

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, www.taikomojikalbotyra.lt.

## **Suaugusio lietuvių gimtosios kalbos žodyno dydis**

### **Karolina Andriuškevičiūtė**

Vilniaus universitetas

karolina.andriuskeviciute@alumni.vu.lt

### **Loreta Vilkienė**

Vilniaus universitetas

loreta.vilkiene@flf.vu.lt

### **Anotacija**

Straipsnio tikslas – atsakyti į klausimus, koks yra suaugusio lietuvių gimtosios kalbos žodyno dydis (lemomis) ir kokie sociolingvistiniai veiksniai gali daryti jam įtakos. Tirtos 132 respondentų žodyno dydžio testo atliktys ir atsakymai į sociolingvistinį klausimyną. Duomenys apdoroti statistine *R-Studio* programa, atliktas *T* testas ir regresinė analizė. Nustatyta, kad jaunesnio amžiaus suaugusių gimtakalbių (apie 22 m. amžiaus) lietuvių kalbos žodyno dydžio vidurkis apytiksliai 57 490 lemų. Vyresnio amžiaus gimtakalbių (apie 49 m. amžiaus) lietuvių kalbos žodyno dydžio vidurkis maždaug 60 770 lemų. Regresinė duomenų analizė atskleidė, kad įtaką žodyno dydžiui daro šie veiksniai: magistrantūros pakopos išsilavinimas, mokamų svetimųjų kalbų skaičius ir grožinės literatūros skaitymas.

**Raktažodžiai:** *žodyno dydis, gimtoji kalba, sociolingvistinių veiksnių įtaka*

### **1. Įvadas**

Žodynas yra sudėtingas konstruktas, vienas iš kalbos komponentų, kiekvienam žmogui reikalingas norint sėkmingai komunikuoti, perduoti informaciją ar išreikšti savo jausmus. Taigi galima spėti, kad kuo daugiau žodžių žmogus yra sukaupęs, tuo tiksliau, geriau išreiškia savo mintis. Tai pagrindžia ir įvairūs tyrimai: vieni mokslininkai pabrėžia, kad žodyno dydis gali net atspindėti bendrą kalbos mokėjimo lygį ar lemti mokymosi sėkmę (Milton, Treffers-Daller 2013), o kiti teigia, kad sukauptas žodžių skaičius daro tiesioginę įtaką teksto suvokimo

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, www.taikomojikalbotyra.lt.

gebėjimų kokybei (Nation 2006). Vadinasi, žmogaus žodynas gali būti prielaida geriau ar sunkiau suprasti ir produkuoti sakinį ir rašytinį diskursus, apskritai – sėkmingai pasiekti savo komunikacinių tikslų. Visa tai rodo, kad žodyno dydžio tyrimai yra aktualūs.

Mokslininkai įvairiais aspektais aktyviai tyrinėja tiek žmogaus gimtosios, tiek negimtosios kalbos žodyno dydį. Pavyzdžiui, nemažai dėmesio sulaukia amžiaus veiksnio įtakos žodyno dydžiui tyrimai (Keuleers et al. 2015, Brysbaert et al. 2016, Segbers, Schroeder 2017), taip pat analizuojama, kaip žodyno dydį veikia išsilavinimas, lytis (Coxhead et al. 2015, Keuleers et al. 2015), skaitymo veiklos aktyvumas (Milton, Treffers-Daller 2013) ir pan. Pasakytina, kad Lietuvoje žodyno dydžio tyrimų dar nėra daug: 2017–2019 m. Vilniaus universiteto Taikomosios kalbotyros instituto darbuotojų Loretos Vilkienės, Lauros Vilkaitės-Lozdienės ir Justinos Bružaitės-Liseckienės buvo vykdytas mokslinis projektas *Lietuvių kalba Vilniaus lietuviškose, lenkiškose ir rusiškose gimnazijose: mokėjimo kokybė, kalbinės nuostatos ir motyvacija* (rėmė VLKK), kurio vienas iš uždavinių buvo nustatyti gimtakalbių lietuvių ir negimtakalbių lenkiškų bei rusiškų Vilniaus gimnazijų trečių klasių mokinių lietuvių kalbos žodyno dydį ir patikrinti, kokie sociolingvistiniai veiksniai daro jam įtaką (plačiau žr. Vilkienė et al. 2019). Taigi per šį projektą pirmą kartą buvo išmatuotas lietuvių kalbos vartotojo žodyno dydis. Rezultatai parodė, kad 16–17 metų gimtakalbio Vilniaus gimnazisto žodyno dydis maždaug 52 790 žodžių (lemų).

Šio straipsnio autorės tęsia pradėtus gimtakalbio lietuvių kalbos vartotojo žodyno dydžio tyrimus ir kelia **tikslą** atsakyti į klausimus, koks yra suaugusio lietuvių gimtosios kalbos žodyno dydis ir kokie sociolingvistiniai veiksniai daro jam įtaką. Kad tikslas būtų pasiektas, analizuojamos 132 tiriamųjų žodyno dydžio testo atliktys ir atsakymai į sociolingvistinį klausimą. Atliekant tyrimą buvo naudotas Vilkienės, Vilkaitės-Lozdienės ir Bružaitės-Liseckienės sudarytas žodyno dydžio testas (žr. Vilkienė et al. 2019), kurio vienetas – lema. Sociolingvistinės apklausos klausimynas buvo parengtas šio straipsnio autorių. Duomenys apdoroti statistine *R-Studio* programa, statistinis skirtingų tiriamųjų grupių duomenų reikšmingumas patikrintas atlikus *T* testą. Ieškant atsakymo į klausimą, kurie sociolingvistiniai veiksniai darė reikšmingą įtaką žodyno dydžio rezultatams, atlikta regresinė analizė. Už pagalbą atliekant šią analizę straipsnio autorės dėkingos kolegei Laurai Vilkaitei-Lozdieni.

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

Jau planuojant tyrimą buvo numatyta apklausti dvi respondentų grupes – jaunesnio ir vyresnio amžiaus gimtakalbius. Jaunesniosios grupės (toliau – JG) duomenys buvo renkami iš studentų Vilniaus universiteto Filologijos ir Fizikos fakultetuose. Iš viso šią grupę sudarė 68 studentai (31 mergina ir 37 vaikinai), kurių amžiaus vidurkis – 22 metai. Vyresniosios tiriamųjų grupės (toliau – VG) duomenys buvo renkami naudojant „sniego gniūžtės“ metodą, taigi ši grupė yra heterogeniškesnė. Iš viso ją sudaro 64 tiriamieji (30 moterų ir 34 vyrai), kurių amžiaus vidurkis 49 metai. Skiriasi ne tik šių asmenų išsilavinimo ir darbo sritys, tačiau ir išsilavinimo laipsnis (profesinis, aukštasis bakalauro, magistro laipsnis ir t. t.). Kiekviena šių grupių – tiek JG, tiek VG – dar buvo suskaidyta lyties požiūriu į pogrupius: jaunesnioji moterų grupė (toliau – JMG), jaunesnioji vyrų grupė (toliau – JVG), vyresnioji moterų grupė (toliau – VMG) ir vyresnioji vyrų grupė (toliau – VVG). Taigi tiriamieji sugrupuoti pagal lytį, nes norėta išsiaiškinti, ar šis veiksnys turi įtakos žodyno dydžiui.

## 2. Žodyno dydis ir jo tyrimai

Šiame straipsnyje aprašomas tyrimas, kuriuo siekta nustatyti, koks gimtakalbių lietuvių kalbos žodyno dydis, kartais dar vadinamas leksikonu. Mokslininkai sutaria, kad žodyno dydis – tai žmogaus mokamų (tiek pasyviai, tiek aktyviai vartojamų) tam tikros kalbos žodžių skaičius (Read 2000: 31). Manoma, kad žodyno dydis (angl. *size*) yra glaudžiai susijęs ir su žodyno gyliu (angl. *depth*) (Schoonen, Verhallen 2008: 215), t. y. su žodžių mokėjimo kokybe, apimančia labai platų žinių apie žodį spektrą. John’as Read’as, remdamasis Paul’u Nation’u, teigia, kad žodžio mokėjimas – tai ir jo reikšmės žinojimas, gebėjimas tą žodį parašyti ir pasakyti, gramatiškai taisyklingai jį pavartoti įvairiuose kontekstuose ir pan. (Read 2000: 26). Batia Laufer ir Zahava Goldstein pažymi, kad žodyno dydžio matavimas yra gana paviršutiniškas, kadangi testuojami žodžiai yra pateikiami vieni, be konteksto (Laufer, Goldstein 2004: 400–401). Taigi būtų galima teigti, kad naudingiau yra matuoti žodyno gylį, tačiau tai užima daug laiko ir gali sukelti sunkumų siekiant gauti objektyvius duomenis ir išvadas, kadangi kol kas dar nėra sukurta tinkamų priemonių tam atlikti (Vilkienė ir kt. 2019: 23)<sup>1</sup>. O žodyno dydis matuojamas paprasčiau, tokių tyrimų įvairiose šalyse esama ne vieno.

<sup>1</sup> Šiame darbe remiamasi Vilkienės ir kt. (2019) dar nepublikuota mokslinė studija, tačiau šio darbo autorės gavo leidimą ja naudotis.

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

Tyrėja Averil Coxhead su kolegomis teigia, kad pirmieji mokslininkų rezultatai, gauti matuojant žmonių anglų kalbos žodyno dydį, buvo netikslūs, nes į tyrimus be jokios atrankos buvo įtraukti visi žodžiai (ir vietovardžiai, ir asmenų vardai, pavardės, ir santrumpos, ir pan.), be to, tai buvo daroma net nesvarstant, kaip iš viso žodis galėtų būti apibrėžiamas (Coxhead et al. 2015). Taip pat neretai atliekant žodyno dydžio tyrimus, buvo skaičiuojama ne tai, kiek žmogus moka žodžių, o kiek jis moka žodžių lizdų, kuriuos sudaro pamatinis žodis ir jo dariniai kartu su visomis kaitybinėmis formomis (Read 2000: 18; Vilkienė et al. 2019: 25). Pasak Norbert'o Schmitt'o, žmogus, žinantis pamatinę žodžio formą, gali palyginti nesunkiai nuspėti kitos to žodžio formos reikšmę (pavyzdžiui, žinodamas žodį *siūti*, gali suprasti formas *susiuvo*, *susiuva* ir pan.) (Schmitt 2010: 189). Šiuo atveju yra kalbama apie besimokančius svetimosios kalbos. Vadinasi, gimtakalbiui tokia žodžių atpažinimo užduotis yra gerokai lengvesnė ir yra didelė tikimybė, kad, žinodamas pamatinę žodžio formą, jis žinos ir išvestines to žodžio formas. Vis dėlto Schmitt'as (2010) pabrėžia, kad dėl praktinių sumetimų gali būti sudėtinga žodyno dydį matuoti žodžių lizdais, nes ne visada galima vienareikšmiai juos nustatyti. Taigi pastaruoju metu mokslininkai žodyno dydžio vienetu laiko lemą, t. y. pagrindinę žodžio formą, kuri gali būti kaitoma (Schmitt 2010: 189; Brysbaert et al. 2016: 2). Tai reiškia, kad žmogus gali suprasti kaitybinės lemos formas, tačiau „<...> skirtingos darybinės formos jau sudaro atskiras lemas“ (Vilkienė et al. 2019: 25).

Žodyno dydžio matavimai buvo organizuojami ir vykdomi įvairiai. Pavyzdžiui, Georg'as Hartmann'as 1946 m. atliko tyrimą, kuriame dalyvavo 106 anglų kalbos gimtakalbiai studentai. Jie turėjo paaiškinti 50 žodžių reikšmes. Į šio žodyno dydžio testą pateko žodžiai, paimti iš žodyno (*Merriam Webster's New International Dictionary 1934*), kurį sudarė 400 000 žodžių. Tyrimui buvo pasirinkta po vieną žodį iš kas keturiasdešimto žodyno puslapio. Į žodžių sąrašą buvo įtraukti ir tikriniai daiktavardžiai, ir kaitybinės bei darybinės žodžių formos, ir sudurtiniai žodžiai. Atlikęs tyrimą Hartmann'as padarė išvadą, kad anglakalbių gimtakalbių studentų žodyno dydis yra lygus 215 000 žodžių (Milton, Treffers-Daller 2013; Brysbaert et al. 2016: 6). Šiuolaikiniai tyrėjai teigia, kad toks žodyno dydis yra gerokai pervertintas. Tokius, manytina, per didelius žodyno dydžio rezultatus lėmė tai, kad į testą įtraukti žodžiai nebuvo tinkamai

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

atrinkti (paprastai į žodyno dydžio testus neįtraukiami tikriniai daiktavardžiai, kaitybinės formos, santrumpos ir pan.). Dėl šios priežasties minėtas tyrimas buvo kritikuotas, teigiant (Coxhead et al. 2015), kad jo nustatyti žodyno dydžio rezultatai nėra tikslūs. Tačiau pabrėžtina, kad tai buvo pirmieji žingsniai, leidę toliau tobulinti žodyno dydžio testavimą.

Remdamiesi ankstesniais tyrimais, mokslininkai Robins Goulden, Paul'as Nation'as ir John'as Read'as (1990) ištestavo 20 išsilavinusių Naujosios Zelandijos anglų kalbos gimtakalbių, kurių amžius buvo apie 22 metus (Goulden et al. 1990). Šie mokslininkai atrinko 208 000 lemų iš tarptautinio Webster'io anglų kalbos žodyno (*Webster's Third New International Dictionary* 1961). Šiuos žodžius jie suskirstė į tris grupes: pamatiniai žodžiai (54 000), darybinės žodžių formos (64 000), sudurtiniai žodžiai (67 000) ir lemos, kurios nepateko nė į vieną iš šių grupių (22 000) (Brysbaert et al. 2016, 5–6). Toks suskirstymas padėjo pamatą tolimesnei Nation'o žodžių klasifikacijai į žodžių lizdus. Toliau į minėtą tyrimą buvo nuspręsta įtraukti 54 000 pamatinių žodžių, iš kurių buvo atrinkti 250 žodžių, o į žodyno dydžio testą įtraukta 50 žodžių. Tiriamųjų grupė turėjo pažymėti tuos žodžius, kuriuos žinojo. Taip buvo nustatyta, kad tyrime dalyvavusių gimtakalbių žodyno dydis siekė nuo 13 200 iki 20 700 žodžių, o žodyno dydžio vidurkis buvo lygus 17 200 žodžių (Goulden et al. 1990). Taigi matyti, kad tokie rezultatai yra gerokai mažesni už ankstesnių tyrimų rezultatus, kurie siekė iki 215 000 žodžių. Tokį žodyno dydžio skirtumą galėjo lemti Goulden ir jos kolegų (1990) metodika.

Minėti tyrimai buvo atlikti analizuojant anglų kalbos gimtakalbių žodyną, tačiau pasakytina, kad matuotas ir vokiečių kalbos gimtakalbio žodyno dydis. Jutta Segbers ir Sacha Schroeder'is (2017) tyrė vokiečių vaikų gimtosios kalbos žodyno dydį. Šie tyrėjai į žodyno dydžio testą iš viso įtraukė 100 žodžių, iš kurių 24 buvo pseudožodžiai, t. y. netikri žodžiai, sukurti lingvistų siekiant gauti kuo objektyvesnius rezultatus ir išvengti respondentų savęs pervertinimo atliekant žodyno dydžio testą. Kadangi testuojamieji buvo ir skirtingo amžiaus vaikai (nuo pirmos iki aštuntos klasės), ir suaugusieji, buvo sukurtos 5 skirtingos testų versijos, pritaikytos kiekvienai iš amžiaus grupių. Atlikę tyrimą šie mokslininkai nustatė, kad vokiečių gimtakalbių vaikų žodyno dydis yra lygus 5 900 lemų, o suaugusiųjų – 73 000 lemų (Segbers, Schroeder 2017: 315). Svarbu pasakyti, kad šiame tyrime dalyvavusių suaugusiųjų amžius buvo apie

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

22 metus. Tačiau, suprantama, kad ir vėliau žodynas yra vis dar kaupiamas. Pasak kai kurių mokslininkų, gimtakalbiai per metus priaugina apie 1 000 žodžių lizdų ir penkerių metų vaikas turi sukaupęs apie 4 000–5 000 žodžių lizdų (Nation, Waring 1997: 7). Vilkienės ir kt. pabrėžiama, kad nors negimtakalbiai per metus taip pat gali papildyti savo žodyną 1 000 žodžių lizdų, vis dėlto jų žodyno dydis bus mažesnis negu gimtakalbių (Vilkienė et al. 2019: 24).

Kitas plataus masto žodyno dydžio matavimas buvo atliktas mokslininkų Emmanuel'io Keuleers'o, Michaël'io Stevens'o, Paweł'o Mander'os ir Marc'o Brysbaert'o (2015), kurie tyrė nyderlandų kalbos gimtakalbius. Šiame tyrime dalyvavo apie 300 000 respondentų (Keuleers et al. 2015), kurie turėjo atsakyti ir į klausimus apie savo lytį, amžių, kalbų mokėjimą, gyvenamąją vietą ir pažymėti savo dominuojančią ranką (ar yra dešiniarankis, ar kairiarankis<sup>2</sup>). Taip pat respondentai turėjo atlikti žodyno dydžio testą, sukurtą pagal Paul'o Meara'os „Yes/No“ modelį. Atliekant šį testą respondentams reikėjo atsakyti *taip*, jeigu žodį žinojo, ir *ne*, jeigu nežinojo, ką žodis reiškia. Svarbu pridurti, kad į šį testą buvo įtraukta ne tik dažniau ir rečiau vartojamų žodžių, bet ir pseudožodžių.

Keuleers'o ir kt. tyrimo rezultatai patvirtino vieną iš iškeltų hipotezių, kad žodyno dydis yra glaudžiai susijęs su amžiumi, t. y., kuo žmogus vyresnis, tuo didesnis gali būti jo žodynas. Pridurtina, kad aptariamo tyrimo dalyvių amžius svyravo nuo 12 iki 72 metų. Be to, nustatyta, kad 13 metų paaugliai vidutiniškai moka apie 29 000 žodžių, o respondentams, kuriems yra daugiau nei 70 metų – apie 43 000 žodžių (lemų). Remiantis šio testo rezultatais buvo padaryta išvada, kad žodynas per 57 gyvenimo metus (nuo 13 iki 70 metų) priauga 14 000 žodžių (Keuleers et al. 2015: 1674). Šiuos duomenis galima palyginti su Marc'o Brysbaert'o ir jo kolegų (2016) atliktu amerikiečių anglų kalbos gimtakalbių žodyno dydžio rezultatais, kurie rodo, kad 20 metų žmogaus žodyno dydis yra 42 000 lemų, o 60 metų JAV gyvenančio žmogaus žodyno dydis – 48 200 lemų. Kaip matyti iš šių duomenų, žodynas per 40 metų praplečiamas maždaug 6 200 lemų (Brysbaert et al. 2016: 8–9). Vadinasi, galima spėti, kad

<sup>2</sup> Paprastai šis aspektas įtraukiamas į psicholingvistinius tyrimus. Į aptariamą tyrimą rankos dominavimas buvo įtrauktas siekiant nustatyti, ar dominuojanti ranka turi įtakos reakcijos greičiui. Atlikus tyrimą nustatyta, kad rankos dominavimas neturėjo jokios įtakos žodyno dydžio variacijai (Keuleers et al. 2015: 1668, 1678).

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbota.lt](http://www.taikomojikalbota.lt).

amžius yra vienas iš žodyno dydžiui įtaką darančių veiksnių, į kurį reikia atsižvelgti tiriant žodyno dydį.

Paminėtina ir tai, kad Keuleers'o ir jo kolegų (2015) žodyno dydžio tyrimo rezultatai parodė, kad taip pat svarbus veiksnys yra ir studijos aukštojoje mokykloje. Šie mokslininkai teigia, kad didžiausias žodyno šuolis įvyksta bakalauro studijų pakopoje (žodynas prasiplečia 3,73 proc. arba 2 000 žodžių) bei magistrantūros studijų pakopoje (žodynas prasiplečia 2,88 proc. arba 1 500 žodžių) (Keuleers et al. 2015: 1674). Žinoma, minėti mokslininkai tokią įžvalgą siejo tik su savo tyrimu, tad analizuojant, kokio dydžio yra kitų kalbų gimtakalbių žodynas, jį turėtų būti tikrinama.

Pridurtina, kad Keuleers'as ir kt. (2015) tyrė svetimųjų kalbų mokėjimo įtaką gimtosios kalbos žodyno dydžiui. Šie tyrėjai išklė hipotezę, kad didesnis mokamų kalbų skaičius gali neigiamai veikti gimtąją kalbą, tai reiškia, kad asmens, kuris moka daugiau svetimųjų kalbų, gimtosios kalbos žodyno dydis gali būti mažesnis negu to, kuris moka mažiau kalbų (Keuleers et al. 2015: 1668). Vis dėlto Keuleers'o ir kolegų tyrimas atskleidė, dvi tendencijas: pirma, paaiškėjo, kad gimtosios kalbos žodynas yra didesnis tų asmenų, kurie moka daugiau svetimųjų kalbų, ir antra, didesnis žodynas buvo tų tiriamųjų, kurie mokėjo anglų, vokiečių arba prancūzų kalbas, giminingas nyderlandų kalbai. Taigi iš esmės svetimųjų kalbų mokėjimas yra prielaida auginti ir gimtosios kalbos žodyną.

Keuleers'o ir kt. (2015) tyrime buvo analizuota ir lyties veiksnio įtaka žodyno dydžiui. Rezultatai parodė, kad lyties veiksnys nelemia reikšmingų skirtumų tarp vyrų ir moterų žodyno dydžio. Tokių lyties veiksnio įtaką žodyno dydžiui tikrinančių tyrimų būta ir daugiau. Pavyzdžiui, Averil Coxhead ir jos kolegų atliktame tyrime, kuriame dalyvavo 13–18 metų mokiniai iš Naujosios Zelandijos, taip pat nebuvo užfiksuotas statistiškai reikšmingas skirtumas tarp merginų ir vaikinių žodyno (Coxhead et al. 2015). Tačiau priešingas rezultatas buvo gautas tiriant lietuvių mokinių žodyno dydį (Vilkienė et al. 2019). Lietuvių kalbos žodyno dydžio tyrime dalyvavo 16–17 metų gimtakalbiai. Paaiškėjo, kad merginų ir vaikinių žodyno dydžiai

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

skiriasi statistiškai reikšmingai. Merginų, kurių gimtoji kalba yra lietuvių, žodyno dydžio vidurkis yra 54 284, o vaikinų – 50 759 žodžiai (lemomis).

Dar vienas veiksnys, apie kurio įtaką kalba mokslininkai, yra skaitymo įpročiai. Kelių mokslininkų (Milton, Treffers-Daller 2013) buvo atliktas tyrimas su 178 anglakalbiais studentais, kurių amžiaus vidurkis buvo 18–19 metų. Vienas iš šio tyrimo tikslų buvo nustatyti, ar skaitymas ir jo dažnumas koreliuoja su žodyno dydžiu, todėl studentų buvo prašoma atsakyti, kaip dažnai jie skaito grožinę literatūrą bei publicistiką. Šie mokslininkai kėlė hipotezę, kad skaitymas (ypač grožinės literatūros ir publicistikos) daro teigiamą poveikį žodyno dydžiui. Vis dėlto tyrimas atskleidė, kad skaitymo ir žodyno dydžio koreliacija nėra didelė. Tačiau šių mokslininkų buvo pasiūlyta toliau tikrinti šį veiksni.

Atliekant šiame straipsnyje aprašomą tyrimą taip pat buvo tikrinta, kurie sociolingvistiniai veiksniai: lytis, amžius, išsilavinimas, mokamų svetimųjų kalbų skaičius, skaitymas savo malonumui, tai, kad respondentas dirba ar nedirba, – daro reikšmingos įtakos suaugusių lietuvių kalbos gimtakalbių žodynui. Apie tai plačiau rašoma 5.4. skyriuje.

### 3. Žodyno dydžio testai

Įvairių šalių mokslininkai sutaria, kad žodyno dydis gali atspindėti bendrą žmogaus kalbos mokėjimo lygį ir net lemti bendrą mokymosi sėkmę (Milton, Treffers-Daller 2013). Taigi, kaip jau buvo minėta, mokslininkai ieško būdų, kaip žodyno dydį išmatuoti. Paprastai tai daroma specialiai sukurtais testais, kuriais galima nustatyti, maždaug kiek žodžių moka asmuo. Dažniausiai tokie testai matuoja žmogaus gebėjimą atpažinti žodžio formą ir susieti ją su reikšme (Coxhead et al. 2015). Vienas iš tokių žodyno dydžio matavimo testų yra VST (*Vocabulary Size Test*)<sup>3</sup>, sukurtas Nation'o ir Beglar'o (2007). Pirmoji šio testo versija leido nustatyti žodyno dydį iki 14 000 žodžių lizdų ir buvo taikyta matuojant negimtakalbių anglų kalbos žodyną. Atlikus šį testą buvo nustatyta, kad negimtakalbio, studijuojančio anglų kalbą pirmojoje studijų pakopoje, žodyno dydis lygus 5 000–6 000 žodžių lizdų, o negimtakalbio

---

<sup>3</sup> Šis žodyno dydžio testas prieinamas tokiu adresu: <https://www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-tests>.



Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuivio gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

doktoranto, studijuojančio anglų kalbą, – 9 000 žodžių lizdų (Schmitt 2010: 199). Šio tipo testai yra sukurti testuoti ne tik anglų kalbos vienakalbius, bet ir dvikalbius, pavyzdžiui, anglų – arabų, anglų – korėjiečių, anglų – rusų ir pan. dvikalbius. Tokiame teste yra pateikiamas žodis su keturiomis jo reikšmės pasirinkimais: testuojamo žodžio apibrėžtimis (viena tinkama ir 3 netinkamos) – gimtakalbiams, o negimtakalbiams greta angliško žodžio, kurio mokėjimas tikrinamas, – 4 jų gimtosios kalbos žodžiai, iš kurių tik vienas yra tinkamas. Testuojamasis turi pažymėti atsakymą. Šio tipo teste reikia atpažinti ne tik žodžio formą, bet ir jo reikšmę. Paminėtina, kad iš esmės šis testas tikrina izoliuoto žodžio reikšmės supratimą. Vienu iš tokio testo trūkumų yra laikoma ilga jo atlikimo trukmė (Nation 2012). Be to, manoma, kad šis testas neparodo, ar testuojamasis galėtų savarankiškai pavartoti duotus žodžius konkrečiame kontekste.

Kitas žodyno dydžio matavimo testo formatas – *taip / ne* testai (angl. *yes / no tests* arba *checklist tests*). Pagrindinis tokio testo pranašumas yra tas, kad jis greitai atliekamas (Schmitt 2010: 200). Šiame teste pateikiamas žodžių sąrašas ir testuojamasis turi pažymėti žodžius, kuriuos žino. Tokio tipo testų atlikimas neužima daug laiko, kadangi testuojamieji neturi žodžių paaiškinti ar kitaip įrodyti, kad juos supranta. Tačiau manoma, kad tokie testai, kaip ir jau aptartas VST, neparodo, kaip gerai yra mokamas žodis, todėl galima nustatyti tik minimaliausią žodžio mokėjimo lygį (Meara 1994: 6).

Taip pat mokslininkai atkreipė dėmesį į tai, kad *taip / ne* testų atveju matyti neobjektyvus testuojamųjų įsivertinimas. Kadangi *taip / ne* testai netikrina bendrų žinių apie pasaulį (Schmitt 2010: 200), visada yra tikimybė, kad vieni asmenys bus linkę save pervertinti ir žymės žodžius, kurių nežino, o kiti atvirkščiai – linkę save nuvertinti, todėl žymės suprantą žodžius tik tada, jeigu juos patys dažnai vartoja ar žino visas jų reikšmes (Meara 1994: 9). Dėl šios priežasties siekiant gauti kuo objektyvesnius rezultatus, į tokį testą įtraukiama pseudožodžių, t. y. kalboje neegzistuojančių darinių, kurie sukuriami taip, kad būtų kuo panašesni į testuojamosios kalbos žodžius. Pridurtina, kad nors kol kas dar nėra vieningai sutarta, koks skaičius pseudožodžių yra tinkamiausias tokiuose žodyno dydžio testuose, dažniausiai netikri žodžiai sudaro apie 25–33 proc. viso testo žodžių (Schmitt 2010: 200).

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvio gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbota.lt](http://www.taikomojikalbota.lt).

Esama įvairių *taip/ne* formato testų. Vienas jų – LexTale<sup>4</sup> (*Lexical Test for Advanced Learners of English*). Šis testo tipas buvo sukurtas siekiant nustatyti asmenų, kurie dalyvauja, psicholingvistiniuose eksperimentuose, kalbos lygį (Izura et al. 2014: 51) ir taikytas testuojant anglų, vokiečių ir olandų (Lemhöfer, Broersma 2012), taip pat prancūzų (Brybaert 2013) ir ispanų (Izura, Cuertos, Brybaert 2014) kalbų žodyno dydį. Manoma, kad toks testas yra tinkamas testuoti tiek gimtakalbius, nes išvengiama lubų efekto<sup>5</sup>, tiek negimtakalbius, nes išvengiama grindų efekto<sup>6</sup> (Izura et al. 2014: 52). Be to, *taip/ne* formato yra ir X\_lex (Meara ir Milton 2005) ir Y\_lex (Meara ir Miralpeix 2006) testai. Abu jie skirti matuoti negimtakalbių žodyno dydį: pirmasis – pradedančiųjų, o antrasis – pažengusiųjų. Yra sukurti tokie įvairių kalbų žodyno dydį matuojantys testai (plačiau žr. Vilkienė et al. 2019).

Taigi esama įvairių žodyno dydžio matavimo testų, o kiekvienas jų turi tiek trūkumų, tiek pranašumų, tačiau vis dėlto tokiais testais galima nustatyti apytikslį tam tikros kalbos vartotojų žodyno dydį.

#### 4. Lietuvių kalbos žodyno dydžio testas

Kaip jau minėta, iki šiol lietuvių kalbos vartotojo žodyno dydis nėra labai plačiai tirtas, tačiau Vilniaus universiteto Taikomosios kalbotyros instituto darbuotojų grupės (žr. Vilkienė et al. 2019) yra sukurtas žodyno dydžio matavimo testas. Kaip teigiama Vilkienės, Vilkaitės-Lozdienės ir Bružaitės-Liseckienės (2019), sudarant lietuvių kalbos žodyno dydžio testą, buvo siekiama derinti LexTALE ir X\_lex (Meara, Milton 2005) bei Y\_lex (Meara, Miralpeix 2006) formatų testų pranašumus, nes lietuviško testo kūrimo tikslas buvo jį taikyti testuojant ne tik lietuvių gimtakalbius, bet ir negimtakalbius mokinius. Lietuviškam žodyno dydžio testui žodžiai buvo atrinkti iš *Dabartinės lietuvių kalbos tekstyno* (KLC 2011)

<sup>4</sup> LexTale testas prieinamas šiuo adresu: <http://www.lextale.com/>.

<sup>5</sup> Lubų efektas – efektas atsiranda, kai, pavyzdžiui, testas yra per lengvas ir didžioji dalis rezultatų yra arti maksimalios rezultatų ribos. Tuomet nėra aišku, kokius rezultatus testuojamieji būtų pasiekę, jeigu maksimaliausia (lubų) riba būtų buvusi nustatyta aukščiau (Tavakoli 2012: 53).

<sup>6</sup> Grindų efektas atsiranda, kai, pavyzdžiui, testas yra per sudėtingas ir didžioji dalis rezultatų yra arti minimaliausias ribos (Tavakoli 2012: 232).

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

publicistikos dalies, o tiksliau iš šios dalies lemuoto dažninio sąrašo. Taigi žodyno dydis tyrime skaičiuotas lemomis<sup>7</sup>. Be to, šiam testui buvo pasirinktas būtent publicistikos dalies žodžių sąrašas, nes daryta prielaida, kad publicistinis stilius apima labai įvairų žodyną, būdingą įvairiems kalbos stiliumams. Iš viso *Dabartinės lietuvių kalbos tekstyno* publicistikos dalies žodžių sąrašą sudarė 771 163 lemos, tačiau buvo atsisakyta tikrinių daiktavardžių, akronimų, romėniškų skaitmenų, santrumpų ir pan. Į lietuvišką žodyno dydžio testą iš viso buvo įtraukta 180 žodžių: 120 žodžių ir 60 pseudožodžių (plačiau žr. Vilkienė et al. 2019).

Kaip jau buvo minėta, lietuvių kalbos žodyno dydis anksčiau nebuvo tirtas, taip pat yra nedaug tyrimų ir su pseudožodžiais<sup>8</sup>. Dėl šios priežasties į lietuvišką žodyno dydžio testą įtraukti pseudožodžiai buvo sukurti pačių minėtų mokslininkų. Pabrėžtina, kad jie išbandyti su gimtakalbiais lietuviais, kurie turėjo įvertinti, kurie iš netikrų žodžių galėtų būti lietuvių kalbos žodžiai. Į žodyno dydžio matavimo testą buvo įtraukta 60 žodžių.

Svarbu paminėti, kad šiame teste užduotis buvo suformuluota klausiant žodžio reikšmės, o ne tik formos atpažinimo. Kaip ir pristatytuose kitų kalbų žodyno dydžio testuose, šiame teste tikrintas minimalus žodžio mokėjimas, t. y. žodžio siejimas su reikšme. Žodyno dydis skaičiuotas pritaikius tokią formulę (Vilkienė et al. 2019):

$$\text{Testo rezultatas} = \text{pažymėtų žodžių skaičius} - \text{pažymėtų netikrų žodžių skaičius} * 667.$$

Iš formulės matyti, kad yra suskaičiuojami visi pažymėti žodžiai, iš jų atimamas pažymėtų pseudožodžių skaičius ir gautas rezultatas dauginamas iš koeficiento 667. Šis koeficientas – tai skaičius atkarpu, į kurias specialios programos automatiškai buvo suskirstytas tyrimo pamatu buvęs lemų sąrašas. Iš kiekvienos šios atkarpos į testą pateko po vieną žodį. Šis žodyno dydžio testas, jau išbandytas Vilniaus gimnazijų tyrime, buvo pasitelktas ir atliekant šiame straipsnyje aprašomą tyrimą.

<sup>7</sup> Straipsnio tekste „lema“ ir „žodis“ vartojami sinonimiškai, „žodis“ suprantamas kaip „lema“.

<sup>8</sup> Eglė Krivickaitė ir Ineta Dabašinskienė (2013) yra vienos pirmųjų mokslininkų Lietuvoje, savo tyrimuose naudojusios pseudožodžius. Jos sukūrė netikrų žodžių pakartojimo testą, kurį sudarė 24 lietuviški pseudožodžiai (Dabašinskienė, Krivickaitė 2013: 243–244).

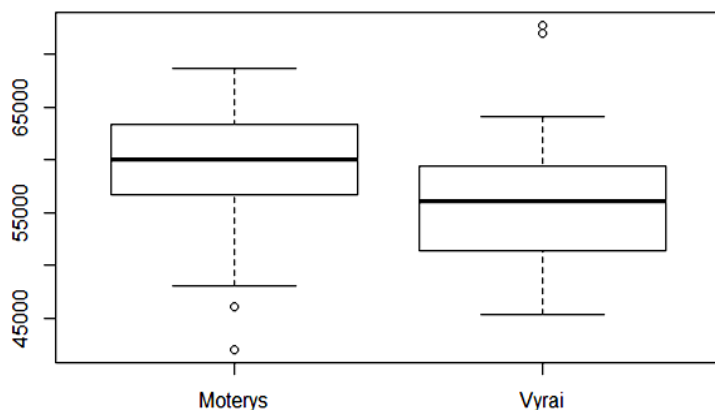
Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, www.taikomojikalbotyra.lt.

## 5. Tyrimo rezultatai

Ieškant atsakymo į tikslo klausimą, pirmiausia buvo keltas uždavinys patikrinti, kokie yra jaunesnės ir vyresnės grupės respondentų žodyno dydžio testo rezultatai. Be to, kaip jau buvo minėta, atliekant tyrimą buvo aiškinamasi, ar rezultatai gali skirtis nelygu respondentų lytis. Vėliau visų tiriamųjų grupių rezultatai buvo bendrai išanalizuoti atliekant regresinę analizę, kad būtų patikrinta, kurie sociolingvistiniai veiksniai daro didžiausią įtaką žodyno dydžiui. Tokie tyrimo žingsniai ir aprašomi tolesniuose poskyriuose.

### 5.1. Jaunesniosios grupės rezultatai

Atlikus JG žodyno testo rezultatų analizę, paaiškėjo, kad šios grupės moterų (JMG) ir vyrų (JVG) žodyno dydžiai skiriasi (žr. 1 pav.).



1 pav. Jaunesniosios moterų ir vyrų grupės žodyno dydžio pasiskirstymas

Iš 1 paveikslo akivaizdu, kad pirmiausia skiriasi abiejų grupių rezultatų mediana: JMG žodyno dydžio mediana – 60 030 žodžiai, o JVG – 56 695 žodžiai, taigi JVG ji yra gerokai žemesnė, o skirtumas tarp abiejų grupių rezultatų medianų – 3 335 žodžiai. Kaip matyti, JMG žodyno dydžio vidurkis (59 040 žodžių) ir mediana yra aukštesni negu JVG (vidurkis – 56 208 žodžiai). Šių dviejų grupių žodyno dydžio vidurkių skirtumo statistinis reikšmingumas buvo patikrintas *Stjudento T-kriterijumi*. Atlikus *T* testą paaiškėjo, kad vis dėlto jaunesniosios moterų ir vyrų grupės žodyno dydžio vidurkių skirtumas statistiškai nereikšmingas ( $p = 0,0648$ ).

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, www.taikomojikalbotyra.lt.

Be to, abiejų grupių duomenų analizė atskleidė, kad rezultatai įvairuoja atsižvelgiant į tai, ar respondentai ne tik studijuoja, bet ir dirba. Paaiškėjo, kad dirbančių JMG žodyno dydis yra didesnis negu nedirbančių (žr. 1 lentelę). Taigi būtų galima manyti, kad kuo platesnė veiklos sfera, pavyzdžiui, aptariamam atveju – ir studijos, ir darbas, – tuo didesnis asmens žodynas.

**1 lentelė.** Žodyno dydžio vidurkis pagal lytį ir darbinę veiklą

	<b>JVG žodyno dydžio vidurkis</b>	<b>JMG žodyno dydžio vidurkis</b>
<b>Dirba</b>	55 194	59 951
<b>Nedirba</b>	58 080	57 933

Tačiau kitokie rezultatai buvo gauti išanalizavus JVG duomenis: pasirodė, kad dirbančių vaikinių žodyno dydis mažesnis negu nedirbančių. Be to, matyti, kad nedirbančių JMG ir JVG žodyno dydžio vidurkių skirtumas yra labai mažas. Paaiškinti tokius rezultatus yra sudėtinga, galima tik numanyti, kad dirbančių merginų žodynas dydis yra didesnis, nes naujų žodžių išmokstama susiduriant su kalba ne tik universitete, skaitant literatūrą, bet ir dirbant įvairiose srityse, tiesiogiai nesusijusiose su studijuojamu dalyku. O vaikinai, galbūt dirba artimoje studijoms srityje, pavyzdžiui, mokslo ar informacinių technologijų, todėl susiduria su mažesne žodyno įvairove. Įžvalga, kad žodynas galbūt gali priklausyti nuo to, ar studentas dirba, ar ne, buvo tikrinama ir atliekant regresinę analizę (žr. 5.4. poskyrį).

Taip pat buvo keltas klausimas, ar skaitymas savo malonumui daro įtaką žodyno dydžiui (plg. Milton, Treffers-Daller 2013). Tyrimo dalyvių buvo klausta ne tik apie grožinės literatūros skaitymą, bet ir apie naujienų portalų bei žurnalų ir laikraščių skaitymo įpročius, nes daryta prielaida, kad bet kuri skaitymo veikla gali turtinti asmens žodyną. Analizuojant duomenis buvo atskleista skaitymo įtaka žodyno dydžiui. Iš tyrimo rezultatų matyti, kad JMG skaito savo malonumui dažniau negu JVG. Be to, galima daryti prielaidą, kad studijuodamos filologės privalo skaityti daug įvairios literatūros, o fizikai galbūt skaito mažiau. Todėl tai galėtų būti viena iš priežasčių, lėmusių didesnę žodyno dydžio vidurkių skirtumą tarp skaitančių dažnai ar visada ir tarp skaitančių retai ar niekada neskaitančių. Taigi skaitymo veiksnio įtaka taip pat buvo tikrinama atliekant regresinę analizę (žr. 5.4. poskyrį).

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, www.taikomojikalbotyra.lt.

Kaip jau minėta anksčiau, remiantis Brysbaert'o ir kt. tyrimu (2016), kuris atskleidė, kad gimtosios kalbos žodyno dydis gali būti didesnis tų asmenų, kurie moka daugiau svetimųjų kalbų, šiame straipsnyje aprašomo tyrimo respondentų buvo klausiama, kiek svetimųjų kalbų jie moka. Pirminė duomenų analizė leido manyti, kad ryškaus svetimųjų kalbų mokėjimo poveikio abiejų aptariamų grupių žodyno dydžiui nebuvo (žr. 2 lentelę).

**2 lentelė.** *JMG ir JVG žodyno dydžio vidurkiai ir svetimiosios kalbos*

	<b>1 svetimoji kalba</b>	<b>2 svetimiosios kalbos</b>	<b>3 svetimiosios kalbos</b>	<b>4 svetimiosios kalbos</b>
<b>JMG</b>	52 026	59 719	59 974	58 696
<b>JVG</b>	54 738	58 529	54 560	49 358

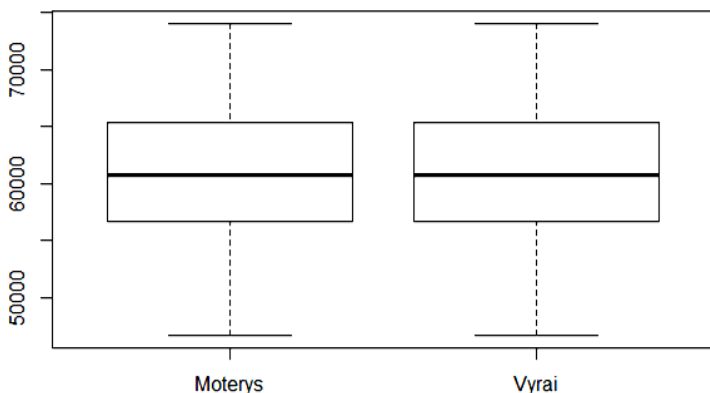
Iš 2 lentelės matyti, kad JMG ir JVG žodyno dydžio vidurkiai pagal mokamų svetimųjų kalbų skaičių skiriasi, bet aiškios tendencijos įvardyti nebuvo įmanoma, kol nebuvo atlikta regresinė duomenų analizė.

Palyginus jaunesniosios grupės rezultatus su kitų kalbų tyrėjų analogiškais tyrimais, atliktais taikant tokią pačią metodiką, galima daryti prielaidą, kad skirtingų kalbų gimtakalbių žodyno dydis įvairuoja. Pavyzdžiui, vokiečių atliktas tyrimas atskleidė, kad 22 metų vokiečių kalbos gimtakalbis moka apie 73 000 žodžių (lemų) (Segbers, Schroeder 2017), JAV dvidešimtmetis gimtakalbis – 42 000 (lemų) (Brysbaert et al. 2016), o šis tyrimas, kaip jau minėta, rodo, kad 22 metų gimtakalbis lietuvis moka apie 57 400 žodžių (lemų). Žinoma, pabrėžtina, kad tokie duomenys gauti ištyrus santykiškai nedidelę respondentų grupę ir turi būti tikrinami.

## **5.2. Vyresniosios grupės rezultatai**

Atlikus VG duomenų analizę ir atsižvelgus į lyties veiksnių buvo nustatyta, kad tiek vyresnių moterų, tiek vyrų žodyno dydis yra labai panašus. Atlikus *T* testą, paaiškėjo, kad VMG ir VVG žodyno dydžio vidurkių skirtumas yra statistiškai nereikšmingas ( $p = 0.7239$ ) (žr. 2 pav.).

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvio gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).



**2 pav.** VMG ir VVG žodyno dydžio pasiskirstymas

Iš 2 paveikslo akivaizdu, kad beveik nesiskiria abiejų lyčių grupių rezultatų mediana: VMG – 60 697 žodžiai, o VVG – 60 030 žodžių, taigi skirtumas yra mažas – 667 žodžiai. Atkreiptinas dėmesys, kad žodyno dydžio vidurkiai taip pat beveik nesiskiria: VMG – 60 484, o VVG – 61 027 žodžiai. Praktiškai sutampa ir viršutinė bei apatinė žodyno dydžio ribos.

Remiantis mokslininkų įžvalgomis (Keuleers et al. 2015), kad išsilavinimas gali turėti įtakos žodyno dydžiui, analizuota, ką rodo tiriamųjų duomenys. Rezultatai apibendrinti 3 lentelėje.

**3 lentelė.** Žodyno dydis ir išsilavinimas

	Vidurinis išsilavinimas	Profesinis išsilavinimas	Aukštesnysis išsilavinimas	BA laipsnis	MA laipsnis	Daktaro laipsnis
<b>VMG</b>	51 359*	60 697*	58 563	60 363	62 888	–
<b>VVG</b>	–	58 453	58 483	60 520	65 766	59 363

\* Pažymėjo tik viena moteris

Iš 3 lentelės akivaizdu, kad žodyno dydžio vidurkių rezultatai, atsižvelgiant į išsilavinimo veiksnį, yra labai panašūs abiejų lyčių grupėse. Tačiau pabrėžtina tai, kad tiek VMG, tiek VVG, žodyno dydžio vidurkiai kyla priklausomai nuo išsilavinimo laipsnio. Pavyzdžiui, magistro laipsnį turinčių VVG tiriamųjų žodyno dydis yra 5 516 žodžių didesnis negu VVG tiriamųjų, kurie turi bakalauro laipsnį. O tiriamųjų, kurie turi bakalauro laipsnį, žodyno dydis (tiek VMG, tiek VVG) yra didesnis negu tų, kurie turi aukštesnį išsilavinimą. Taigi iš šių duomenų matyti

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

tendencijos, apie kurias kalba Keuleers'as ir kt. (2015): aukštasis išsilavinimas, taip pat aukštojo išsilavinimo laipsnis turi įtakos žodyno dydžiui ir, be to, vieni didžiausių žodyno dydžio šuolių įvyksta būtent bakalauro ir magistrantūros studijų metais. Šį įžvalgą taip pat buvo tikrinama atliekant regresinę analizę.

Atlikus tyrimą, išvelgta ir dar viena tendencija. Kaip jau minėta, šiame straipsnyje buvo remtasi Keuleers'as ir jo kolegų (2015) įžvalga, kad kelių svetimųjų kalbų mokėjimas gali turėti teigiamos įtakos gimtosios kalbos žodyno dydžiui. Abiejų vyresniųjų grupių – VMG ir VVG – duomenys patvirtina tokią mintį (žr. 4 lentelę).

**4 lentelė.** VMG ir VVG žodyno dydžio vidurkiai ir svetimųjų kalbų skaičius

	1 svetimoji kalba	2 svetimosios kalbos	3 svetimosios kalbos
<b>VMG</b>	58 758	59 196	65 651
<b>VVG</b>	58 938	60 897	64 603

Iš 4 lentelės akivaizdu, kad VMG ir VVG didžiausias žodyno dydžio vidurkis yra tų tiriamųjų, kurie moka tris svetimąsias kalbas, mažesnis – tų, kurie moka dvi svetimąsias kalbas, o mažiausias tų tiriamųjų, kurie pažymėjo moką vieną svetimąją kalbą. Vadinasi, šiame darbe aprašomas vyresniųjų grupės tyrimas atspindi Keuleers'o ir jo kolegų (2015) įžvalgą. Atkreiptinas dėmesys į tai, kad Keuleers'o ir kt. tiriamieji dažniausiai mokėjo giminingas kalbas, o šio tyrimo atvejis yra kitoks – respondentų mokamos svetimosios kalbos nėra giminingos lietuvių kalbai. Tokios mokamų svetimųjų kalbų skaičiaus įtakos gimtosios kalbos žodynui reikšmingumas taip pat buvo tikrinamas regresine analize (žr. 5.4).

VG grupės rezultatų analizė atskleidė ir tai, kad didesnis žodyno dydžio vidurkis yra respondentų, kurie dažniau skaito, o ypač grožinę literatūrą (žr. 5 lentelę).



Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

**5 lentelė.** VMG ir VVG žodyno dydžio vidurkis ir skaitymas savo malonumui

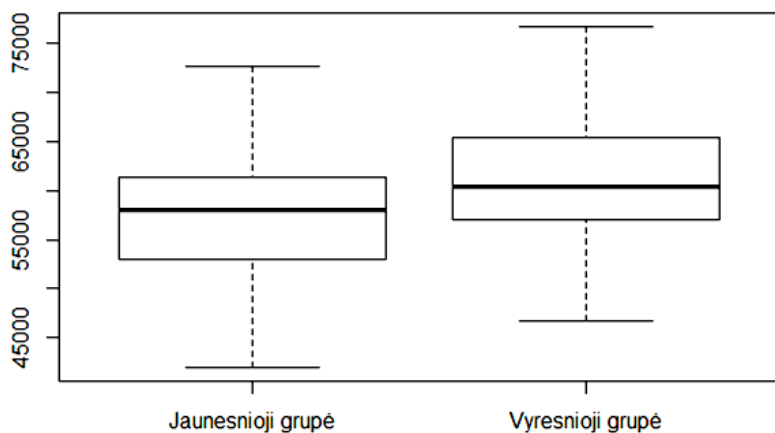
		<b>Visada, dažnai skaitančių žodyno dydžio vidurkis</b>	<b>Kartais skaitančių žodyno dydžio vidurkis</b>	<b>Retai skaitančių, niekada neskaitančių žodyno dydžio vidurkis</b>
<b>VMG</b>	<b>Grožinė literatūra</b>	60 648	61 630	53 026
	<b>Naujienų portalai internete</b>	61 486	56 695	–
	<b>Žurnalai, laikraščiai</b>	60 030	63 031	51 581
<b>VVG</b>	<b>Grožinė literatūra</b>	58 904	64 820	58 981
	<b>Naujienų portalai internete</b>	60 923	68 034	56 861
	<b>Žurnalai, laikraščiai</b>	61 736	59 963	59 918

Iš 5 lentelės matyti dar viena tendencija, susijusi skaitymu malonumui. Abiejų tiriamųjų grupių respondentų žodyno dydžio vidurkiai mažesni tuo atveju, kai skaitoma retai ar niekada to nedaroma. Iš to galima daryti prielaidą, kad skaitymas gali turėti įtakos žodyno dydžiui, tai irgi buvo tikrinama atliekant regresinę analizę.

### 5.3. Skirtingų grupių rezultatų palyginimas

Remiantis mokslininkų atliktais tyrimais (Keuleers et al. 2015, Brysbaert et al. 2016, Segbers ir Schroeder 2017), kurie atskleidė, kad amžius gali turėti įtakos žodyno dydžiui, t. y. amžiui didėjant, didėja ir žodynas, buvo nuspręsta palyginti skirtingų amžiaus grupių tiriamųjų rezultatus būtent šiuo aspektu. Atlikus jaunesnių ir vyresnių respondentų grupių žodyno dydžio testų rezultatų analizę, buvo rasti šių grupių žodyno dydžio skirtumai (žr. 3 pav.).

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbota.lt](http://www.taikomojikalbota.lt).



**3 pav.** Jaunesniosios ir vyresniosios grupės žodyno dydžio pasiskirstymas

Iš 3 paveikslo akivaizdu, kad skiriasi abiejų grupių medianos: jaunesniosios grupės žodyno dydžio mediana – 58 029 žodžiai, vyresniosios – 60 364 žodžiai, taigi skirtumas – 2 335 žodžiai. Taip pat skiriasi ir žodyno dydžio vidurkiai: JG – 57 499 žodžiai, o VG – 60 772 žodžiai. Siekiant nustatyti šio skirtumo statistinį reikšmingumą, buvo atliktas *T* testas ir paaiškėjo, kad abiejų grupių žodyno dydžio vidurkių skirtumas yra statistiškai reikšmingas ( $p = 0,002^{***}$ ). Pridėtina ir tai, kad VG pasiekta minimali ir maksimali riba yra aukštesnė negu JG. Didžiausias JG žodyno dydis lygus 72 703 žodžiams, o VG – 76 705 žodžiams, taigi skirtumas – 4 002 žodžiai. Mažiausias JG žodyno dydis – 42 021, o VG – 46 690 žodžių, skirtumas – 4 669 žodžiai. Kaip matyti iš žodyno testo rezultatų, VG žodyno dydis yra didesnis negu JG.

Šie rezultatai taip pat buvo palyginti su Vilkienės et al. (2019) atliktu tyrimu su mokiniais (žr. 6 lentelę).

**6 lentelė.** Žodyno dydis pagal amžių

	Žodyno vidurkis	Maksimalus žodynas	Minimalus žodynas
Mokiniai (apie 17 m.)	52 799	64 032	32 016
Studentai (apie 22 m.)	57 499	72 703	42 021
Suaugusieji (vidurkis 49 m.)	60 772	76 705	46 690

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

Iš 6 lentelės akivaizdu, kad žodynas auga su amžiumi. Kaip matyti iš šių duomenų, mokinių ir studentų žodyno dydžio vidurkių skirtumas – 4 700 žodžių, o mokinių ir suaugusiųjų (49 m.) – 7 973 žodžiai. Iš to galima daryti prielaidą, kad vyresniame amžiuje žodyno dydis auga, bet lėčiau<sup>9</sup>. Be to, palyginus šių amžiaus grupių žodyno dydžio vidurkių skirtumus, matyti, kad žodyno dydžio skirtumas yra mažiausias (3 273 žodžiai) tarp abiejų suaugusiųjų grupių (22 m. ir 49 m.). Žinoma, didžiausias žodyno dydžio skirtumų atotrūkis matomas tarp mokinių ir vyresniosios grupės (49 m.). Tai pat reikėtų turėti omenyje, kad didžioji dalis šių suaugusiųjų tiriamųjų yra išsilavinę, o studentai – mokosi universitete, taigi dar nėra įgiję nei BA išsilavinimo, vadinasi, tai gali turėti įtakos žodyno dydžiui. Kita vertus, iš šių duomenų įžvelgtina tendencija, kad vis dėlto amžius ir per tą laiko tarpą sukaupta patirtis (svetimųjų kalbų mokėjimas, išsilavinimas, darbas tam tikroje srityje ir t. t.) daro poveikį žodynui. Taigi spėtina, kad įvairių sociolingvistinių veiksnių visuma galėtų veikti asmens žodyno dydį. Tačiau vis dėlto atkreiptinas dėmesys į tai, kad amžiaus veiksnys yra itin glaudžiai susijęs su išsilavinimu: didesnė dalis VG respondentų buvo ir aukštesnio išsilavinimo.

#### 5.4. Sociolingvistinių veiksnių įtaka žodyno dydžiui

Siekiant nustatyti, kurie iš tirtų sociolingvistinių veiksnių daro įtaką žodyno dydžiui, atlikta regresinė analizė<sup>10</sup>. Tam buvo kartu sudėti abiejų grupių rezultatai. Į pradinį modelį buvo įtraukti visi šie nepriklausomi kintamieji: respondentų lytis, amžius, mokamų svetimųjų kalbų skaičius, išsilavinimas, skaitymo įpročiai (grožinės literatūros, naujienų portalų, laikraščių, žurnalų ir kt.) savo malonumui, taip pat – dirbama ar nedirbama. Pradedant nuo šio pradinio modelio, buvo pažingsniui po vieną atmetami nereikšmingi veiksniai, kol modelyje liko tik reikšmingi kintamieji. Kaip priklausomas kintamasis buvo žodyno dydžio testo rezultatas, išreikštas mokamų žodžių skaičiumi. Gavus galutinį modelį patikrinta, ar reikšmingos galimos sąveikos tarp modelio kintamųjų. Atlikus pirminę analizę, paaiškėjo, kad amžiaus ir išsilavinimo veiksniai yra artimai susiję: šio tyrimo tendencija – vyresni respondentai turi aukštesnį išsilavinimą. Dėl

<sup>9</sup> Yra atlikta tyrimų (pvz., Brysbaert ir kt. 2016; Keuleers ir kt. 2015), kurių dalies respondentų amžius buvo 70–72 metai. Šių tyrimų rezultatai rodo, kad asmens žodynas auga net ir tokiame amžiuje. Kita vertus, kol kas dar nėra atlikta platesnio masto tyrimų, kurių respondentai būtų buvę vyresnio nei 70 metų amžiaus. Taigi dar nėra išsiaiškinta, kaip keičiasi, ar keičiasi žodyno dydis žmogaus, perkopusio per 70 metų ribą.

<sup>10</sup> Plačiau apie regresinę analizę žr.

[http://www.lidata.eu/index.php?file=files/mokymai/trast/trast.html&course\\_file=trast\\_turinys.html](http://www.lidata.eu/index.php?file=files/mokymai/trast/trast.html&course_file=trast_turinys.html).

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

stiprios šių veiksnių koreliacijos, abiejų veiksnių palikimas modelyje galėtų būti problemiškas. Be to, amžiaus veiksnys pradiniam modelyje reikšmingas nebuvo, taigi galutiniame modelyje liko tik išsilavinimo veiksnys. Taip pat paaiškėjo, kad iš pradinio modelio turi būti pašalinti lyties, naujienų portalų, laikraščių bei žurnalų ir kt. skaitymo kintamieji, o paliktas tik grožinės literatūros skaitymas. Jokios įtakos žodyno dydžiui taip pat nerodė tai, ar respondentas dirba, ar ne. Galutinis modelis apibendrintas 7 lentelėje.

**7 lentelė.** Žodyno dydžio rezultatų modelis

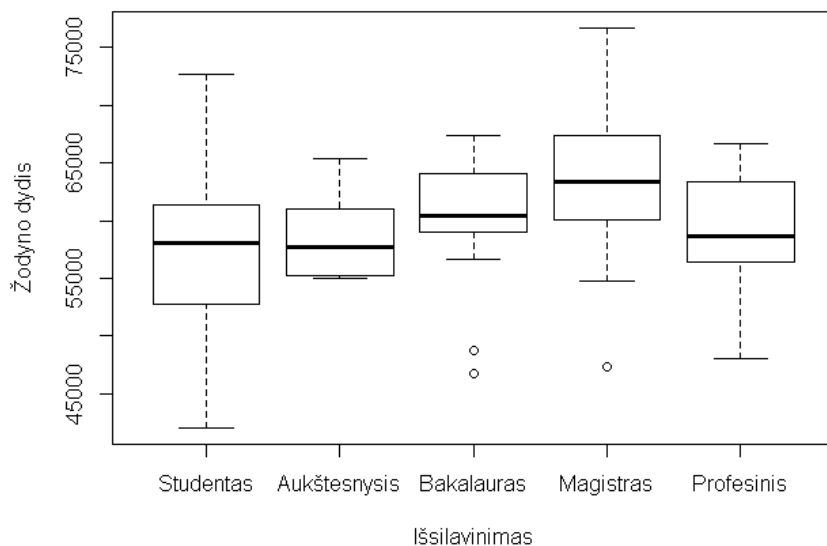
<b>Fiksuoti modelio parametrai</b>	<b>Parametrų įverčiai (standartinė paklaida), parametro reikšmingumas</b>
b <sub>0</sub> (laisvasis narys)	51650,7 (1580,2), $p = 0,000^{***}$
b <sub>1</sub> (mokamų svetimųjų kalbų skaičius)	1840,7 (643,0), $p = 0,005^{**}$
b <sub>2</sub> (išsilavinimas – aukštesnysis)	571,3 (2166,0), $p = 0,792$
b <sub>3</sub> (išsilavinimas – BA)	2887,3 (1508,2), $p = 0,058$
b <sub>4</sub> (išsilavinimas – MA)	5119,3 (1446,7), $p = 0,001^{***}$
b <sub>5</sub> (išsilavinimas – profesinis)	2989,5 (1829,6), $p = 0,105$
b <sub>6</sub> (grožinės literatūros skaitymas kartais)	4313,5 (1348,7), $p = 0,002$
b <sub>7</sub> (grožinės literatūros skaitymas dažnai ar visada)	2905,1 (1252,8), $p = 0,022^*$

$p < 0,001$  – \*\*\*;  $p < 0,01$  – \*\*;  $p < 0,05$  – \*

Iš 7 lentelės akivaizdu, kad reikšmingi yra išsilavinimo (žr. 4 pav.), svetimųjų kalbų mokėjimo (žr. 5 pav.), ir grožinės literatūros skaitymo (žr. 6 pav.) veiksniai. Pabrėžtina, kad išsilavinimo požiūriu statistiškai modelyje išsiskiria tik magistro laipsnį turinti grupė. Kitaip tariant, tik magistro pakopos išsilavinimas daro įtaką žodyno dydžiui<sup>11</sup>. O skaitymo aspektu reikšmingas ir įprotis grožinę literatūrą skaityti kartais, dažnai ir visada.

<sup>11</sup> Primintina, kad atskirai netirtas daktaro laipsnį turinčių asmenų žodynas. Taigi gali būti tikrinama, ar ši aukštesnė išsilavinimo pakopa daro įtaką žodyno dydžiui.

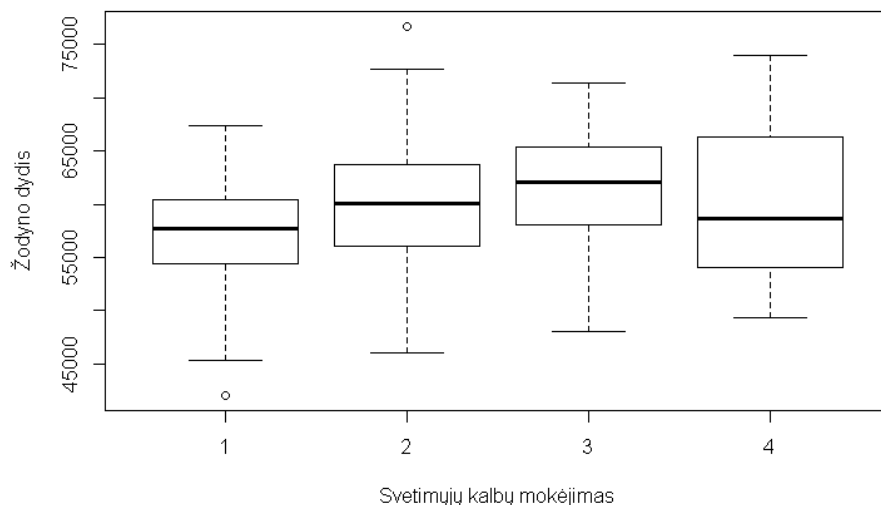
Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).



**4 pav.** Išsilavinimas ir žodyno dydžio rezultatų pasiskirstymas

Kaip matyti iš 4 paveikslo, didžiausias žodynas yra tų asmenų, kurie yra įgiję magistro laipsnį: tiek rezultatų mediana, tiek viršutinė riba gerokai pranoksta visų kitų išsilavinimo grupių rezultatus. Ir nors statistiškai modelyje (žr. 6 lentelę) išsiskiria tik magistro išsilavinimą turinčių asmenų grupė, iš 4 paveikslo aišku, kad turinčių bakalauro išsilavinimą žodyno dydžio rezultatų mediana, nors yra žemiau nei magistro (atliekant regresinę analizę į šią grupę buvo įtraukti ir 2 tiriamieji, turintys daktaro laipsnį, nes atskirai tirti daktaro laipsnį turinčių asmenų atvejį nepakako duomenų) išsilavinimo grupės, bet lenkia ir studentų, ir aukštesniojo bei profesinio (atliekant regresinę analizę į šią grupę įtrauktas ir 1 vidurinį išsilavinimą turintis asmuo) išsilavinimo grupių rezultatus. Tačiau ši įžvalga turėtų būti tikrinama atlikus papildomą analizę, o toks uždavinys šiame tyrime nebuvo keliamas. Universiteto studentų rezultatų mediana yra tokia pati, kaip aukštesniojo ar profesinio išsilavinimo grupių. Taigi galima daryti išvadą, kad tiriamosios medžiagos analizė rodo, kad kuo aukštesnis išsilavinimas, tuo didesnis žodynas. Matyti, kad tai susiję ir su amžiaus veiksniu, kuris negali būti visiškai atsietas nuo išsilavinimo (plg. Keuleers et al. 2015).

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

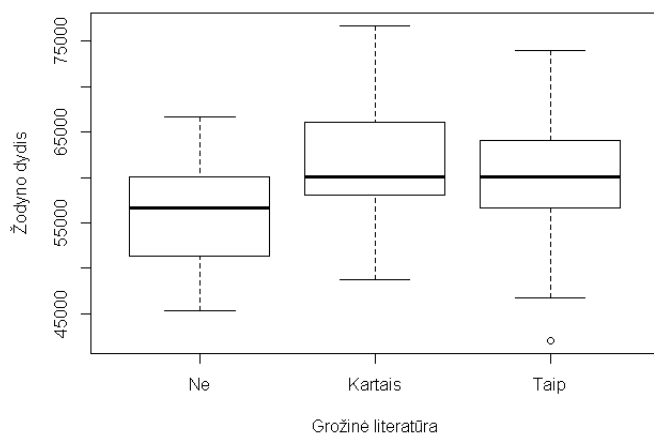


**5 pav.** Mokamų svetimųjų kalbų skaičius ir žodyno dydžio rezultatų pasiskirstymas

Dar akivaizdžiau žodyno dydžio sąsajos matyti su mokamų svetimųjų kalbų skaičiumi. Iš 5 paveikslą matyti, kad 1, 2 ar 3 svetimąsias kalbas mokančių grupės rezultatų medianos tarytum sudaro kylančių laiptų pakopas. Kitaip tariant, kuo daugiau svetimųjų kalbų mokama, tuo didesnis tos tiriamųjų grupės daugumos gimtosios kalbos žodynas. Tai patvirtina ir jau aptartas Keuleers'o ir kt. (2015) išvalgas. Tiesa, stačiakampė diagrama, vaizduojanti 4 svetimąsias kalbas mokančių grupės rezultatus, negali patvirtinti tokios išvalgos, nes šioje grupėje buvo iš viso 3 žmonės. Taigi išvelgti tendencijos iš esmės nebuvo galima.

Atliekant regresinę analizę respondentų atsakymai į klausimą, ar jie skaito grožinę literatūrą savo malonumui, buvo parengti taip: į vieną sulieti atsakymai *niekada* ir *retai* – pavadinta *ne* –, taip pat sulieti ir atsakymai, kad skaitoma *dažnai*, *visada* – pavadinta *taip*. Kaip atskira grupė analizuoti atsakymai *kartais*.

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).



**6 pav.** Grožinės literatūros skaitymas ir žodyno dydžio rezultatų pasiskirstymas

Kaip matyti iš 6 paveikslo, tų, kurie teigė, kad nors kartais arba dažnai, visada skaito grožinę literatūrą, žodyno dydžio mediana yra panaši ir aukštesnė nei tų, kurie tai daro retai arba visai neskaito. Pastarosios grupės viršutinė žodyno dydžio riba yra gerokai žemiau nei kitų dviejų grupių. Primintina, kad Milton ir Treffers-Daller (2013) buvo iškelta hipotezė, jog grožinės literatūros ir net publicistikos tekstų skaitymas galbūt daro įtaką žodyno dydžiui, tačiau vėliau padaryta išvada, kad įtaka nėra labai didelė. Šiame straipsnyje aprašomų duomenų analizė rodo, kad publicistikos skaitymas statistiškai reikšmingos įtakos nedaro, tačiau grožinės literatūros skaitymas bent kartais yra reikšmingas veiksnys žodyno dydžio atžvilgiu.

## 6. Išvados

Atlikus tyrimą, galima daryti tokias išvadas:

1. Jaunesnio amžiaus suaugusių (apie 22 m.) gimtakalbių, studijuojančių aukštojoje mokykloje, lietuvių kalbos žodyno dydžio vidurkis apytiksliai 57 490 žodžių (lemų), tačiau nustatyta ir žodyno dydžio variacija nuo 42 021 iki 72 703 žodžių (lemų). Matyti, kad tai gana platus – apie 30 000 žodžių – rėžis. Vyresnio amžiaus suaugusių (amžiaus vidurkis 49 m.) įvairaus išsilavinimo gimtakalbių lietuvių kalbos žodyno dydžio vidurkis yra maždaug 60 770 žodžių (lemų), o variacija nuo 46 690 iki 76 705 žodžių (lemų), taigi rėžis taip pat apie 30 000 žodžių. Kaip akivaizdu iš duomenų, VG žodyno dydžio vidurkis skiriasi nuo JG apie 3 000 žodžių.

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

Analogiškai keliais tūkstančiai žodžių didesnis VG ir minimalaus, ir maksimalaus žodyno dydis. Palyginus tiriamųjų suaugusiųjų apytikslį žodyno vidurkį su mokinių (16–17 m.) žodyno dydžio vidurkiu (Vilkienė et al. 2019), paaiškėjo, kad suaugusio žmogaus žodynas auga ne taip greitai, kaip tai vyksta laikotarpiu, kai baigiama bendrojo ugdymo mokykla ir imama studijuoti. Palyginimui: 16–17 metų gimtakalbio mokinio, gyvenančio Vilniuje, apytikslis žodyno vidurkis 52 790 žodžių, o variacija nuo 32 016 iki 64 032 žodžių. Taigi akivaizdu, kad nuo 16 iki 22 metų asmens žodyno vidurkis gali padidėti apie 4 700 žodžių, minimalus žodynas gali paaugti apie 10 000, o maksimalus – maždaug 8 680 žodžių. Tai yra gerokai spartesnis augimas negu tas, kuris buvo rastas palyginus JG ir VG žodyno dydžius.

2. Regresinė duomenų analizė atskleidė, kad tiriamuoju atveju įtaką žodyno dydžiui darė šie veiksniai: magistro pakopos išsilavinimas, didesnis mokamų svetimųjų kalbų skaičius ir grožinės literatūros skaitymas kartais, dažnai ar visada. Kaip buvo atrasta atliekant regresinę analizę, amžiaus ir išsilavinimo veiksniai stipriai koreliuoja, todėl jų abiejų palikimas modelyje galėtų būti problemiškas. Primintina, kad amžiaus veiksnys pradiniam modelyje nebuvo reikšmingas, todėl galutiniame modelyje liko tik išsilavinimo veiksnys.

Taigi šiame straipsnyje aprašomo tyrimo duomenys iš dalies patvirtino Keuleers'o ir kt. (2015) įžvelgtą tendenciją, kad kuo aukštesnis asmens išsilavinimo laipsnis, tuo didesnis jo žodynas. Tačiau šiame tyrime atlikta regresinė analizė atskleidė, kad tik magistro pakopos išsilavinimas daro įtaką žodyno dydžiui, nors taip pat buvo atrasta, kad bakalaurų žodynas yra didesnis negu tų, kurie turi aukštesnį išsilavinimą ar profesinį išsilavinimą, tačiau ši įžvalga turi būti tikrinama atliekant papildomą analizę.

Taip pat buvo patvirtinta Keuleers'o ir kt. (2015) mintis, kad žodynas yra didesnis tų asmenų, kurie moka daugiau svetimųjų kalbų. Kaip atskleidė duomenų analizė, mažiausias žodynas buvo tų tiriamųjų, kurie mokėjo vieną kalbą. Didėjant mokamų kalbų skaičiui, didėjo ir respondentų žodynas. Tačiau Keuleers'o ir kt. (2015) buvo teigiama, kad tokia žodyno augimo sąsaja su svetimųjų kalbų mokėjimu yra tada, kai svetimosios kalbos giminingos žmogaus gimtajai kalbai. Šiame straipsnyje aprašomo tyrimo duomenys leidžia manyti, kad nebūtinai mokamos kalbos turi būti giminingos gimtajai kalbai: tiriamieji, kurių gimtoji kalba yra lietuvių, deklaravo mokantys anglų, vokiečių, net japonų ir kt. kalbas. Taigi lemiamą įtaką, spėtina, iš esmės daro tik mokamų kalbų skaičius.



Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, www.taikomojikalbotyra.lt.

Milton, Treffers-Daller (2013), tyrusių publicistikos ir grožinės literatūros skaitymo įtaką žodyno dydžiui, nebuvo išvelgta stipri koreliacija tarp skaitymo dažnumo ir žodyno dydžio, tačiau minėtos mokslininkės ragino ir toliau nagrinėti šią sąsają. Šiame tyrime atlikta regresinė analizė atskleidė, kad publicistikos skaitymas nedaro reikšmingos įtakos žodyno dydžiui, tačiau tokią įtaką daro grožinės literatūros skaitymas. Kitaip tariant, žodynas buvo didesnis tų tiriamųjų, kurie nors kartais arba dažnai, visada skaito grožinę literatūrą.

Baigiant pasakytina, kad šio straipsnio autorės gerai supranta, kad tyrimas atliktas su palyginti nedidele respondentų imtimi, tad visos išvados ir išvalgos turi būti tikrinamos plečiant tiriamųjų grupes, pavyzdžiui, į jaunesnę grupę įtraukiant ir kitų specialybių studentus, o vyresniųjų grupę dar labiau skaidant į smulkesnius pogrupius pagal mažesnes amžiaus atkarpas. Be to, vyresniųjų tiriamųjų grupės turėtų būti sudaromos ir atsižvelgus į išsilavinimo veiksnį. Taip pat svarbu pažymėti, kad į straipsnyje aprašytą vyresnio amžiaus tiriamųjų grupę nepateko asmenys, peržengę 65 metų ribą. Tai galėtų būti dar viena tikslinė tiriamųjų grupė, kuri leistų atidžiau patikrinti amžiaus veiksnio įtaką žodynui. Tačiau vis dėlto pasakytina, kad šis tyrimas ir išvados – dar vienas žingsnis atliekant lietuvių gimtosios kalbos žodyno dydžio matavimą ir gali būti įdomus ir naudingas srities mokslininkams.

## Literatūra

- Brysbaert, M., M. Stevens, P. Mandera, E. Keuleers. 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*, 1–11.
- Coxhead, A., P. Nation, D. Slim. 2015. *Measuring Vocabulary Size of Native Speakers of English in New Zealand Secondary Schools*. Springer, New Zealand Association for Research in Education. Vol. 50, No.1, 121–135.
- Goulden, R., P. Nation, J. Read. 1990. How large can a receptive vocabulary be? *Applied Linguistics*, 11 (4). Oxford University Press, 341–363.
- Izura, C., Cuetos, F. ir Brysbaert, M. 2014. Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, 35, 49–66.

- Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).
- Keuleers, E., M. Stevens, P. Mandera, M. Brysbaert. 2015. Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The quarterly journal of experimental psychology*. Vol. 68, No. 8, 1665–1692.
- Krivickaitė, E., I. Dabašinskienė. 2013. Lietuvių kalbos fonologijos įsisavinimas: netikrų žodžių kartojimo eksperimentinis tyrimas. *Kalbos Kultūra*, 86, 237–255.
- Laufer, B., Z. Goldstein. 2004. Testing vocabulary knowledge: size, strength and computer adaptiveness. *Language learning*, 54 (3), 399–436.
- Meara, P. 1994. The complexities of simple vocabulary tests. Prieiga internetu: <http://www.lognostics.co.uk/vlibrary/meara1994a.pdf>
- Milton, J., J. Treffers-Daller. 2013. Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4 (1).
- Nation, P., R. Waring. 1997. Vocabulary size, text coverage and word lists. *Vocabulary: Description, Acquisition and Pedagogy*. N. Schmitt, M. McCarthy (eds.), Cambridge: Cambridge University Press, 6–19.
- Nation, I. S. P. 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63 (1), 59–82.
- Nation, P. 2012b, October 23. The Vocabulary Size Test. Prieiga internetu: <https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf>
- Read, J. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Schoonen, R., M. Verhallen. 2008. The assessment of deep word knowledge in young first and second language learners. *Language testing*, 25, 211–236.
- Schmitt, N. 2010. *Researching Vocabulary. A Vocabulary Research Manual*. University of Nottingham, UK, Palgrave Macmillan.
- Segbers, J., S. Schroeder. 2017. How many words do children know? A corpus-based estimation of children's total vocabulary size. *Language Testing*, 34 (3), 297–320.
- Tavakoli, H. 2012. *A Dictionary of Research Methodology and Statistics in Applied Linguistics*. Iran: Ranhama Press.
- Vilkienė L., L. Vilkaitė-Lozdienė, J. Bružaitė-Liseckienė. 2019. *Lietuvių kalba Vilniaus lietuviškose, lenkiškose ir rusiškose gimnazijose: mokėjimo kokybė, kalbinės nuostatos ir motyvacija*. Vilnius: Vilniaus universitetas. Publikuojama.

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

## **Suaugusio lietuvių gimtosios kalbos žodyno dydis**

Karolina Andriuškevičiūtė, Loreta Vilkienė

### **Santrauka**

Šio straipsnio tikslas atsakyti į klausimus, koks yra suaugusio lietuvių gimtosios kalbos žodyno dydis ir kokie sociolingvistiniai veiksniai gali daryti jam reikšmingos įtakos. Analizuojamos 132 tiriamųjų žodyno dydžio testo atliktys ir atsakymai į sociolingvistinį klausimyną. Duomenys apdoroti statistine *R-Studio* programa, atliktas *T* testas ir regresinė analizė. Tiriamieji buvo suskirstyti dvi grupes. Jaunesniųjų tiriamųjų grupėje (amžiaus vidurkis 22 m.) buvo 68 asmenys: 31 mergina ir 37 vaikinai. Vyresniųjų grupę (amžiaus vidurkis 49 m.) sudarė 64 asmenys: 30 moterų ir 34 vyrai.

Nustatyta, kad jaunesnio amžiaus suaugusių gimtakalbių lietuvių kalbos žodyno vidurkis apytiksliai 57 490 žodžių (lemų), o žodyno dydžio variacija nuo 42 021 iki 72 703 žodžių (lemų). Vyresnio amžiaus suaugusių gimtakalbių lietuvių kalbos žodyno vidurkis yra maždaug 60 770 žodžių (lemų), o variacija nuo 46 690 iki 76 705 žodžių (lemų). Regresinė duomenų analizė atskleidė, kad įtakos žodyno dydžiui turi šie veiksniai: magistrantūros pakopos išsilavinimas, mokamų svetimųjų kalbų skaičius ir grožinės literatūros skaitymas. Tyrimo duomenys atskleidė, kad magistro laipsnį turinčių respondentų žodynas yra didesnis nei visų kitų tiriamųjų. Taip pat atrasta, kad žodynas yra didesnis tų asmenų, kurie moka daugiau svetimųjų kalbų, kurios nebūtinai turi būti giminingos gimtajai kalbai. Regresinė analizė atskleidė, kad reikšmingą įtaką žodyno dydžiui daro ir grožinės literatūros skaitymas: žodynas buvo didesnis tų tiriamųjų, kurie nors kartais arba dažnai, visada skaito grožinę literatūrą. Tyrimas atliktas su palyginti nedidele respondentų imtimi, tad visos išvados ir įžvalgos turi būti tikrinamos plečiant tiriamųjų grupes.

Andriuškevičiūtė, K., L. Vilkienė. 2019. Suaugusio lietuvių gimtosios kalbos žodyno dydis. *Taikomoji kalbotyra*, 12: 154–181, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

## **Vocabulary size of an adult Lithuanian native speaker**

Karolina Andriuškevičiūtė, Loreta Vilkienė

### **Summary**

This paper aims to answer the question about the vocabulary size of an adult native speaker of the Lithuanian language and to check which sociolinguistic factors have a significant impact on the size of the vocabulary. The material for this study consists of 132 responses given to a sociolinguistic survey and a vocabulary size test. The data was analysed using the statistical R-Studio program. T-test and regression analysis were also applied in the research. Participants of the study were divided into two age groups (the younger and the older) as the adult age range is very broad. In total, the younger group consisted of 68 participants: 31 females and 37 males. The average age of this group is about 22. The older group consisted of 64 participants: 30 females and 34 males. The average age of this group is about 49.

The study has revealed that the vocabulary of a younger native speaker (about 22 years old) is about 57,490 words and that the vocabulary size varies between 42,021 and 72,703 words. It has been found that the vocabulary of an older native speaker (about 49 years old) is about 60,770 and that the vocabulary size varies between 46,690 and 76,705 words. Regression analysis data shows that such factors as education, knowledge of foreign languages and reading fiction have a significant impact on the size of vocabulary. The study has also revealed that the size of the vocabulary is significantly larger for participants with a Master's degree. Furthermore, the vocabulary size is larger for those individuals who know several languages which may not necessarily share any vocabulary with the Lithuanian language. Regression analysis data also indicates that reading fiction can significantly influence the size of vocabulary and that the vocabulary is larger for those participants who read sometimes, often or daily. It must be noted that the sample of participants in this study is relatively small, so all the findings and insights need to be verified by expanding the study groups.

**Keywords:** vocabulary size; native language; the impact of sociolinguistic factors

Įteikta 2019 m. rugsėjis  
Publikuota 2019 m. lapkritį