70 m. – 43 000
Brysbaert, Stevens, Mandera, Keuleers, 2016	anglų	20 m. 60 m.	42 000 48 200
Segbers, Schroeder, 2017	vokiečių	nuo 6 m. iki 14 m. nuo 18 m. iki 25 m.	5 900 73 000
Vilkienė, Vilkaitė-Lozdienė, Bružaitė-Liseckienė, 2019	lietuvių	17 m.	52 800
Andriuškevičiūtė, 2019	lietuvių	22 m. 49 m.	57 400 60 700
Pažereckaitė, 2020	lietuvių (I ir IV kurso anglų k. filologijos studentų)	19 m. 22 m.	57 100 60 300

Kaip matyti iš 1 lentelės, žodyno dydžio tyrimai buvo atlikti su įvairaus amžiaus žmonėmis. Pavyzdžiui, Švedijoje buvo analizuojamas žodyno dydis vaikų, kurie dar tik pradeda lankyti mokyklą (apie 5–6 m. amžiaus). Šiam tyrimui atlikti buvo naudotas testas, parengtas remiantis *OrdiL*<sup>4</sup> sulemuotu tekstynu, kuris sudarytas iš įvairių dalykų mokymo vadovėlių (Lindberg, Johansson, Kokkinakis, Järborg, Holmegaard, 2007). Šis tyrimas atskleidė, kad tokio amžiaus gimtakalbis švedas vidutiniškai moka apie 8 000–10 000 žodžių, o besimokant kasmet vaiko žodynas vidutiniškai pasipildo maždaug 3 000 žodžių (Lindberg, 2006). Šio tyrimo rezultatus galima palyginti su gautais analizuojant šiek tiek vyresnio amžiaus vokiečių kalbos gimtakalbių žodyno dydį (Segbers, Schroeder, 2017). Mokslininkai Jutta Segbers ir Sascha Schroeder (2017) testavo 6–14 m. vaikus ir paauglius. Šių tyrėjų testas buvo parengtas remiantis *childLex*<sup>5</sup> tekstynu, kuris yra sudarytas iš 500 vaikiškų vokiškų knygų tekstų. Šio tyrimo rezultatai atskleidė, kad tiriamųjų žodyno dydžio vidurkis yra maždaug 5 900 lemų. Pastebėtina, kad nors vaikai buvo vyresni nei minėtame Švedijoje atliktame tyrime, nustatytas mažesnis jų žodyno dydis. Šio straipsnio autorės mano, kad įtakos tam galėjo turėti žodyno dydžio testui

<sup>4</sup> Testą, kuris buvo naudojamas tyrime, galima rasti šiuo adresu: [https://gupea.ub.gu.se/bitstream/2077/20503/1/gupea\\_2077\\_20503\\_1.pdf](https://gupea.ub.gu.se/bitstream/2077/20503/1/gupea_2077_20503_1.pdf) (p. 165–170).

<sup>5</sup> *ChildLex* tekstyną galima rasti šiuo adresu: <http://alpha.dlexdb.de/pages/help/dbs/childlex/>.

sudaryti naudoti skirtingo pobūdžio tekstynai: vienas buvo susijęs su mokomąja literatūra (įvairūs mokyklose naudojami vadovėliai), o kitas – iš grožinės literatūros tekstų (vaikams skirtų knygų). Tame pačiame tyrime Segbers ir Schroeder'is (2017) analizavo ir suaugusių asmenų, kurių amžius buvo nuo 18 iki 25 m., žodyno dydį. Rezultatų analizė atskleidė, kad suaugusio gimtakalbio vokiečio žodyno dydžio vidurkis yra apie 73 000 lemų.

Žodyno dydis buvo testuotas ir Naujojoje Zelandijoje. Mokslininkai Robins Goulden, Paul'as Nation'as ir Johan'as Read'as (1990) atliko žodyno testą su 20 aukštąjį universitetinį išsilavinimą turinčių žmonių, kurių gimtoji kalba – anglų. Tiriamųjų amžius buvo apie 22 metus. Paminėtina, kad šis testas buvo sudarytas remiantis Webstr'io anglų kalbos žodynu (*Webster's Third New International Dictionary*, 1961). Tyrime dalyvavę studentai turėjo pažymėti tuos žodžius, kuriuos mokėjo, suprato. Dalyviams atlikus testą paaiškėjo, kad šio amžiaus žmonių žodyno dydis svyravo nuo 13 200 iki 20 700 žodžių, o vidurkis buvo apie 17 200 lemų. Matyti, kad vidurkis gerokai mažesnis palyginti su kitų kalbų rezultatais, kurių tiriamųjų amžius taip pat buvo apie 20 m. Šio straipsnio autorės spėja, kad galbūt tokį žodyno dydžio skirtumą lėmė Goulden et al. (1990) pasirinkta metodika, kuri skyrėsi nuo kitų minėtų tyrimų. Tačiau galima ir prielaida, kad galbūt skirtingų kalbų gimtakalbių žodynas yra skirtingo dydžio.

Marc'as Brysbaert'as, Michaël'as Stevens'as, Paweł'as Manderas ir Emmanuel'is Keuleers'as (2016) atliko tyrimą Jungtinėse Amerikos Valstijose su amerikiečiais anglų kalbos gimtakalbiais. Šio tyrimo rezultatai parodė, kad 20 m. amerikietis vidutiniškai moka 42 000 lemų, o 60 m. asmens žodyno vidurkis yra apie 48 200 lemų. Taigi šie mokslininkai atskleidė, kad per 40 m. žmogaus žodyno dydis paauga maždaug 6 200 lemų (Brysbaert et al., 2016). Šiuos rezultatus galima palyginti su Nyderlanduose Keuleers'o ir jo kolegų (2015) atliktu tyrimu, kur taip pat buvo analizuotos dvi amžiaus grupės. Šiame tyrime dalyvavo 278 590 dalyvių: 46,57 proc. buvo augę Belgijoje (šalyje viena iš oficialiųjų kalbų yra olandų), o 53,43 proc. – Nyderlanduose. Visų tiriamųjų gimtoji kalba buvo olandų. Dalyvių amžius svyravo nuo 12 iki 72 m.. Tyrimas atskleidė, kad olandų kalbos gimtakalbis, būdamas 13 m., moka vidutiniškai apie 29 000 žodžių, o apie 70 m. amžiaus – maždaug 43 000 žodžių. Vadinas, per 57 metus žmogaus žodynas padidėja apie 14 000 lemų (Keuleers et al., 2015). Iš aptartų tyrimų akivaizdu, kad žmogaus žodyno dydis su amžiumi didėja. Tačiau matyti, kad olandų kalbos gimtakalbių, kuriems atliekant tyrimą buvo 70 m., žodyno dydžio vidurkis yra šiek tiek mažesnis už anglakalbių gimtakalbių, kurių amžius yra apie 60 m. Galima atsargiai manyti, kad galbūt žodyno dydis ima mažėti nuo maždaug 60 m. Tačiau pabrėžtina, kad lyginami skirtingų kalbų gimtakalbiai, tad tvirtų išvadų daryti negalima. Vadinas, galima išvelgti tendenciją, kad žodynas kaupiamas su amžiumi, tačiau kol kas nėra atlikta konkrečių tyrimų, kurie įrodytų, jog nuo tam tikro amžiaus žmogaus žodynas ima mažėti (Andriuškevičiūtė ir Vilkienė, 2019).

Iš to, kas buvo pasakyta, matyti, kad skirtingų kalbų gimtakalbių žodynas gali būti skirtingo dydžio. Be to, pastebėtina, kad kiekvienos šalies gimtakalbis, kaip rodo įvairūs



tyrimai, savo žodyną per tam tikrą laiką praturtina nevienodu žodžių skaičiumi. Sunku pasakyti, kodėl taip yra, tačiau galima spėti, kad rezultatams, kaip jau buvo minėta, įtakos gali turėti ir tyrėjų naudojamos nevienodos metodikos, žodyno dydžio testai rengiami remiantis savo prigimtimi skirtingais šaltiniais. Taip pat įtakos gali turėti tiriamųjų imtis. Dauguma čia paminėtų tyrimų yra ne itin didelės imties, tad dėl to rezultatai gali būti netikslūs.

Gimtakalbių ir negimtakalbių lietuvių kalbos vartotojų žodyno dydis taip pat yra matuojamas. Jau anksčiau minėtoje Vilkienės et al. (2019) mokslo studijoje aprašomas tyrimas, kuriame buvo analizuotas Vilniaus gimnazijų trečių klasių mokinių žodyno dydis. Kadangi šiame darbe aktualus tik gimtakalbio žodynas, toliau pateikiami tik lietuvių gimtakalbių rezultatai, aptarti minėtoje studijoje. Iš Vilniaus gimnazijų lietuvių ugdomąja kalba tyrime dalyvavo 69 merginos ir 58 vaikinai, iš viso – 127 mokiniai. Dalyvių amžiaus vidurkis buvo 16,78 m. Tyrimo rezultatai atskleidė, kad vidutinis tokio amžiaus merginų žodyno dydis yra 54 300, o vaikinų – 50 500 žodžių. Taigi aptariamas tyrimas leido teigti, kad merginų žodynas yra statistiškai reikšmingai didesnis nei vaikinų.

Kitas žodyno dydžio tyrimas buvo atliktas Karolinos Andrieuškevičiūtės (2019). Jame buvo analizuotos dvi lietuvių kalbos gimtakalbių amžiaus grupės: pirmos grupės amžiaus vidurkis buvo 21,68 m., o antros – 49,31 m. Jaunesnių tiriamųjų grupę sudarė 31 moteris ir 37 vyrai, o vyresnių – 30 moterų ir 34 vyrai. Tad iš viso tyrime dalyvavo 132 respondentai. Šis tyrimas atskleidė, kad vidutinis jaunų moterų žodyno dydis yra apie 59 000 žodžių, o vyrų – 56 200 žodžių. Paminėtina, kad respondentės moterys buvo filologės, o vyrai – fizikai. Abi grupės gana specifinės. Tai, kad jaunesniosios grupės moterys buvo filologės, galėjo lemti aukštesnius jų žodyno dydžio testo rezultatus. Vyresniosios grupės moterų vidutinis žodyno dydis buvo 60 500 žodžių, o vyrų – 61 000 žodžių. Matyti, kad abiejų grupių vidurkiai skiriasi nedaug, statistiškai nereikšmingai.

Specifinę grupę – VU anglų filologijos studentus – analizavo ir Greta Pažereckaitė (2020). Ji ištyrė 30 pirmakursių ir 30 ketvirtakursių studentų. Šio tyrimo duomenų analizė atskleidė, kad pirmakursių žodyno dydžio vidurkis yra apie 57 000 lemų, o ketvirto kurso studentų – apie 60 300. Matyti, kad studentų žodyno dydis panašus į Andrieuškevičiūtės (2019) tyrime gautus studentų grupės rezultatus. Apskritai pastebėtina, kad minėtų tyrimų rezultatai daug aukštesni palyginti su, pavyzdžiui, Nyderlanduose atlikto tyrimo duomenimis. Galima spėti, kad tam įtakos galėtų turėti gana skirtingos tiriamųjų imtys: Nyderlanduose atliktas tyrimas yra didelio masto, o kiti aptariami tyrimai yra daug mažesnių imčių. Be to, Pažereckaitės tyrimo respondentai buvo filologai.

Apžvelgus įvairių kalbų gimtakalbių tyrimus galima matyti, kad skirtingų kalbų gimtakalbių žodyno dydis gana varijuoja. Tačiau, be abejo, konkrečios kalbos gimtakalbių žodynas taip pat yra nevienodas. Pavyzdžiui, Vilkienės et al. (2019) tyrimas atskleidė, kad gimtakalbių Vilniaus gimnazijų trečių klasių mokinių minimalus lietuvių kalbos žodynas apie 32 000 lemų, o didžiausias – apie 63 700 lemų, vidurkis – apie 53 800. Taigi vidurkis dar neatskleidžia, koks yra konkretaus gimtakalbio žodyno dydis.

## Žodynui įtaką darantys veiksniai

Matuodami žmogaus kurios nors kalbos žodyno dydį, tyrėjai sprendžia ir klausimus, kokie veiksniai gali daryti įtaką asmens žodyno dydžiui. Šiame straipsnyje analizuojama įvairių veiksmų įtaka tiriamų mokinių žodyno dydžiui. Tyrimui atlikti naudotame sociolingvistiniame klausimyne respondentų buvo prašoma nurodyti savo lytį, atsakyti į klausimus, kaip dažnai jie skaito įvairius tekstus (grožinę literatūrą, internetinius naujienų portalus, laikraščius ar žurnalus) ir kaip dažnai jie žiūri filmus, *YouTube* kanalus lietuviškai bei klausosi lietuviškų radijo stočių. Remiantis įvairiais tyrimais yra teigiama, kad šios veiklos gali daryti įtaką žodyno dydžiui, tad toliau šie tyrimai yra aptariami detaliau.

Pavyzdžiui, mokslininkai tiria lyties veiksmo įtaką žodyno dydžiui. Tyrėjos Robin Scarcella ir Cheryl Zimmerman (1998) analizavo studentų anglų kalbos, kaip svetimšios, akademinį žodyną, o tyrimo rezultatai parodė, kad vaikinų žodynas buvo didesnis nei merginų. Mercedes Díez Prados (2010), atlikusi tokį pat tyrimą su negimtakalbiais, neaptiko merginų ir vaikinų žodyno dydžio skirtumo. Tačiau tai nereiškia, kad šis veiksnys neturėtų būti tikrinamas kituose tyrimuose. Pavyzdžiui, Andriuškevičiūtės ir Vilkienės (2019) tyrime taip pat teigiama, kad lytis neturi įtakos žodyno dydžiui. Tačiau Vilkienės et al. (2019) buvo nustatyta, kad mokinių grupėje lyties veiksnys daro statistiškai reikšmingą įtaką žodyno dydžiui. Tad mokslininkai vis dar kelia klausimą, ar esama reikšmingų sąsajų tarp žodyno dydžio ir lyties, bet kol kas vienareikšmio atsakymo nepateikia. Šiame straipsnyje aprašomame tyrime taip pat buvo keltas klausimas, ar skiriasi mokinių žodyno dydis lyties atžvilgiu.

Dar viename mokslininkų Mabel L. Rice ir Lindos Woodsmall (1988, 422–423) tyrime teigiama, kad žmogus gali atsitiktinai išmokti naujų žodžių (gimtosios arba svetimšios kalbos) žiūrėdamas televiziją ar filmus. Tyrimas buvo atliktas su vaikais, kurie žiūrėjo 12 minučių animacinį filmą. Į filmą buvo įtraukta 20 gimtosios kalbos žodžių, kuriuos norėta, kad vaikai įsimintų. Visi tie žodžiai filme buvo paminėti mažiausiai penkis kartus. Buvo sudaryta ir kontrolinė grupė, kuri žiūrėjo animacinį filmą, į kurį nebuvo įtraukti tikrinami žodžiai. Po to vaikai turėjo atlikti nuotraukų atpažinimo užduotį: tiriamasis turėjo pasakyti, kas vaizduojama nuotraukoje (nuotraukose buvo filmo kadrai), o tam reikėjo pavartoti tikslius žodžius. Vaikai, kurie žiūrėjo filmą su tikslingai įtrauktais žodžiais, užduotį atliko geriau nei kontrolinė grupė. Tai leidžia daryti prielaidą, kad žmogus gali išmokti naujų žodžių žiūrėdamas filmus, televiziją. Šiame tyrime naudotame klausimyne taip pat buvo klausimas apie filmų žiūrėjimą lietuvių kalba.

Be to, mokslininkų buvo stebėta ir skaitymo įtaka žodyno dydžiui. Pavyzdžiui, Andriuškevičiūtės (2019) tyrime teigiama, kad tų respondentų, kurie savo malonumui skaito visada arba dažnai (grožinę literatūrą, publicistiką), žodyno dydis yra didesnis nei tiriamųjų, kurie skaito retai arba niekada to nedaro. Todėl šiame tyrime taip pat buvo įtraukti klausimai apie skaitymo polinkius ir stebima jų įtaka žodynui.



Mokslininkai savo darbuose mini ir kitokius veiksnius, darančius įtaką žmogaus žodyno dydžiui: tai ir namų aplinka (Huttenlocher et al., 1991; Hart, Risley, 1995), tėvų ir vaikų bendravimo polinkiai (Pancsofar ir Vernon-Feagans, 2006), tėvų ligos (Paulson et al., 2006; 2009), žmonių išsilavinimas (Keuleers et al., 2015; Andriuškevičiūtė ir Vilkienė, 2019; Pažereckaitė, 2020), mokamų svetimųjų kalbų skaičius (Keuleers et al., 2015; Andriuškevičiūtė ir Vilkienė, 2019), amžius (Keuleers et al., 2015; Segbers ir Schroeder, 2017; Andriuškevičiūtė ir Vilkienė, 2019) ir kt. Šiame tyrime tokie veiksniai nebuvo analizuojami, todėl plačiau neaptariami.

## Tyrimo metodika ir dalyviai

Atliekant šiame straipsnyje aprašomą tyrimą duomenims rinkti buvo pasitelktas lietuvių kalbos žodyno dydžio testas (plačiau žr. Vilkienė et al., 2019, 28–30), publikuotas Vilkienės, Vilkaitės-Lozdienės ir Bružaitės-Liseckienės mokslo studijoje *Lietuvių kalba Vilniaus lietuviškose, lenkiškose ir rusiškose gimnazijose: mokėjimo kokybė, kalbinės nuostatos ir motyvacija* (ibid., 78–79). Taip pat, atsižvelgus į minėtoje studijoje pateiktą sociolingvistinį klausimyną (ibid., 77–78), buvo parengti klausimai ir šiam tyrimui tikslingai renkant duomenis mažuose miestuose ir miesteliuose. Pabrėžtina, kad buvo atrinkti tik tie klausimai, kurie aktualūs gimtakalbiams. Taip elgtis paskatino siekis palyginti Vilniaus ir mažų miestų mokinių rezultatus, patikrinti, ar vienodi sociolingvistiniai veiksniai daro įtaką abiejų grupių žodyno dydžiui. Taip pat, kaip jau buvo minėta, siekta, sudėjus mažuose miestuose atlikto tyrimo rezultatus su Vilniaus mokinių rezultatais, patikslinti maždaug 17 m. amžiaus lietuvių kalbos gimtakalbio žodyno vidurkį.

Taigi iš viso analizuotos 242 vienuoliktų klasių mokinių atliktys. Šio tyrimo tikslais keturiuose mažuose miestuose buvo surinkta 115 mokinių atliktųjų: 62 merginų ir 53 vaikinių. Minėtos studijos autorės pasidalijo savo neanalizuotais Vilniaus miesto duomenimis, kad straipsnio autorės galėtų juos patikrinti tiksliniais aspektais. Taigi straipsnyje analizuoti ir 127 dviejų Vilniaus gimnazijų mokinių – 69 merginų ir 58 vaikinių – duomenys (plačiau apie šią grupę žr. Vilkienė et al., 2019, 13–20). Iš skaičių matyti, kad abi grupės, galima sakyti, buvo subalansuotos lyties aspektu. Be to, abiejų grupių – Vilniaus ir mažų miestų – mokinių amžius varijavo nuo 16 iki 18 m., o amžiaus vidurkis buvo apie 17 m. Tačiau reikia pridurti, kad tiek Vilniuje, tiek mažuose miestuose atliktų tyrimų atveju buvo pasirinkta ne reprezentatyvioji, o patogioji imtis, t. y. tyrime dalyvavo tų mokyklų, kurios sutiko dalyvauti tyrimuose, mokiniai.

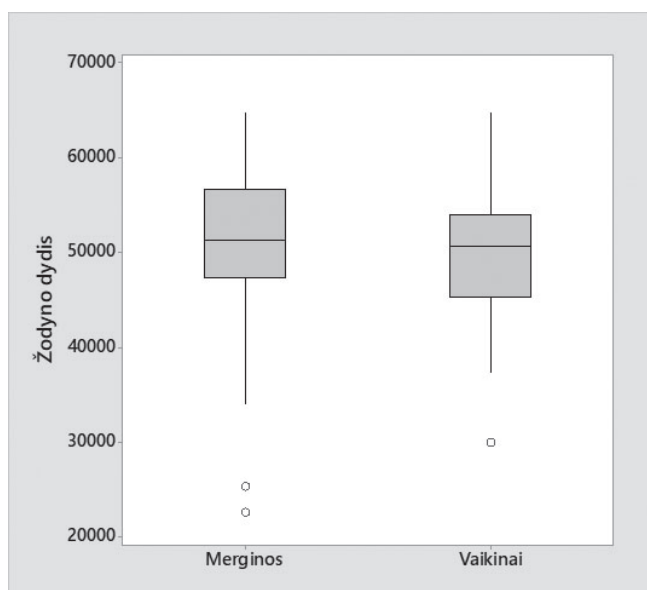
Tiek mažų miestų, tiek Vilniaus mokinių duomenys vėliau buvo suvesti į *MS Excel* lentelines skaičiuokles ir apdoroti statistine *Minitab 17* programa. Kad būtų atsakyta, ar skirtumų tarp grupių (gyvenamosios vietos ir lyties atžvilgiu) rezultatų vidurkiai skiriasi statistiškai reikšmingai, buvo atliktas *T* testas. O kad būtų patikrinta, kurie sociolingvistiniai veiksniai daro reikšmingą įtaką žodyno dydžiui, pasitelkta regresinė analizė.

## Tyrimo rezultatai

### *Lytis ir žodyno dydis*

Vilkienės et al. studijoje (2019, 33) teigiama, kad Vilniaus mokyklų vienuoliktokų rezultatai skiriasi nelygu lytis: merginų vidutinis žodynas buvo apie 54 300, o vaikinų – apie 50 800 lemų. Taigi skirtumas 3 500 lemų. Atlikus dvipusį *T* testą, skirtą dviejų grupių rezultatų vidurkių statistiniam reikšmingumui patikrinti, paaiškėjo, kad lietuviškų gimnazijų merginų ir vaikinų žodyno vidurkiai skiriasi statistiškai reikšmingai ( $p = 0,005^{***}$ ).

Šio straipsnio autorėms atlikus konceptualų tyrimo pakartojimą mažuose miestuose, buvo gauti kitokie rezultatai (žr. 1 pav.).

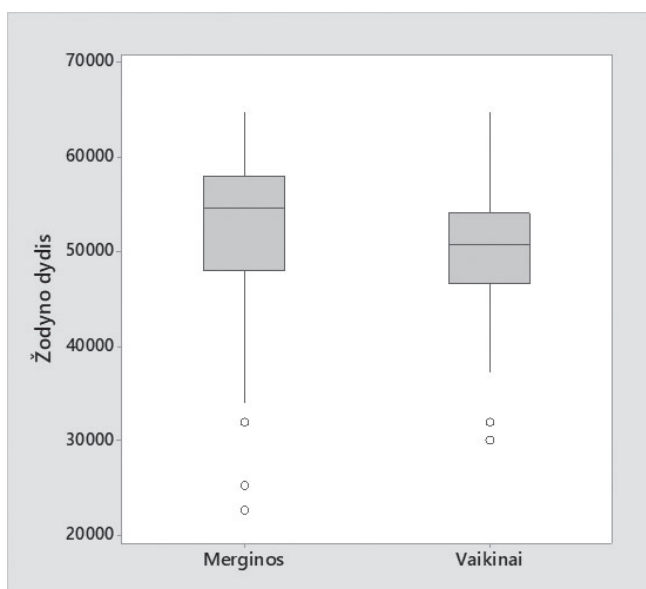


1 pav. Mažų miestų merginų ir vaikinų žodyno dydžio pasiskirstymas

Duomenų analizė atskleidė, kad mažų miestų merginų ir vaikinų žodyno dydis skiriasi nedaug. Mažiausias žodynas merginų atveju yra dviejų respondenčių – 22 678 ir 25 346, o vaikinų grupėje – vieno tiriamojo 30 015 lemų. Skirtumas tarp mažiausio merginos ir mažiausio vaikinų žodyno dydžio yra apie 7 340 lemų. Pastebėtina, kad ir merginų, ir vaikinų atskleistas toks pat didžiausias žodynas, t. y. 64 699 lemos, tačiau pabrėžtina, kad abiejose grupėse tokį rezultatą pasiekė tik po vieną respondentą. Medianos taip pat skiriasi nedaug: merginų grupės mediana yra 51 026 lemos, o vaikinų – 50 692. Vadinasi, daugumos tiriamųjų žodynas yra maždaug tokio paties dydžio. Taip pat iš 1 paveikslą akivaizdu, kad merginų grupės rezultatų trečias kvartilis yra didesnis nei vaikinų, vadinasi, didesnės merginų dalies rezultatai yra šiek tiek geresni nei vaikinų. Taip pat patikrinti tiriamųjų grupių vidurkiai: merginų žodyno dydžio vidurkis yra apie

50 500 lemu, o vaikinių – maždaug 49 400. Vadinas, skirtumas yra apie 1 100 lemu. Šis žodyno dydžio vidurkių skirtumo statistinis reikšmingumas buvo patikrintas *Studento T-kriterijumi*. Analizė atskleidė, kad rezultatai nėra statistiškai reikšmingi ( $p = 0,424$ ). Vadinas, negalima teigti, jog nedidelių miestų 11 klasių mokinių rezultatai pagal lytį yra statistiškai reikšmingai skirtingi. Kitaip tariant, šios tiriamųjų grupės rezultatai lyties požiūriu yra gana panašūs, skirtingai nei Vilniaus miesto analogiškos grupės atveju.

Tuomet šio straipsnio autorių buvo nuspręsta patikrinti, ar lyties veiksnys išlieka statistiškai reikšmingai veiksmingas, kai sujungiami Vilniaus ir mažų miestų rezultatai. Primintina, kad tokiu atveju merginų grupę sudarė 131 respondentė, o vaikinių – 111. Atlikus merginų ir vaikinių žodyno dydžio testo rezultatų analizę paaiškėjo, kad jų žodyno dydžio rezultatai skiriasi (žr. 2 pav.).



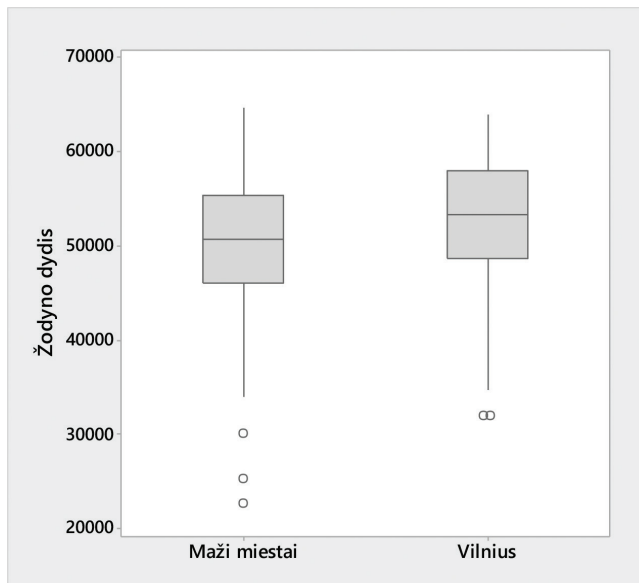
2 pav. Visų merginų ir vaikinių žodyno dydžio pasiskirstymas

Iš 2 paveikslo matyti, kad abiejose grupėse buvo pavienių mažiausių žodyno dydžio rezultatų. Merginų grupėje mažiausias žodyno dydis buvo 22 678 lemos, o vaikinių – 30 015. Skirtumas tarp mažiausio merginos ir mažiausio vaikinų žodyno dydžio yra apie 7 340 lemu. Pastebėtina, kad ir merginų, ir vaikinių atskleistas toks pat didžiausias žodynas, t. y. 64 699 lemos, tačiau pabrėžtina, kad abiejose grupėse tokį rezultatą pasiekė po vieną respondentą. Taip pat matyti, kad merginų trečias kvartilis yra gerokai aukščiau nei vaikinių, vadinas, merginos žodyno dydžio testą atliko geriau nei vaikinai. Be to, akivaizdu, kad skiriasi abiejų grupių mediana: merginų žodyno dydžio mediana yra 54 694 lemos, o vaikinių – 50 692. Pastebėtina, kad vaikinių ji gerokai žemesnė nei merginų. Abiejų grupių medianų skirtumas yra 4 002 lemos. Lyginamų grupių žodyno dydžio vidurkiai taip pat skiriasi: merginų – 52 540, o vaikinių – 50 230. Matyti, kad

merginų ir mediana, ir vidurkis yra aukštesni už vaikinų. Šis žodyno dydžių vidurkių skirtumas buvo patikrintas *Stjudento T-kriterijumi*. Atlikus *T* testą paaiškėjo, kad merginų žodyno dydžio vidurkis yra statistiškai reikšmingai didesnis už vaikinų ( $p = 0,011$ ). Iš to, kas buvo šiame skyriuje pasakyta, galima daryti išvadą, kad Vilniaus merginų rezultatai lemia visos merginų grupės geresnius žodyno dydžio rezultatus.

### *Gyvenamoji vieta ir žodyno dydis*

Kaip buvo minėta, atliekant tyrimą buvo siekiama patikrinti, ar gyvenamoji vieta – mažas miestas ir sostinė Vilnius – gali turėti įtakos mokinio žodyno dydžiui. Siekiant atsakyti į šį klausimą, buvo sugretinti nedidelių miestų ir sostinės tirtų vienuoliktų klasių mokinių žodyno dydžio testo rezultatai, kurie apibendrinti 3 paveiksle.



3 pav. Mažų miestų ir Vilniaus žodyno dydžio testų rezultatų pasiskirstymas

Iš duomenų analizės matyti, kad skiriasi lyginamų grupių rezultatų mediana: mažų miestų žodyno dydžio mediana yra 50 692 lemos, o Vilniaus didesnė – 53 360, taigi mažų miestų ji yra žemesnė, o skirtumas – apie 2 670 lemų. Taip pat matyti, kad Vilniaus mokinių rezultatų trečias kvartilis yra aukščiau už mažų miestų trečią kvartilį, vadinasi, Vilniaus mokiniai žodyno dydžio testą atliko geriau nei tos pačios klasės mokiniai, kurie gyvena ne sostinėje. Be to, skiriasi ir žodyno dydžio vidurkiai: mažų miestų – apie 50 000, o Vilniaus – maždaug 52 800 lemų. Matyti, kad skirtumas yra apie 2 800 lemų. Siekiant nustatyti šio skirtumo statistinį reikšmingumą, buvo atliktas *T* testas ir paaiškėjo, kad šių grupių rezultatų vidurkiai skiriasi statistiškai reikšmingai ( $p = 0,002$ ). Vadinasi, galima teigti, kad sostinėje gyvenančių mokinių žodynas yra didesnis nei mažų miestų mokinių.

Taip pat matyti, kad abiejose grupėse yra pavienių rezultatų. Mažuose miestuose išsiskyrė trys pavieniai atvejai, o Vilniuje tokių buvo du. Mažiausias žodynas mažuose miestuose buvo 22 678 lemos – šis rezultatas buvo merginos, o Vilniuje – 32 016 lemy, šie atvejai buvo du: merginos ir vaikino. Iš skaičių matyti, kad mažiausias Vilniuje minimalus žodynas yra didesnis už minimalų mažame mieste 9 300 lemy. Didžiausias žodynas Vilniuje buvo vaikino, sukaupusio 64 032 lemas, o mažuose miestuose vos didesnis – 64 699. Tokį žodyną atskleidė ir mergina, ir vaikinas. Matyti, kad didžiausias rezultatas iš esmės nesiskiria. Akivaizdu, kad mažų miestų mokinių rezultatų vidurkis mažesnis dėl to, kad yra pavienių itin mažų žodynų. Be to, galima teigti, kad būtent Vilniaus merginų rezultatai lemia, kad sostinės mokinių žodyno dydis yra didesnis nei mažų miestų mokinių.

### *Sociolingvistinių veiksnių įtaka žodyno dydžiui*

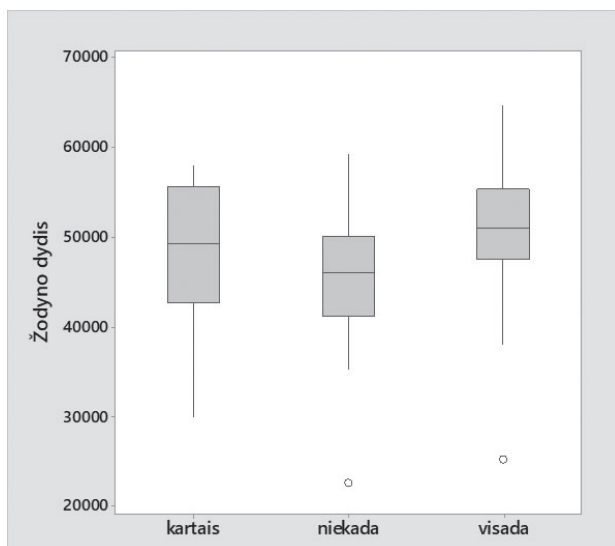
Kad būtų atskleista, kurie sociolingvistiniai veiksniai daro reikšmingą įtaką mokinių žodynui, buvo atlikta regresinė analizė. Pirmiausia šiuo aspektu buvo tiriama mažų miestų mokinių grupė. Priklausomasis kintamasis buvo žodyno dydžio testo rezultatas. Į pradinį modelį buvo įtraukti tokie nepriklausomi kintamieji: lytis, skaitymo įpročiai laisvalaikiu lietuvių kalba (grožinės literatūros, naujienų portalų, laikraščių ar žurnalų), filmų ir *YouTube* kanalų žiūrėjimas lietuviškai ir lietuviškų radijo stočių klausymas. Analizė buvo pradėta nuo šio modelio, kiekviename žingsnyje po vieną buvo atmetami tie veiksniai, kurie nebuvo reikšmingi, kol galutiniame modelyje liko tik reikšmingi kintamieji. Atlikus regresinę analizę paaiškėjo, kad pradiniame modelyje nebuvo reikšmingas grožinės literatūros, naujienų portalų, laikraščių, žurnalų skaitymas savo malonumui. Taip pat iš modelio buvo pašalinti lietuviškų radijo stočių klausymo ir *YouTube* kanalų žiūrėjimo kintamieji, nes jie nebuvo reikšmingi. Galutinis modelis pavaizduotas 2 lentelėje.

2 lentelė

*Žodyno dydžio rezultatų modelis: mažų miestų atvejais*

<b>Fiksuoti modelio parametrai</b>	<b>Parametrų įverčiai (standartinė paklaida), parametro reikšmingumas</b>
$b_0$ (laisvasis narys)	45070 (1922), $p = 0,000$
$b_1$ (filmų žiūrėjimas – kartais)	3032 (2595), $p = 0,245$
$b_2$ (filmų žiūrėjimas – visada)	6122 (2076), $p = 0,004$

Taigi galutiniame modelyje liko tik filmų žiūrėjimas lietuvių kalba. Matyti, kad tik filmų žiūrėjimas lietuvių kalba visada arba dažnai yra statistiškai reikšmingas veiksnys, darantis įtaką tiriamosios grupės žodyno dydžiui (žr. 4 pav.).



4 pav. Filmų žiūrėjimas lietuviškai ir žodyno dydžio rezultatų pasiskirstymas. Mažų miestų atvejais

Kaip matyti iš 4 paveikslo, tų respondentų, kurie teigė filmus lietuvių kalba žiūrintys visada arba dažnai, rezultatų mediana ir viršutinė riba yra gerokai aukščiau nei tų, kurie tai daro retai arba niekada neužsiima šia veikla. Pastebėtina, kad mokinių, kurie visada arba dažnai žiūri filmus lietuviškai, žodyno dydžio vidurkis yra apie 51 100 lemu, o tų, kurie tai daro retai arba niekada to nedaro, – 45 400. Primintina, kad ir ankstesniuose tyrimuose, pavyzdžiui, Rice ir Woodsmall (1988), buvo taip pat pastebėta, kad televizijos ir filmų žiūrėjimas turi įtakos žodyno dydžiui. Galima teigti, kad mažų miestų atveju ši įžvalga pasitvirtina.

Šio straipsnio autorių taip pat buvo siekiama patikrinti, ar Vilniuje gyvenančių mokinių žodyno dydžiui gali daryti įtaką toks pat veiksnys kaip ir mažų miestų mokiniams, todėl taip pat buvo atlikta šios grupės duomenų regresinė analizė. Priklausomasis kintamasis ir šiuo atveju, žinoma, buvo žodyno dydžio testo rezultatas. Į pradinį modelį buvo įtraukti tokie pat nepriklausomi kintamieji, kaip ir mažų miestų analizės atveju. Kiekvienu analizės žingsniu buvo atmetami tie sociolingvistiniai veiksniai, kurie nebuvo statistiškai reikšmingi. Galutinis modelis pavaizduotas 3 lentelėje.

### 3 lentelė

Žodyno dydžio rezultatų modelis: Vilniaus atvejais

Fiksuoti modelio parametrai	Parametrų įverčiai (standartinė paklaida), parametro reikšmingumas
$b_0$ (laisvasis narys)	54375 (790), $p = 0,000$
$b_1$ (lytis – vaikas)	-3384 (1169), $p = 0,004$



Iš duomenų analizės rezultatų matyti, kad Vilniaus atveju reikšmingas yra tik lyties veiksnys – vaikinų žodyno dydis yra statistiškai reikšmingai mažesnis už merginų ( $p = 0,004$ ). Vilniuje filmų žiūrėjimas, priešingai nei mažų miestų atveju, nebuvo reikšmingas veiksnys.

Siekiant nustatyti, kurie iš sociolingvistinių veiksnių galėtų daryti reikšmingą įtaką visos grupės – tiek mažų miestų, tiek Vilniaus mokinių kartų sudėjus – žodyno dydžiui, taip pat atlikta regresinė duomenų analizė. Priklausomasis kintamasis, suprantama, buvo žodyno dydžio testo rezultatas. Į pradinį modelį buvo įtraukti visi jau minėti sociolingvistiniai veiksniai. Galutinis modelis pateiktas 4 lentelėje.

#### 4 lentelė

*Žodyno dydžio rezultatų modelis: mažų miestų ir Vilniaus bendri rezultatai*

<b>Fiksuoti modelio parametrai</b>	<b>Parametrų įverčiai (standartinė paklaida), parametro reikšmingumas</b>
$b_0$ (laisvasis narys)	52540 (623), $p = 0,000$
$b_1$ (lytis – vaikas)	-2311 (921), $p = 0,013$

Iš duomenų analizės akivaizdu, kad šiuo atveju tik lyties veiksnys regresinėje analizėje yra reikšmingas. Matyti, kad vaikinų žodyno vidurkis yra statistiškai reikšmingai mažesnis už merginų ( $p = 0,013$ ). Iš gautos analizės pastebėtina, kad vaikinų žodynas yra mažesnis maždaug 2 300 lemų. Vadinas, šiame tyrime pasitvirtina ankstesnių mokslininkų (pavyzdžiui, Vilkienė et al., 2019) prielaida, kad lytis gali turėti įtakos žodyno dydžiui. Tačiau primintina, kad mažų miestų atveju tai nebuvo išreikšta.

## Išvados

Išanalizavus 242 lietuvių gimtakalbių 11 klasių mokinių atliktis galima teigti, kad maždaug 17 metų amžiaus lietuvių kalbos gimtakalbio žodyno dydis varijuoja nuo 22 678 iki 64 699 lemų, o vidurkis – apie 51 400 lemų. Primintina, kad Vilkienės et al. (2019) tyrime buvo nustatyta, kad Vilniaus mokyklų tokio paties amžiaus mokinių žodyno vidurkis yra 52 800 lemų. Taigi sujungus mažų miestų bei Vilniaus duomenis paaiškėjo, kad vidutinis 17 metų amžiaus gimtakalbio žodynas yra 1 400 lemų mažesnis, nei rodė Vilniaus duomenys. Tokį sumažėjimą nulėmė mažų miestų rezultatai. Tačiau pabrėžtina, kad šie duomenys gauti tiek Vilniuje, tiek mažuose miestuose išanalizavus ne reprezentatyviąją, o patogiąją imtį, taigi negali būti apibendrinti visai populiacijai.

Tyrime dalyvavusių mažų miestų mokinių, kurių amžius apie 17 metų, žodyno vidurkis yra apie 50 000 lemų, t. y. apie 2 800 lemų mažesnis, nei buvo nustatytas Vilniaus mokinių atveju (52 800 lemų). Atlikus  $T$  testą paaiškėjo, kad Vilniaus mokinių žodyno

dydžio vidurkis yra statistiškai reikšmingai didesnis nei tirtų mažų miestų mokinių ( $p = 0,002$ ). Vadinasi, galima daryti išvadą, kad gyvenamoji vieta gali daryti įtaką mokinio žodyno dydžiui: kaip rodo aprašomo tyrimo duomenys, mokinių, gyvenančių sostinėje, žodynas yra reikšmingai didesnis. Vykdam tarptautinį tyrimą PISA (2018) buvo pastebėta, kad gyvenamoji vieta turi įtakos penkiolikmečių skaitymo gebėjimams. Pasirėmus šio straipsnio autorių duomenimis, galima pratęsti mintį, kad mažų miestų mokinių skaitymo gebėjimai yra menkesni galbūt ir dėl to, kad mažesnis yra šių mokinių žodynas. Tačiau šią išvargą reikėtų tikrinti toliau kaupiant duomenis, analizuojant reprezentatyvią imtį.

Į vieną grupę sujungus Vilniaus ir mažų miestų duomenis, paaiškėjo, kad merginų grupės žodyno dydžio vidurkis – apie 52 540 (tik Vilniaus merginų atveju 54 300), o vaikinių – apie 50 230 lemų (tik Vilniaus vaikinių atveju 50 692). Pastebėtina, kad sujungus tiriamųjų duomenis tik merginų grupės žodyno dydžio vidurkis akivaizdžiai sumažėjo – apie 1 850 lemų. Taip nutiko dėl mažesnių mažų miestų respondenčių rezultatų. Sujungus Vilniaus ir mažų miestų duomenis ir atlikus  $T$  testą nustatyta, kad merginų žodyno dydžio vidurkis yra statistiškai reikšmingai didesnis už vaikinių ( $p = 0,011$ ). Tai leidžia daryti atsargią prielaidą, kad PISA (2018) tyrimuose pastebėta tendencija, jog merginų skaitymo gebėjimai yra geresni nei vaikinių, gali būti susijusi su tuo, kad merginų žodynas yra didesnis. Bet paminėtina, kad mažų miestų ir miestelių atveju  $T$  testas neparodė statistiškai reikšmingo vidurkių skirtumo tarp merginų ir vaikinių žodyno dydžio. Taigi skaitymo gebėjimų tyrimo (PISA, 2018) geresnius merginų rezultatus, spėtina, lemia didelių miestų, pavyzdžiui, Vilniaus merginų kompetencija. Ši išvarga toliau turi būti tikrinama plečiant tiriamųjų imtį.

Siekiant padaryti nuodugnesnes išvadas, kurie sociolingvistiniai veiksniai turi reikšmingos įtakos žodyno dydžiui, buvo atlikta regresinė duomenų analizė, po kurios paaiškėjo, kad šiame tyrime tiek visos tiriamųjų grupės atžvilgiu, tiek Vilniaus miesto mokinių požiūriu žodyno dydžiui reikšmingą įtaką turi lyties veiksnys: merginų žodyno dydis yra statistiškai reikšmingai didesnis ( $p = 0,013$  visos grupės atžvilgiu;  $p = 0,004$  Vilniaus atveju). O kiti sociolingvistiniai veiksniai akivaizdžios įtakos nedarė. Atskirai mažų miestų atveju reikšmingas pasirodė esąs filmų žiūrėjimo lietuvių kalba veiksnys ( $p = 0,004$ ), o lyties veiksnys įtakos nedarė. Tačiau tokius rezultatus reikėtų tikrinti patitelkus didesnę respondentų imtį. Vis dėlto visos grupės atžvilgiu šio tyrimo rezultatai patvirtina Vilkenės et al. (2019) pateiktą išvargą, kad lytis gali būti reikšmingas veiksnys žodyno dydžiui.

Baigiant pabrėžtina, kad šiame straipsnyje aprašomo tyrimo respondentų imtis buvo patogioji, o ne reprezentatyvioji, tad remiantis tokios imties rezultatais negalima daryti išvadų apie visą populiaciją. Tai yra šio tyrimo ribotumas. Vis dėlto ir šio tyrimo rezultatai papildė ir pakoreguoja jau turimus duomenis. Ateityje gvildenant mokinių žodyno dydžio klausimą būtų galima tikslingai analizuoti ir tai, ar žodyno dydis kinta mažėjant miestui.

## Literatūra

- Andriuškevičiūtė, K. (2019). Gimtakalbio lietuvių žodyno dydis. *Magistro darbas*. Vilnius: Vilniaus universitetas.
- Andriuškevičiūtė K., & Vilkienė, L. (2019). Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12, 154–181.
- Brysbart, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, 7, 1–11.
- Enström, I. (2004). Ordförråd och ordinlärning–med särskilt fokus på avancerade inlärare [*Žodynas ir žodžių mokymasis – ypatingas dėmesys pažengusiems vartotojams*]. *Svenska som andraspråk-i forskning, undervisning och samhälle*, 171–195.
- Enström, I. (2010). *Ordens värld: svenska ord-struktur och inlärning* [Žodžių pasaulis: švedų žodžių struktūra ir mokymasis]. Stockholm: Hallgren & Fallgren.
- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31(2), 105–119.
- Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied linguistics*, 11(4), 341–363.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Heinle: Paul H Brookes Publishing.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: A relation to language input and gender. *Developmental Psychology*, 27 (2), 236–248.
- Jakaitienė, E. (2009). *Leksikologija: studijų knyga*. Vilnius: Vilniaus universiteto leidykla.
- Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: challenging conventions. *TESOL Quarterly*, 50(4), 976–987.
- Keuleers, E., Stevens, M., Mandera, P., & Brysbart, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, 68(8), 1665–1692.
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365–391.
- Lindberg, I. (2006). Bedömning av skolrelaterat ordförråd [Su mokykla susijusio žodyno įvertinimas]. In Olofsson (Eds), *Symposium. Bedömning, flerspråkighet och lärande*. [Vertinimas, daugiakalbystė, mokymasis] (pp. 83–107). Stokholmas: HLS leidykla.
- Lindberg, I., Johansson Kokkinakis, S., Järborg, J., & Holmegaard, M. (2007). OrdiL–en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år [OrdiL – tekstyno pagrindu parengtas vidurinės mokyklos vadovėlių žodyno tyrimas]. Geteborgo universitetas: Švedų, kaip antrosios kalbos, institutas.
- Maskor, Z. M., & Baharudin, H. (2016). Receptive vocabulary knowledge or productive vocabulary knowledge in writing skill, which one important? *International Journal of Academic Research in Business and Social Sciences*, 6 (11), 2222–6990.

- Melka, F. (1997). Receptive vs. productive aspects of vocabulary. *Vocabulary: Description, Acquisition and Pedagogy*, 33(2), 84–102.
- Murphy, M. L. (2010). *Lexical meaning*. Cambridge: Cambridge University Press.
- Nation, I. S. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- OECD PISA. (2018). EBPO PISA 2018: rezultatai ir tendencijos. *Ataskaita*. Vilnius: Nacionalinis egzaminų centras.
- Pancsofar, N., Vernon, & Feagans, L. (2006). Mother and father language input to young children: Contributions to later language development. *Journal of Applied Developmental Psychology*, 27, 571–587.
- Paulson, J. F., Dauber, S. E., & Leiferman, J. A. (2006). Individual and combined effects of maternal and paternal depression on parenting behavior. *Pediatrics*, 118, 659–668.
- Paulson, J. F., Keefe, H. A., & Leiferman, J. A. (2009). Early parental depression and child language development. *Journal of Child Psychology and Psychiatry*, 50(3), 254–262.
- Pažerekaitė, G. (2020). Anglų filologijos studentų gimtosios lietuvių kalbos žodyno dydis. *Baigiamasis gretutinių studijų darbas*. Vilnius: Vilniaus universitetas.
- Peter, R., Abbuhl, R., Adams, R., Bowles, M., Connor-Linton, J., Doyle, P. G., Grenfell, M., & Trites, L. (2008). Replication studies in language learning and teaching: Questions and answers. *Language Teaching*, 41(1), 1–14. <https://doi.org/10.1017/S0261444807004727>
- Prados, M. D. (2010). Gender and L1 influence on EFL learners' lexicon. In R. M. J. Catalan (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 44–73). London: Palgrave Macmillan.
- Rice, M. L., & Woodsmall, L. (1988). Lessons from television: Children's word learning when viewing. *Child Development*, 59(2), 420–429.
- Scarcella, R., & Zimmerman, C. (1998). Academic words and gender: ESL student performance on a test of academic lexicon. *Studies in Second Language Acquisition*, 20(1), 27–49.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. UK: Palgrave Macmillan, Springer. <https://doi.org/10.1057/9780230293977>
- Segbers, J., & Schroeder, S. (2017). How many words do children know? A corpus-based estimation of children's total vocabulary size. *Language Testing*, 34 (3), 297–320.
- TreffersDaller, J., & Milton, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4 (1), 151–172.
- Vilkienė, L., Vilkaitė-Lozdienė, L., & Bružaitė-Liseckienė, J. (2019). *Lietuvių kalba Vilniaus lietuviškose, lenkiškose ir rusiškose gimnazijose: mokėjimo kokybė, kalbinės nuostatos ir motyvacija*. Vilnius: Vilniaus universiteto leidykla.
- Vilkienė, L., Vilkaitė-Lozdienė, L., Juknevičienė, R., Bružaitė-Liseckienė, J., Geben, K., & Ryvitytė, B. (2020). Jei tai nepakartojama, ar tai patikimas mokslas? Atviras laiškas apie tyrimų kartojimo svarbą kiekybiniuose tyrimuose. *Taikomoji kalbotyra*, 14, 33–38.
- Webb, S. (2013). Depth of vocabulary knowledge. *The Encyclopedia of Applied Linguistics*. Chichester: John Wiley and Sons Ltd.

---

# The Size of the Native Lithuanian Vocabulary of the 11<sup>th</sup> Grade Students

Rugilė Miškinytė<sup>1</sup>, Loreta Vilkienė<sup>2</sup>

<sup>1</sup> Vilnius University, Universiteto g. 5, LT-01131 Vilnius, rugile.miskinytee@gmail.com

<sup>2</sup> Vilnius University, Institute of Applied Linguistics, Department of Lithuanian Studies, Universiteto g. 5, LT-01131 Vilnius, Loreta.Vilkiene@flf.vu.lt

---

## Summary

The aim of this research was to measure the vocabulary size of native speakers of Lithuanian language attending the 11<sup>th</sup> grade in small cities and to compare the results with the research conducted in Vilnius schools. This paper aimed to find out whether the place of residence and the gender factor have a statistically significant influence on the size of the vocabulary and discover if any sociolinguistic factors can impact it. The research material consisted of 242 responses given to a sociolinguistic survey and vocabulary size test. T-test and regression analysis were applied in the research.

The results of the study indicated that the average vocabulary of an 11<sup>th</sup> grade student consists of approximately 51,400 lemmas. Vilnius school students have an average vocabulary size of about 52,800 lemmas; the vocabulary size of small towns students is lower – 50,000 lemmas. The T-test statistical analysis has revealed that the difference in the mean size of the vocabulary across genders is statistically significant ( $p = 0,011$ ). It was found out that the vocabulary size of students living in Vilnius is statistically significantly bigger than that of students in the smaller cities ( $p = 0,002$ ). The regression analysis indicated that gender can affect the vocabulary size. The analysis showed statistical significance ( $p = 0,004$ ) that watching movies in the Lithuanian language can impact the vocabulary size of students living in small towns.

---

**Keywords:** *vocabulary size, native speaker, Lithuanian language, 11<sup>th</sup> grade student, gender factor.*

---

Gauta 2020 05 28 / Received 28 05 2020  
Priimta 2020 07 08 / Accepted 08 07 2020