

**Jolanta Kovalevskaitė**

Vytauto Didžiojo universitetas, Lietuva

**Erika Rimkutė**

Vytauto Didžiojo universitetas, Lietuva

## **MOKOMASIS LIETUVIŲ KALBOS VARTOSENOS LEKSIKONAS – NAUJA TEKSTYNO PAGRINDU PARENGTA LEKSINĖ BAZĖ**

---

*Kol nemoki žodžio vartoti, tol jo nežinai.* (Jimmie Hill, 2000, p. 60)

**Santrauka.** Šiame straipsnyje pristatomas tekstynų lingvistikos metodu parengtas lietuvių kaip svetimosios kalbos mokymui ir mokymuisi skirtas išteklius – *Mokomasis lietuvių kalbos vartosenos leksikonas*. Tai – pirma tokio pobūdžio lietuvių kalbos leksinė bazė, kurios antraštynas ir žodžių vartosenos aprašas pagrįstas konkrečiu tekstynu – *Mokomojo tekstyno* rašytine dalimi (ją sudaro apie 620 tūkst. žodžių, sakytinės dalies dydis – 50 tūkst. žodžių). Leksikone pateikta 3 700 vienažodžių ir keliažodžių leksinių vienetų (sudėtinių pavadinimų, frazeologizmų, posakių). Dažniausiems ir visuose A1–B2 kalbos lygiuose pavartotiems žodžiams (jų yra apie 700) parengti ilgieji aprašai – juose pateikiami su konkrečiomis reikšmėmis susieti vartosenos modeliai, kuriuose užfiksuota leksinė, semantinė ir gramatinė informacija. Kiekvienam vartosenos modeliui pateikta pavyzdžių, kuriais iliustruoti visi modeliuose užfiksuoti vartosenos požymiai. Retesniems dariniams (vediniams ir dūriniams), pavartotiems tekстыne ir susijusiems su dažniausiai antraštyno žodžiais, parengti trumpieji aprašai su pavyzdžiais. Iš viso leksikone yra 2 542 dariniai, susieti su 940 pamatinių žodžių. Visiems antraštyno leksiniams vienetams pateikiama 28 550 pavyzdžių, skirtų produkciniais gebėjimams ugdyti. Straipsnyje išsamiai aptariami *Mokomojo lietuvių kalbos vartosenos leksikono* požymiai, privalumai ir trūkumai, daug dėmesio skiriama adaptuotam vartosenos modelių metodui pristatyti. Tikimasi, kad ši leksinė bazė, tobulinama pagal vartotojų poreikius, užpildys dažną esamų lietuvių kalbos leksikografinių išteklių spragą – gerų produkavimui skirtų pavyzdžių stoką, o kartu su esamais ištekliais – vadovėliais, žodynais ir tekstynais – bus naudingas mokymo ir mokymosi šaltinis pažengusiųjų kalbos vartotojų grupei.

**Pagrindinės sąvokos:** leksikonas; vartosenos modelių metodas; tekstynas; tekstynų lingvistika; mokomoji leksikografija; lietuvių kalba.

### **Įvadas**

Vykdamt projektą „Užsienio baltistikos centrų ir Lietuvos mokslo ir studijų institucijų bendradarbiavimo skatinimas“ (Nr. 09.3.1-ESFA-V-709-01-0002) parengtos elektroninės mokomosios priemonės (žr. portalą kalbu.vdu.lt). Viena iš šių priemonių – *Mokomasis lietuvių kalbos vartosenos leksikonas* (toliau – leksikonas). **Leksikonas yra leksinė duomenų bazė, kurioje pateiktas**

**konkrečiu tekstynu pagrįstas kalbos vartosenos aprašas, skirtas praktinėms kalbos mokymo(si) reikmėms. Leksikonas orientuotas į pažengusių vartotojų grupę (A2–B2 kalbos lygi) ir dėstytojus, vadovėlių rengėjus.**

Lietuvių kaip užsienio kalbos mokymosi poreikis yra labai išaugęs, vis dėlto iki šiol ji daugiau buvo dėstoma pradedančiųjų lygio kursuose, todėl šiai tikslinei grupei tinkamos mokymosi medžiagos ir išteklių yra daugiau negu pažengusiems ir aukštesnių kalbos lygių studentams (pvz., vadovėlių B lygiui daug mažiau negu A lygiui).

Visada svarbu žinoti, kokiais principais remiantis sudarytas leksikografinis išteklius. Taigi ir šio straipsnio tikslas – pristatyti leksikoną, jo rengimo metodiką, naudotus išteklius, išbandytus naujus sprendimus. Siekiame pradėti diskusiją su būsimais vartotojais: studentais, dėstytojais, vadovėlių rengėjais ir kt.; aptarsime ne tik naujas galimybes, bet ir ribotumus. Dabartinių lietuvių kalbos žodynų kontekste leksikonas gali atrodyti gana netipiškas dėl kai kurių sudarymo aspektų, todėl norime išsamiau juos paaiškinti.

**Pirmiausia, leksikonas parengtas remiantis konkrečiu tekstynu, o leksikono antraštynas neperimtas iš jokio kito žodyno**, vadinasi, tai visiškai naujas leksikografinis resursas (angl. *created from scratch*). Iki šiol lietuvių leksikografijoje tekstynai dažniausiai naudoti tik pavyzdžiams rinkti, pvz., rengiant *Bendrinės lietuvių kalbos žodyną* (toliau – BLKŽ). Leksikonui sudaryti naudotas *Mokomasis tekstynas*<sup>1</sup> (toliau – tekstynas) (žr. Kovalevskaitė et al., 2020b; Boizou et al., 2020).

Šis tekstynas sudarytas 2017–2019 m. vykdant anksčiau minėtą projektą. Tekstynė yra autentiškos lietuvių kalbos vartosenos tekstų, atrinktų pagal kriterijus, kad būtų suprantami ir aktualūs skirtingų lygių besimokantieji. Visi tekstai pagal *Bendruosius Europos kalbų mokymosi, mokymo ir vertinimo metmenis*<sup>2</sup> suklasifikuoti į A1, A2, B1 ir B2 kalbos lygius. Tekstyną sudaro skirtingų kalbos atmainų (rašytiniai ir transkribuoti sakytiniai) tekstai: 111 000 žodžių A1–A2 lygio tekstų (96 000 rašytinės kalbos ir 15 000 sakytinės kalbos); 558 000 žodžių B1–B2 lygio tekstų

---

<sup>1</sup> <https://kalbu.vdu.lt/mokymosi-priemones/mokomasis-tekstynas/>

<sup>2</sup> <https://www.coe.int/en/web/common-european-framework-reference-languages>

(523 000 rašytinės kalbos ir 35 000 sakytinės kalbos); iš viso 669 000 žodžių. Šis tekstynas skiriasi nuo mokinių tekstynų, kurie yra dažniau kalbos mokymui panaudojamas resursas (apie lietuvių ir latvių kalbos mokinių tekstynus plačiau žr. Darģis et al., 2020; Ruzaitė et al., 2020; Kovalevskaitė et al. 2022).

**Antra, leksikonas nėra gatavas mokomasis žodynas, tai – leksinė bazė, tačiau ją galima laikyti aktyviojo tipo mokomojo žodyno prototipu.** Lietuvių leksikografijoje leksinė duomenų bazė naudota rengiant BLKŽ, bet kaip antrinė, o ne pirminė žodyno rengimo fazė (Murmulaitytė, 2009). Rengiant leksikoną buvo siekiama sukurti duomenų bazę, kurioje lietuvių kalbos kaip svetimosis dėstytojai ir lietuvių kalbos jau pramokę studentai rastų kuo daugiau informacijos apie žodžio vartoseną – autentišką, būdingą dabartinei lietuvių kalbai ir aktualią kalbos mokymuisi. Todėl tekstynas tapo leksikono pagrindu – tekstynas yra šaltinis, kuriame nustatyta aprašomųjų žodžių gramatinė bei leksinė aplinka ir pateikta vartosenos modelių pavidalu.

Vartosenos modeliais atskleidžiamos ir aprašomo žodžio reikšmės. Laikydami tekstynų lingvistikos principų, reikšmę suvokiame kaip žodį su artimiausiu jo kontekstu ir manome, kad žodžio reikšmė išryškėja iš jo vartosenos, todėl **leksikone skyrėme reikšmes, bet nepateikėme reikšmių aiškinimų.** Šis naujas principas yra trečiasis leksikono aspektas, skiriantis nuo kitų žodynų. Įprasta, kad mokomuosiuose žodynuose apibrėžtys pateikiamos žodžiais iš pagrindinio žodyno (angl. *basic vocabulary*). Pavyzdžiui, *Mokomąjį lietuvių kalbos žodyną* (Norkaitienė et al., 2000) galima naudoti pagrindinio žodyno žodžių reikšmėms išsiaiškinti. Kita vertus, jame neretai trūksta daugiau vartosenos pavyzdžių, leksinio junglumo informacijos. Kiti aiškinamieji lietuvių kalbos žodynai taip pat ne visada atspindi dabartinę tipinę vartoseną, juose daug sudėtingos ir neretai jau senstelėjusios leksikos, trūksta gerų pavyzdžių. Leksikone vartosenos pavyzdžių pateikiama gausiai, iliustruojami visi tekстыne pastebėti leksinės ir gramatinės aplinkos dėsniniai. *Mokomasis lietuvių kalbos žodynas* ir leksikonas gali būti vienas kitą papildantys mokomosios leksikografijos ištekliai: leksikonas – aktyviojo tipo, žodynas – labiau pasyviojo tipo.

Leksikonas, kurį rengė penkių tekstynų lingvistų komanda, yra apie dvejus metus trukusio teorinio ir praktinio darbo rezultatas. Antroje straipsnio

dalyje „Leksikono sudarymo principai ir metodika“ pristatysime teorinio leksikografinio darbo rezultata – leksikono sudarymo principus: taikytą vartosenos modelių analizės metodiką, medžiagos atranką ir tekstyno analizės būdus. Trečioje dalyje „Leksikono ypatybės“ aptarsime leksikono ypatybes, kurias nulėmė taikyti sudarymo principai; paminėsime privalumus ir trūkumus norėdamos atkreipti leksikografų dėmesį, kokių kilo iššūkių ir kaip juos sprendėme. Reikia pripažinti, kad antraštyno sandarą ir vartosenos aprašą labai lėmė būtent naudotas tekstynas, todėl, kalbėdamos apie leksikono duomenis, trečiojoje dalyje neretai juos susiesime su naudoto tekstyno požymiais.

### **Leksikono sudarymo principai ir metodika**

Tyrimuose apie negimtakalbių leksinę kompetenciją ir leksikos mokymąsi minima, kad mokėti vartoti tam tikrą žodį reiškia žinoti jo formą, reikšmę ir suprasti vartoseną, pvz., kaip rašomas ir tariamas, kokios jo kaitybinės formos, su kokiais žodžiais ir kokiuose junginiuose šis žodis įprastai vartojamas ir pan. (plačiau Nation, 2013, cituojama iš Barclay et al., 2019, p. 803–804). Leksikone pateikiama žodžio<sup>3</sup> **vartosenos informacija apima duomenis apie:**

- žodžio formą (kaip rašomas, kaip tariamas, kaip kirčiuojamas ir transkribuojamas);
- žodžio kaitybą (morfologinės formos, kuriomis tekстыne vartojamas aprašomas žodis) ir darybą (konkrečios reikšmės vediniai ir dūriniai, kurie yra tekстыne);
- žodžio reikšmę: reikšmė matyti iš vartosenos modelio (-ių), kuriame (-iuose) fiksuojama reikšmės gramatinė (gramatinis junglumas) ir leksinė aplinka (ne rečiau kaip 3 kartus su aprašomuoju žodžiu tekстыne pavartoti žodžiai – kolokatai);
- konkrečios žodžio reikšmės vartosenos pavyzdžius, kuriuose atsispindi vartosenos modeliuose užfiksuota informacija.

---

<sup>3</sup> Didžiąją leksikono antraštyno dalį sudaro vienažodžiai leksiniai vienetai, tačiau į leksikono antraštyną įtraukta ir keliažodžių leksinių vienetų (tokių pastoviųjų žodžių junginių kaip sudėtiniai pavadinimai, frazeologizmai, posakiai).

---

Leksikone antraštyne pateikiami keturių kalbos dalių žodžiai: veiksmažodžiai, daiktavardžiai, būdvardžiai irrieveiksmiai (apie tai, kaip žodžiai atrinkti į antraštyną, žr. dalį „Leksikono ypatybės“). Pateikti keli ir kitų kalbos dalių žodžiai, pvz., skaitvardžiai, kurie pagal taikytą antraštyno atrankos metodą pateko tarp dažniausiai vartojamų žodžių (tai yra dažni kelintiniai ir daiktavardiškai vartojami kiekiniai skaitvardžiai): *pirmas, antras, trečias, šimtas, tūkstantis, milijonas*. Nekaitomąsias kalbos dalis (pvz., dalelytes, prielinksnius) aprašome tuo atveju, jeigu jų forma sutampa su viena iš savarankiškos kalbos dalies forma (veiksmažodžiu, daiktavardžiu), pvz., *vietoj* (prieveiksmis ir prielinksnis) sutampa su daiktavardžio *vieta* forma. Tokia nekaitomųjų kalbos dalių strategija pasirinkta dėl besimokančiųjų, nes, tarkim, mokantis veiksmažodžio *vadintis*, svarbu žinoti, kad viena iš šio žodžio formų sutampa su dalelytės *vdinasi* forma.

### **Kodėl pasirinkta vartosenos modelių analizė?**

Sudarant leksikoną, taikytas tekstynų lingvistikos metodas ir **laikytasi tekstynų lingvistikos požiūrio į reikšmės vienetą**: reikšmės raiška yra ne pavieniai žodžiai, o žodžiai su artimiausiu kontekstu (Sinclair, 1991). Artimiausią kontekstą galima iširti tekстыne ir nustatytus dėsningumus aprašyti. Leksikone tam pasirinktas vartosenos modelių metodas. Vartosenos modeliai yra su reikšmėmis susieta informacija, kuria parodoma, iš kokių leksinių ir gramatinių elementų įprastai konstruojama konkreti reikšmė.

**Vartosenos modelių analizė** (angl. *Corpus Pattern Analysis* (toliau – CPA), žr. Hanks, 2004) – vienas iš tekstynų inspiruotos (angl. *corpus-driven*) analizės metodų. Dažnai iš žodynų informacijos neaišku, kaip atskirti vieną reikšmę nuo kitos, nes pateikiama per mažai informacijos apie žodžio ar žodžio reikšmės vartosenos kontekstą (Hanks, 2004, p. 91). Pagrindinis šio metodo taikymo tikslas leksikone – pateikti daugiau informacijos apie reikšmei būdingą vartosenos kontekstą, nes tai gali padėti geriau atskirti (vdinasi, suprasti ir vartoti) reikšmes. Šis metodas pagrįstas mintimi, kad skirtingos reikšmės susijusios su skirtingais vartosenos modeliais: Patricko Hankso (2004, p. 90) teigimu, daugelis žodžių glaudžiai susiję tik su labai mažai tam tikrų tipinių

vartosenos modelių. Taikant šį metodą, reikšmė ir nustatoma, ir aprašoma pagal jai būdingą leksinę bei gramatinę aplinką, todėl vartosenos modelis atspindi ir gramatinį, ir leksinį junglumą.

**Leksinis junglumas labai aktualus leksikono tikslinei grupei** – negimtakalbiams, kurie mokosi lietuvių kalbos. Valentingumo žodynai, nurodomi kaip tipiniai junglumo ir žodžių vartosenos žodynai (Jakaitienė, 2005), leksinio junglumo neparodo. Pasak Rūtos Petrauskaitės (Marcinkevičienė, 2000, p. 40–41), valentingumo tyrimo metodologija apsiriboja tik bendrosiomis junglumo tendencijomis – nurodoma, kokių semantinių klasių (ar atskiri) žodžiai gali būti tiriamųjų predikacinių žodžių partneriai, tačiau taip aprašoma tik labai bendra žodžio reikšmė; norint nustatyti individualią reikšmę (tai svarbu praktiniams kalbos vartojimo tikslams), reikalinga visos distribucijos analizė. Tekstynų lingvistikos metodas leidžia sukaupti informaciją apie leksinį junglumą ir parodyti, su kokiais konkrečiais tam tikrų semantinių grupių žodžiais tiriamasis žodis įprastai vartojamas. Specialūs leksinį junglumą atspindintys žodynai yra kolokacijų žodynai, pvz., *Lietuvių kalbos kolokacijų žodynas* (Bielinskienė et al., 2019); kolokacijų žodynų apžvalgą žr. Rimkutė et al., 2012).

Vartosenos modelių analizė taikyta pirmiausia veiksmažodžių aprašams. Pritaikius vartosenos modelių analizę rengiant leksikoną, **junglumas suvokiamas plačiau už valentingumą**: aprašomas sintaksinis junglumas (žr. Labutis, 2002) parodant ir būtinuosius, ir nebūtinuosius partnerius. Svarbu paminėti, kad sintaksinio junglumo aprašas leksikone skiriasi: vienoks jis taikomas veiksmažodžiams, kitoks – daiktavardžiams, būdvardžiams irrieveiksmiams. Šioms kalbos dalims, pasirinktas principas parodyti tik minimaliuosius pastarųjų modelius, t. y. artimiausius sintaksinius partnerius, pvz., junginyje *mąstyti AIŠKIAI apie Obj\_acc*, aprašomasrieveiksmis *aiškiai* susijęs su *mąstyti*, bet prielinksnine konstrukcija išreikštas objektas yra su veiksmažodžiu susijęs dėmuo, todėl jis neįtraukiamas į aprašomojorieveiksmio modelį: [AIŠKIAI] [Pred]. Žinoma, dėl tokio sprendimo skirtingi būdvardžiai, daiktavardžiai irrieveiksmiai turi daug gramatiškai vienodų modelių (pvz., atributas ir modifikatorius), tačiau vartosenos skirtumus atskleidžia semantikos ir leksikos lygmuo.

## Kaip buvo atliekama vartosenos modelių analizė?

Kitaip nei daryta rengiant *Anglų kalbos veiksmažodžių vartosenos modelių žodyną*<sup>4</sup>, leksikone vartosenos modelis sudarytas iš trijų lygmenų: iš gramatinės, semantinės ir leksinės modelio raiškos (laužtiniuose skliaustuose surašyti kiekvieno lygmens komponentai). Skliaustai ir atskiria modelio komponentus, ir kartu susieja trijuose lygmenyse pateikiamą informaciją (žr. 1 lentelę).

### 1 lentelė

*Modelio raiška trimis lygmenimis*

Lygmuo / Reikšmė	1 reikšmė	2 reikšmė
Gramatinis lygmuo	[Sub] [SKAMBINTI_imp] [Obj_ins]	[Sub] [SKAMBINTI] [Obj_ins]
Semantinis lygmuo	[asmuo] [Pred] [iranga]	[asmuo] [Pred] [muzikos instrumentas]
Leksinis lygmuo	[ ] [SKAMBINTI_imp] [telefonas]	[ ] [SKAMBINTI] [pianinas]

**Gramatiniame lygmenyje** fiksuojami komponentai yra **sintaksinės funkcijos**: Sub – subjektas, Obj – objektas, Pred – vientisinis predikatas, PredN – vardažodinis predikatas, PredV – veiksmažodinis predikatas, AtrA – derinamasis atributas, AtrN – nederinamasis atributas, Adv – aplinkybės (šios funkcijos perimtos iš lietuvių kalbos sintaksiškai anotuoto tekstyno ALKSNIS<sup>5</sup>).

Gramatiniame lygmenyje prie sintaksinių kategorijų žymimos ir vartosenai aktualios **morfolginės ypatybės**, pvz.: Obj\_ins – objektas, reiškiamas įnagininko linksniu. Fleksinė lietuvių kalbos prigimtis reikalauja atsižvelgti į jos žodžių formas (Marcinkevičienė, 2000, p. 34); ši informacija svarbi ir negimtakalbiams, todėl iš morfolginių ypatybių modeliuose dažnai nurodoma: linksnis, skaičius, tam tikra veiksmažodžio

<sup>4</sup> <http://www.pdev.org.uk/>

<sup>5</sup> Žr. <https://clarin.vdu.lt/xmlui/handle/20.500.11821/21>; leksikone naudojama dar keletas kitų funkcijų, jos išsamiai aprašytos Bielinskienė et al., 2021. <https://kalbu.vdu.lt/wp-content/uploads/2022/02/Sintaksiniu-funkciju-pazymos.pdf>

forma, pvz., iš vartosenos matyti, kad 1-ąja reikšme veiksmažodis *skambinti* dažniausiai vartojamas liepiamąja nuosaka: *Jei turite informacijos, skambinkite ekrano apačioje nurodytu telefonu*. Modelyje komponentais gali eiti ir konkretūs žodžiai, ir jų formos, pvz.: 1) [Pred] [x valandą AtrN LAIKAS\_ins]: *Pirmasis mačas vyks sekmadienį, 20.15 val. Lietuvos laiku*. 2) [Pred] [lygus\_pl.ins DALIS\_pl.ins]: *Lygiomis dalimis sumaišykite mėtus ir ramunėlių žiedus*.

**Semantiniame lygmenyje** nurodomos leksinės semantinės grupės. Pagal vartosenos modelių analizės metodą paprastai naudojama iš anksto pasirinkta ontologija. Mes rinkomės mišrų būdą: dalį ontologijos pasirinkome iš anksto (metodas *nuo viršaus į apačią*) *Praktinės gramatikos* (Ramonienė et al., 2003) pagrindu, kitą dalį užpildėme analizuodami vartosena (aprašydami žodžius) (metodas *nuo apačios į viršų*). Būdvardžių, veiksmažodžių kategorijų sąrašas buvo baigtinis, o daiktavardžių kategorijų sąrašas – tik pirminis, jis buvo nuolat pildomas atsižvelgiant į tai, kokių semantinių grupių prireikdavo konkretaus žodžio junglumui aprašyti.

Būdvardžiai suskirstyti į tris grupes: fizinius (pvz.: *aukštas, mažas, stiprus*), vertinamuosius (pvz.: *geras, gražus, malonus*) ir rūšinius (pvz.: *medinis, fizinis*). Veiksmažodžiai klasifikuoti į žyminčius aktyvų veiksmą (pvz.: *bėgti, skaityti, daryti*), savaiminį procesą (pvz.: *lyti, snigti, temti*) arba būseną (pvz.: *mylėti, gulėti, tikėtis*). Daiktavardžiai semantiškai klasifikuoti smulkiau pagal tai, kokios ypatybės išryškėdavo iš konteksto, pvz.: asmuo, kūno dalis, tapatybė, vardas, profesija, rūšis, kolektyvas, šalis, augalas, daiktas, vieta, matmuo, veiksmas ir pan. Kai kurios iš šių grupių klasifikuotos dar smulkiau.

**Leksiniame lygmenyje** nurodomi kolokatai. Pavyzdžiui, 1 lentelėje 1 reikšmės modelyje objekto raiška yra įrangos semantinės grupės daiktavardžiai, vienas iš jų – daugiau nei 3 kartus tekstyne pavartotas žodis *telefonas*, todėl jis nurodytas kolokatų eilutėje. 2 reikšmės modelyje kaip objektai vartojami muzikos instrumentai, konkrečiai – *pianinas*. Subjekto pozicijoje abiejuose modeliuose vartojami įvairūs daiktavardžiai iš asmenų semantinės grupės, nė vienas iš jų nepavartotas daugiau nei 3 kartus, todėl kolokatų nenurodyta. Žinoma, daugiau kolokatų galima tikėtis prie dažniau vartojamų žodžių, dažnesniuose modeliuose. Kita vertus, net ir rečiau vartojamų žodžių kolokatų eilutė gali būti užpildyta, jeigu šis žodis turi ribotą



leksinį junglumą (plg. *skambinti* abiejų reikšmių modelių kolokatus). Kolokatai buvo nustatyti automatiškai naudojant įrankį *WordSketch* tekstynų analizės programoje *SketchEngine*<sup>6</sup>, kuri buvo panaudota vartosenos modelių analizei automatizuoti. Kaip tai atlikta, trumpai aptarsime toliau.

Šiame projekte lingvistai, prieš pradėdami konkretaus žodžio modelių aprašymą, jau turėjo bendrą vaizdą apie gramatinį bei leksinį junglumą: tam reikėjo iš dalies **automatizuoti žodžio konteksto analizę**. Tekstynų analizės programoje *SketchEngine* yra įrankis *WordSketch*, galintis padėti nustatyti analizuojamam žodžiui būdingus žodžių junginius. Kadangi mūsų naudotas tekstynas buvo anotuotas tik morfologiškai, bet ne sintaksiškai, siekiant gauti tikslesnius duomenis apie gramatinį junglumą, lietuvių kalbai buvo parengta vadinamoji *SketchGrammar* (plačiau apie jos principus Kovalevskaitė et al., 2020a). Šioje gramatikoje, remiantis morfolagine informacija, buvo nurodyti lietuvių kalbos žodžių junginiams būdingi sintaksiniai ryšiai.

## 1 paveikslas

Informacija apie veiksmožodžio reikšti vartoseną, pateikta *WordSketch*

Panel	Relation	Word	Frequency	Label	
has_nom_noun	sis	7	8.59	reikšti	
	žodis	6	10.42	reikšti	
	tas	6	8.45	reikšti	
	jis	6	5.26	reikšti	
	spalva	4	9.85	reikšti	
	vardas	3	9.94	reikšti	
	kas	3	6.57	reikšti	
	kuris	3	0.14	reikšti	
	šiaudas	2	9.64	reikšti	
	sapnas	2	9.64	reikšti	
has_adv_modifier	nebūtinai	2	10.82	reikšti	
	maždaug	2	9.57	reikšti	
	dažnai	2	7.05	reikšti	
	daug	2	5.56	reikšti	
	kūdikiskai	1	10.6	reikšti	
	iš pradžių	1	9	reikšti	
	išties	1	8.43	reikšti	
	tarsi	1	8.05	reikšti	
	kaip	1	7.76	reikšti	
	anksti	1	7.67	reikšti	
has_ins_noun	idioma	1	11.67	reikšti	
	sportininkas	1	11.54	reikšti	
	amerikietis	1	11.54	reikšti	
	lietuvi	1	11.41	reikšti	
	moteris	1	8.93	reikšti	
	savęs	1	8.24	reikšti	
	laikas	1	7.7	reikšti	
	tas	1	7.15	reikšti	
	has_inf_compl	būti	6	7.4	reikšti
		taupyti	3	11.34	reikšti
gyventi		3	8.29	reikšti	
nenugriūti		1	10.24	reikšti	
nelankyti		1	10.24	reikšti	
nebeturėti		1	10.19	reikšti	
badauti		1	10.19	reikšti	
sapnuoti		1	10.19	reikšti	
nusivilti		1	10.14	reikšti	
tarnauti		1	10.04	reikšti	

1 paveiksle matyti dalis informacijos, pateiktos *WordSketch* apie žodį *reikšti*: iš jos lingvistai susidarė vaizdą, koks leksinis ir gramatinis junglumas būdingas aprašomam žodžiui. Atskirose kortelėse matyti būdingi sintaksiniai ryšiai, pvz.,

<sup>6</sup> <https://www.sketchengine.eu/>

šiam veiksmažodžiui būdingi subjektiniai (*žodis, spalva, vardas reiškia*), aplinkybiniai (*nebūtinai, maždaug, daug reiškia*), objektiniai (*ką reiškia; reiškia taupyti, gyventi*) junginiai. Peržiūrėjus duomenis, aišku, kad dažnai vartojami junginiai, kuriuose subjektas sudaromas su veiksmažodžio bendratimi, pvz.: *ką reiškia gyventi, ką reiškia nusivilti ir ką reiškia būti sportininku, ką reiškia būti lietuve*. Objektas taip pat gali būti reiškiamas veiksmažodžio bendratimi, pvz.: *taupyti šilumą reiškia taupyti pinigus*. Apibendrinus šią informaciją, parengti 4 modeliai, kuriais realizuojama pirmoji *reikšti* reikšmė „žymi, nurodo“ (žr. 2 lentelę). Kai *reikšti* vartojamas subjektiniuose junginiuose, kuriuose gali būti kiekio aplinkybė, taip realizuojama kita *reikšti* reikšmė – „turi vertę“, pvz., *tai man daug reiškia* (tiesa, ši reikšmė retesnė, rastas tik vienas vartosenos modelis) (žr. 2 lentelę).

## 2 lentelė

*Veiksmažodžio reikšti (187 pavartojimo atvejai) aprašas leksikone*

<b>1 REIKŠMĖ</b>
<b>1.1 modelis</b>
<p>Gramatinis lygmuo: [Obj_acc] [REIKŠTI] [Sub_inf][Sub_inf+noun_ins]                  Semantinis lygmuo: [ ] [Pred] [būsena aktyvus][būsena+asmuo]                  Kolokatai: [kas] [REIKŠTI] [būti, gyventi] ][Sub_inf+noun_ins]                  Pavyzdžiai: <i>Noriu išsiaiškinti, ką reiškia būti lietuve – kas tai yra.</i>  <i>Jūs – gyvas pavyzdys, ką reiškia būti amerikiečiu.</i>  <i>Turėjau ... parodyti, ką reiškia būti tikra moterimi.</i>  <i>Ar žinote, ką reiškia gyvent vaikų namuose?</i>  <i>Ką reiškia nelankyti mokyklos dėl socialinių priežasčių?</i>  <i>Savaitės pabaigoje patirsite, ką reiškia nusivilti...</i></p>
<b>1.2 modelis</b>
<p>Gramatinis lygmuo: [Sub_inf] [REIKŠTI] [Obj_inf]                  Semantinis lygmuo: [aktyvus būsena] [Pred] [aktyvus būsena]                  Kolokatai: (nėra)                  Pavyzdžiai: <i>Taupyti šilumą reiškia taupyti pinigus, taupyti pinigus reiškia uždirbti pinigus!</i>  <i>Mokėti džiaugtis nugyventu gyvenimu – reiškia gyventi dukart.</i>  <i>Vilkėti svetimais drabužiais – reiškia keisti savo likimą.</i>  <i>Islamas skelbia, kad tarnauti Alachui reiškia melstis, pasninkauti (kurį laiką nevalgyti) ir padėti vargšams.</i></p>
<b>1.3 modelis</b>
<p>Gramatinis lygmuo: [Sub] [REIKŠTI], kad                  Semantinis lygmuo: [ ] [Pred], kad                  Kolokatai: [tai] [REIKŠTI], kad                  Pavyzdžiai: <i>Pašnekovo kojų pirštai pasukti į jus. Tai reiškia, kad kalbėti su jumis yra</i></p>

*išties įdomu.*

*Aš jus atvežiau ir išvežu. Tai reiškia, kad esu jūsų bendrininkas.*

*– Ką reiškia sapnuoti obuolius? – Reiškia, kad artėja Kalėdos...*

*Jis tik prie durų padėdavo šluotą ir tai reiškė, kad nieko nėra namie.*

#### 1.4 modelis

Gramatinis lygmuo: [Sub] [REIKŠTI]||[Pred+REIKŠTI] [Obj\_acc]

Semantinis lygmuo: [ ] [Pred]||[modalinis+V] [ ]

Kolokatai: [tai] [REIKŠTI]||[ +REIKŠTI] [kas, saulė]

Pavyzdžiai: *Geltona spalva reiškia saulę, šviesą ir gerovę, žalia – gamtos grožį, laisvę ir viltį...*

*Ar gali sužinoti, ką reiškia tas žodis?*

*Ar jau sužinojai, ką tai reiškia?*

*Ką reiškia šis žodis? Šis žodis reiškia „to go“.*

*Aukštaičiams brėkšta rytais, o žemaičiams – vakarais, daiktavardis grobai aukštaičiams reiškia „kaulai“, o žemaičiams – „žarnos“.*

*Kai patiriame įtampą, juokas gali reikšti mūsų susinervinimą.*

**1 reikšmės dariniai** reikšmė, išreikšti, pareikšti#<sup>7</sup>, reikštis#

## 2 REIKŠMĖ

#### 2.1 modelis

Gramatinis lygmuo: [Sub] [Obj\_dat] [Adv] [REIKŠTI]||[Pred+REIKŠTI]

Semantinis lygmuo: [ ] [asmuo] [kiekis] [Pred]||[modalinis+V]

Kolokatai: [tai] [ ] [ ] [REIKŠTI]||[ +REIKŠTI]

Pavyzdžiai: *Man tai daug reiškia.*

*Ką Tau reiškia pokalbis?*

*Jeigu turėtumėte visko atsisakyti ir likusį gyvenimą praleisti laive, kas būtų jūsų pakeleiviai? Tie, be kurių jūs negalite gyventi? Sugalvokite savo sąrašą. Ir padarykit viską, kad jie žinotų, kiek daug jums reiškia.*

**2 reikšmės dariniai** (nėra)

Konkrečios semantinės grupės nurodomos ne visuose 1 reikšmės modeliuose, tačiau iš kolokatų matyti, kad subjekto pozicijoje yra ne asmenų pavadinimai, o dažnai tik apibendrinamasis įvardis *tai* (1.3 modelis) (sietinas su ankstesniu sakiniu), arba dėl prijungiamojo sakinio objekto pozicijoje vartojamas įvardis *ką*: *Ar gali sužinoti, ką reiškia tas žodis?* (1.1, 1.4 modelis). 2 reikšmės 2.1 modelyje subjekto pozicijoje vartojamų žodžių konkrečia semantine grupe apibendrinti nebuvo galima, tačiau objekto pozicijoje yra asmenys: nors kolokatų nėra, bet iš pavyzdžių matyti, kad vartojami įvardžiai *man*, *tau*, *jums*.

2 lentelėje pateikto veiksmažodžio *reikšti* atveju modeliai labai skiriasi jau gramatiniame lygmenyje, tačiau daug tokių žodžių, kurių vartosenos modelių gramatinis lygmuo sutampa, o modelių skirtumus lemia semantinis ir

<sup>7</sup> Grotažyme pažymėta, kai žodžiai susiję darybiškai, bet ne semantiškai.

leksinis lygmuo (plg. *skambinti*, žr. 1 lentelę). Sudarę konkretaus žodžio vartosenos santrauką (*WordSketch*), pagal kolokacijas lingvistai matydavo, kokios reikšmės išryškėja. Pavyzdžiui, aprašant daiktavardį *amžius*, pirmą reikšmę „gyvenimo metai“ atskleidė kolokatai *panašaus amžiaus*, *pensinio amžiaus*, *garbaus amžiaus* ir kt., o antrą reikšmę „šimtmetis“ – dažniausiai romėniškais skaitmenimis užfiksuotas laikotarpis arba kolokatai *amžiaus pradžia*, *pabaiga*, *pusė*, *vidurys* ir kt. Panašios reikšmės kolokatai buvo apibendrinami semantinėmis grupėmis: jos rodo leksiką, kuria perteikiamos konkretaus gramatinio modelio sintaksinės kategorijos, pvz., subjektas.

Automatizuota tekstyno analizė paspartino lingvistų darbą, ypač kai reikėjo apdoroti dažnai vartojamų žodžių informaciją, padėjo nustatyti tikėtinus dėsningumus. Tačiau lingvistams prireikdavo ir papildomos konkordanso analizės, nes *WordSketch* duomenyse parodoma ir netiksliai nustatytų ryšių. *SketchGrammar* nustato santykius tarp dviejų žodžių, tačiau tam tikram žodžiui ar jo reikšmei būdingi dėsningumai gali išeiti už dviejų žodžių junginio ribų, todėl lingvistams reikėdavo tai pastebėti ir, analizuojant konkordansą, suformuoti vartosenos modelį iš platesnio konteksto: to dažnai reikėjo, kai nederinamasis pažymynys būdavo sudarytas su daiktavardžio kilmininku ir su būtiniais priklausomais žodžiais, pvz.: [AtrA AMŽIUS\_sg.gen] [Mod]: *vidutinio / garbingo / ... amžiaus asmuo*. Be to, jeigu tam tikras ryšys nenurodytas *SketchGrammar* gramatikos taisyklėse, kai kurie vartosenos faktai galėjo būti neužfiksuoti, pvz., būdvardžio *aiškus* vartosena įterpinio funkcija. Dėl šios priežasties lingvistai peržiūrėdavo tiriamojo žodžio kaitybines formas ir patikrindavo, kas galbūt neužfiksuota analizuojant automatizuotai.

Taikant vartosenos modelių analizę, susidurta ir su įprastais leksikografinio darbo klausimais, pvz., dėl reikšmių skyrimo – tarkim, veiksmažodžio *kreiptis* leksikone pateikti trys modeliai, susieti su reikšme „kreiptis į įstaigą“ (pvz., *Atsitikus nelaimei, kreipkitės į policiją, informuokite savo šalies atstovybę*), ir du modeliai, susieti su reikšme „kreiptis į asmenį“ (pvz., *Į viršininką ar aukštesnį pareigūną kreipkitės pavarde*). DLKŽ abiem atvejais pateikta viena reikšmė. Konkrečiai su vartosenos modelių metodu susijęs probleminis klausimas – semantinių grupių priskyrimas (tai jau pastebėta ir žvalgomajame tyrime, žr. Kovalevskaitė et al., 2019). Siekiant semantinėmis grupėmis apibendrinti žodžio leksinę aplinką, svarbu nustatyti,

kokio apibendrinimo laipsnio reikia, pvz., aprašant dviejų *rašyti* reikšmių modelius, reikia parodyti, kad reikšmės atsiskiria dėl skirtingų objektų: kai *rašyti* vartojamas įvardyti, kad rašomi kūriniai (pvz., eilėraščiai) ir kai *rašyti* vartojamas įvardyti, kad rašomi tekstai laikraštyje, spaudoje.

Dar vienas vartosenos modelių analizės probleminis aspektas – dėl leksinio junglumo skirtumų net ir tai pačiai semantinei grupei priklausantys konkretaus žodžio kolokatai yra skirtingi (Jezek et al. 2010: 8–11): pvz., iš tos pačios semantinės grupės „aktyvus veiksmas“ prie vieno aprašomo žodžio gali būti nurodomi vienokie veiksmažodžiai, prie kito aprašomo žodžio – kitokie. Be to, dėl plačių semantinių veiksmažodžių ir būdvardžių grupių dažnai vienoje grupėje atsiduria gana įvairių reikšmių žodžiai, todėl, pvz., „aktyvus veiksmas“ semantinei grupei priskirtus veiksmažodžius dar galima būtų semantiškai klasifikuoti smulkiau (plg. kolokatus *dovanoti*, *gardinti*, *valgyti*, nurodytus prie šios semantinės grupės *tikti* modelyje 3 lentelėje).

Modelių sudarymas lingvistams kėlė daug iššūkių. Net ir turint pakankamai vartosenos duomenų, nėra lengva tinkamai juos apibendrinti ir pateikti modelių forma. Pirmiausia buvo svarbu kuo tiksliau atpažinti modelius, suderinti, kad juos vienodai traktuotų skirtingi lingvistai (tam buvo taikyta dviejų lingvistų patikra). Kita vertus, tiriant skirtingų žodžių vartoseną lingvistui buvo ne mažiau svarbu išlikti lanksčiam, nes vienam žodžiui aprašyti gali tikti tradicinės kategorijos, kitam – prireikti kitokių kategorijų. Kai kurie priimti sprendimai, kaip sudaryti modelius, buvo tiesiog susitarimo reikalas (pvz., ar skirtingo tipo predikatus pateikti atskiruose modeliuose, kaip žymėti skirtingos raiškos objektus ir aplinkybes ir pan.).

Pasirinkti modelių aprašo principai lemia, koks yra galutinis rezultatas – kaip atrodo vartosenos modeliai, kaip juos interpretuoti ir kokių ribotumų nepamiršti. Kaip minėta anksčiau, žodžio junglumas leksikone suvokiamas plačiau nei valentingumas. Viena vertus, dėl to galime išsamiau užfiksuoti žodžio vartosenos ypatybes, kita vertus, tada kai kurių dažnų veiksmažodžių vartosenos modeliuose pateikiama daug komponentų ir juos interpretuoti gali būti gana sudėtinga. Pavyzdžiui, prie veiksmažodžio *tikti* (*kas TINKA kam*, žr. 3 lentelę) leksikone pateiktame modelyje yra net 8 komponentai:

### 3 lentelė

*Veiksmazodžio tikti vartosenos modelis*

Lygmuo / Reikšmė	Modelio informacija
Gramatinis lygmuo	[Sub] [TIKTI] [Obj_acc] [Obj_dat] [Adv] [Adv] [prie Obj_gen] [ant Adv_gen]
Semantinis lygmuo	[maistas augalas spalva kūno dalis apranga] [Pred][Pred+aktyvus] [daiktas] [asmuo gėrimas maistas] [būdas] [laikas] [maistas  ] [vieta]
Leksinis lygmuo	[vynas, sultys, padažas žiedas, lapas, vaisius spalva plaukas  ] [TIKTI][TIKTI+dovanoti, gardinti, valgyti] [ ] [tu, aš, vaikas, jis, moteris, visas    ] [labai, puikiai, gerai, ypač, tikrai] [visada, dažnai] [prie žuvies, prie valgių, prie patiekalų] [ ]

Kadangi modelyje fiksuojami ir būtinieji, ir nebūtinieji junglumo partneriai, vadinasi, ir pateiktuose pavyzdžiuose (pateikti 47 pavyzdžiai) matysime ne visus modelyje nurodytus komponentus. Tai priklausys nuo to, apie kokių reikšmių žodžius kalbama. Kai subjektas reiškiamas daiktavardžiais iš maisto semantinės grupės (*vynas, sultys, padažas*), tai objekto gramatinė raiška yra su prielinksniu, ir jis vartojamas su žodžiais iš maisto grupės, pvz.: *Šis **padažas** tinka prie paukštienos **patiekalų***. Objektas naudininku vartojamas tada, kai subjektą žymi spalvą ar aprangą pavadinantys daiktavardžiai: *Daugumai moterų tinka mėlyna **spalva**. Tau labai tinka šis **paltas***. Kaip matome, šiuo atveju objekto pozicijoje – asmenys. To paties objekto (reiškiamo naudininku) pozicijoje gali būti ir kitų semantinių grupių žodžiai: gėrimai, maistas, bet tada subjekto pozicijoje bus maistas arba augalai, pvz.: ***Likeriai** ypač tinka įvairiems **kokteiliams**. Džiovinti jų **lapai ir žiedai** tinka **arbatai***. Iš pavyzdžių matyti, kad būdo arba laiko aplinkybė beveik visada vartojama, kai kalbama, kokie drabužiai tinka asmeniui: *Tau **labai** tinka šis **paltas**. **Visur ir visada** tiks balti marškiniai...*; su maistu – priešingai – gali būti pavartota, bet daug rečiau: *Dažniausia giroms gaminti skinamų kiaulpienių žiedai gaivaus ir kiek saldoko skonio, todėl **puikiai** tinka desertinėms salotoms gardinti*. Prielinksninės raiškos aplinkybė [ant Adv\_gen] žymi vietą ir vartojama kalbant apie maistą: *Kepamosios vištienos dešrelės puikiai tinka kepti **ant žarijų***. Objektas galininku arba naudininku vartojamas tada, kai *tikti* su kitais

veiksmožodžiais sudaro sudėtinį predikatą, ir antrasis jo dėmuo reikalauja būtent galininko (*Iš užsienio atvykusiam svečiui **tinka dovanoti dovanas***) arba naudininko (*Malta **vištienai ir kalakutienai gardinti tinka švieži ir džiovinti raudonėliai, petražolės.***). Viena modelyje dažnai nurodomas keliais skirtingais linksniais reiškiamas objektas. Vartotojas tik iš pavyzdžių gali suprasti, su koku veiksmožodžiu koks objekto linksnis bus tinkamas, todėl toks pateikimas reikalauja daugiau analitinių vartotojo pastangų. Aptartas pavyzdys rodo, kad mokomajam žodynui reikėtų ieškoti vartotojui draugiškesnių informacijos pateikimo būdų.

Vartosenos modelių analizės tikslas – aprašyti **tipinę, arba tikėtiną, konkretaus žodžio vartoseną**. Tai, kas vartosenoje tipiška, dažniausiai padeda nustatyti dažnumas, todėl, sudarant modelius, remtasi modelio ir modelio komponento (kalbos dalies, gramatinės kategorijos, kolokato ir pan.) dažnumu. Pasirinkta trijų pavartojimo atvejų minimali riba. Aprašyti tipišką vartoseną galima turint pakankamai duomenų, todėl vartosenos modelių analizės metodu leksikone aprašyti ne visi antraštyno vienetai, o tik tie, kurių dažnumas tekstyno rašytinėje dalyje buvo 100 ar daugiau pavartojimo atvejų. Atitinkamai leksikone pateikti dalies žodžių vadinamieji ilgieji aprašai, t. y. su vartosenos modeliais, ir trumpieji aprašai – su pavyzdžiais, bet be vartosenos modelių. Parengti 207 veiksmožodžių, 386 daiktavardžių, 87 būdvardžių ir 41rieveiksmio, t. y. kiek daugiau nei 700 žodžių (tai sudaro 20 proc. iš visų 3 700 antraštinių vienetų), vartosenos modelių aprašai.

## **Leksikono ypatybės**

### **Leksikonas – tekstyno pagrindu parengtas leksikografinis išteklius.**

Tekstynai plačiai naudojami leksikografiniuose tyrimuose, nes suteikia daugiau duomenų apie realią vartoseną, o tai padeda nustatyti tokius kalbos dėsningumus, kurie anksčiau visai nepatekdavo į lingvisto akiratį, arba juos buvo sunku pastebėti (plačiau Marcinkevičienė, 2000, p. 45–47). Tačiau taip pat žinoma, kad tekstyne galima rasti tik tai, kas jame yra. Tas pats pasakytina ir apie šiame leksikografiniame projekte naudotą *Mokomąjį tekstyną*.

Leksikonas sudarytas remiantis šio tekstyno rašytinės kalbos pateiktyniu, kurį sudaro dvejopi tekstai: 1) tekstai, rinkti iš vadovėlių, skirtų

lietuvių kalbos besimokantiems kitakalbiamis (jie sudaro 17 proc. ir vadinami vadovėliniais tekstais), ir 2) tekstai, rinkti iš populiariamųjų ir grožinių knygų, naujienų portalų, viešųjų užrašų, instrukcijų, skelbimų, dokumentų ir kt. (jie sudaro 83 proc. ir vadinami nevadovėliniais tekstais). Vadovėliniai ir nevadovėliniai tekstai suklasifikuoti į 29 žanrus (dialogus, pasakojimus, informacinius tekstus ir kt.) ir atitinka keturias grupes pagal komunikacinius tikslus: informacinius, pažintinius, apeliacinius ir meninius. Svarbu pasakyti, kad jų tekstų skaičius yra nevienodas, nors reprezentuojama nemažai skirtingų žanrų: pavyzdžiui, labai daug informacinių tekstų, o apeliacinių – mažai (plačiau apie tekstyno sandarą žr. Kovalevskaitė et al., 2020b). Nors leksikono medžiaga atspindi rašytinės kalbos ypatybes, dialoguose ir subtitruose vartojama kalba atskleidė ir sakybinei vartosenai būdingas, bet leksikografiniuose ištekliuose dar nefiksuotas ypatybes: pvz., daugiafunkcę tam tikrų dažnų veiksmažodžių vartoseną liepiamąja nuosaka *palauk(ite)*, *klausyk*, *žiūrėk* (plg. Jasionytė-Mikučionienė, 2016). Reikia pripažinti, kad jeigu būtų analizuota ir sakinė tekstyno dalis<sup>8</sup>, vartotojams taip pat būtų pateikta svarbių ir mažai aprašytų lietuvių kalbos duomenų: kaip ir kitų kalbų, lietuvių kalbos sakinė kalba skiriasi nuo rašytinės (plg. atributų vartoseną, aprašytą Kamandulytė-Merfeldienė et al., 2016).

Dalis tekstyno medžiagos yra kiek supaprastinta, ne visai natūrali kalba (pvz., diktantai ar pokalbiai iš vadovėlių); kai kurie tekstai galbūt ne tokie aktualūs (pvz., jau gana senoki pažinčių skelbimai) arba pernelyg skiriasi nuo rašytinės kalbos (tarkim, subtitrai, kurių vertimas būna ne itin kokybiškas arba vartojama specifinė leksika). Tačiau atrenkant pavyzdžius, manytume, tos neaktualios dalies pavyko išvengti: atsisakyta vadovėlių tekstuose kartais pasitaikančių nenatūralių sakinių (pvz., *Petriukas eina į mokyklą*), sakinių su pasenusiomis realijomis, kurių nebūtina žinoti, nes tai nėra dabar aktuali kultūrinė istorinė informacija (pvz., *spragilas*, *juodosios beretės*). Negalime teigti, kad tekstynas visiškai atspindi pačią naujausią kalbos vartoseną (seniausias tekstas publikuotas 1990 m.), tačiau dauguma tekstų yra iš 2000–2020 metų laikotarpio.

Dažnu atveju, palyginę, tarkim, to paties žodžio aprašą *Dabartinės*

---

<sup>8</sup> Šia tekstyno dalimi nepasinaudota dėl to, kad, rengiant leksikoną, tekstai sakinės kalbos patekstytiui dar buvo tvarkomi.

---



*lietuvių kalbos žodyne* (toliau – DLKŽ) ir leksikone, matydavom, kad leksikone nustatyti modeliai ir jų vartosenos pavyzdžiai neatspindi DLKŽ, nors yra dažni tokia palyginti mažame tekстыne kaip mūsų naudotasis. Pavyzdžiui, analizuojant žodžio *amžius* vartosena, išryškėjo gana dažnas modelis, kai *amžius* būtinai vartojamas su derinamuoju pažyminiu ir kartu su juo sudaro dvinarį nederinamąjį pažyminį, pvz., plg. leksikos lygmenį viename iš *amžius* vartosenos modelyje: *įvairaus, senyvo, panašaus, garbaus, garbingo, vyresnio, mokyklinio, pensinio, vidutinio, darbingo, x metų AMŽIAUS vaikas, žmogus, draugas, našlė, vyras, vyriškis, moteris, asmuo, grupė*. Sudėtinis atributas [*įvairaus, senyvo, panašaus...* AMŽIUS\_sg.gen] dažnai eina vardine sudėtinio tarinio dalimi: **Esu 30 metų amžiaus...**; *Svečių susodinimo taisyklių galima nesilaikyti, jei svečių pora yra itin garbaus amžiaus*. Leksikone užfiksuotas ir modelis [AMŽIUS] [Pred], kuriame *amžius* atlieka subjekto funkciją, o predikatas išreiškiamas būseną (*būti*) arba aktyvų veiksmą rodančiais veiksmažodžiais. Pavyzdžiui: *Pabendrauti ir smagiai praleisti laiką nori visi, amžius čia neturi reikšmės; ...palankiausias amžius mokytis skaityti yra nuo trejų iki šešerių metų...; Palyginti su 2001 m., vidutinis gyventojų amžius padidėjo 4 metais, dabar vidutinis amžius – 41 metai*.

Paminėto daiktavardžio *amžius* vartosenos ypatybės neatspindėtos DLKŽ, kur galima rasti tokių pavyzdžių: *Kiek jūs amžiaus turite? Visam amžiui užteks. Sulaukė garbingo amžiaus. Pagal savo amžių jis dar gana žvalus*, nors tokių atvejų neužfiksuota naudotame tekстыne. Kai klausama apie amžių, dabartinėje kalboje paprastai vartojama *Kokio Jūs amžiaus?* (o neformaliai: *Kiek tau metų?*). Būtent tokie pavyzdžiai pateikti *Mokomajame lietuvių kalbos žodyne*, kuriame randame dar vieną pavyzdį prie pirmosios reikšmės „gyvenimo metai“: *Jis mano amžiaus*. Leksikono apraše tokios vartosenos neužfiksuota, bet prie *amžius* (dažnumas – 456 pavartojimo atvejai) pateikti šeši vartosenos modeliai, kuriuose išryškėja labai įvairios šio daiktavardžio sintaksinės funkcijos: nederinamojo atributo, subjekto, objekto, kelių aplinkybių. *Mokomajame žodyne* tokia vartosena neaprašyta.

Kaip jau minėta skyriuje „Leksikono sudarymo principai ir metodika“, vartosenos dėsningumams nustatyti svarbu tam tikro elemento pasikartojimas, dažnumas, o tekstynas nėra didelis. Vis dėlto po tyrimo galima pasakyti, kad pagrindiniai žodžio vartosenos dėsningumai gali atsiskleisti net ir tokia

mažame tekстыne. Žinoma, mūsų naudotame tekстыne patikimiausia spręsti apie dažnai pavartotų žodžių ypatybes. Rečiau vartojamiems žodžiams dažnumo kriterijų galime taikyti tik su išlygomis, nes dėl mažos tekstyno apimties retai gali pasirodyti ir tipiški, ir ne tokie tipiški vartosenos modeliai. Rečiau vartojamų žodžių dėsnų tendencijas galime tik užčiuopti, todėl, kaip jau minėta, vartosenos modelių analizė atlikta tik gana dažnai (ne mažiau kaip 100 kartų) vartojamų žodžių. Kai kurie vartosenos atvejai buvo per reti ir dėl to leksikone neparodyti kaip tipinė tam tikro žodžio vartoseną, nors, įvertinus vartojimo dažnumą didesniame tekстыne, veikiausiai būtų pagrindo priskirti juos tipinei tam tikro žodžio vartosenai nurodant:

a) atskirą reikšmę: pvz., leksikone *pristatyti* nurodyta tik reikšmė „supažindinti, informuoti, skelbti“, nes buvo vos vienas ar du atvejai, kai pavartota reikšmė „atvežti“: *ten pas šeimininką yra picerijų brošiūrų ir nemokamai pristato picas*;

arba

b) dar vieną vartosenos modelį, pvz., modelis *galėti + įsivaizduoti* nebuvo suformuotas, nes jo dažnis tekстыne buvo mažiau nei minimalus (t. y. trys pavartojimo atvejai). Nors, pavyzdžiui, patikrinus šį atvejį *Dabartinės lietuvių kalbos tekстыne* (toliau – DLKT), matyti, kad *galima / gali / galite / galiu / galėjau įsivaizduoti* yra dažnas vartosenos modelis, ir, pridėję trūkstumus komponentus (greičiausiai – šalutinio papildinio dėmens jungtuką *kad*), turėtume dar vieną žodžio *įsivaizduoti* modelį. Vis dėlto jeigu konteksto elementas (reikšmė, sintaksinė konstrukcija, konkretus žodis) buvo pavartotas rečiau kaip tris kartus, jo nelaikėme tipine vartoseną.

Vartosenos modelių analizės tikslas – aprašyti tipinę žodžio vartoseną, todėl kai kurie reti vartosenos faktai buvo atmesti dėl to, kad rodė ne vartosenos normas, o tik galimybes (angl. *exploitations*, žr. Hanks, 2004). Pavyzdžiui, tiriant suformuoto vartosenos modelio semantines grupes, gali išryškėti tokių retų atvejų, kai, pvz., modelyje [Sub] [TAPTI+N\_ins] tekстыne radome

pavyzdžių, kai subjektas buvo išreikštas žodžiais, reiškiančiais negyvus objektus, bet tokie pavyzdžiai priskirti pavieniams atvejams: **Išmaniosios technologijos ir televizija** [Sub] **tampa** iki trečdalis Lietuvos vaikų **auklėmis** vasarą. Dažnai **televizorius** [Sub] **tampa** tikru **šeimos nariu**. Šio modelio tipiniai subjektai ir vardažodinio predikato dėmenys yra asmenys, pvz.: *ateityje jis taps puikiu specialistu*.

Leksikone kiekvieno žodžio apraše pateikiamos ir kaitybinės formos, kuriomis šis žodis pavartotas tekstyne. Prie skyriuje „Leksikono sudarymo principai ir metodika“ aptarto veiksmažodžio *reikšti* pateikiamos tokios kaitybinės formos<sup>9</sup>:

**rėiškia** VERB.FIN.PRS.3.IND (158), **rėiškė** VERB.FIN.PST.DFN.3.IND (10), **rėiškiantis** VERB.PTCP.PRS.SG.M.~ACT.NOM (5), **rėikšti** VERB.INF (4), **rėikštų** VERB.FIN.3.COND (2), **rėiškiančių** VERB.PTCP.PRS.PL.M.~ACT.GEN (2), **rėiškusi** VERB.PTCP.PST.DFN.SG.F.~ACT.NOM (1), **rėiškęs** VERB.PTCP.PST.DFN.SG.M.~ACT.NOM (1), **reikšdamà** VERB.~PSD.SG.F (1), **reikš** VERB.FIN.FUT.3.IND (1), **rėiškiančio** VERB.PTCP.PRS.SG.M.~ACT.GEN (1), **reikšdamì** VERB.~PSD.PL.M (1).

Iš karto matyti, kuri viena ar kelios kaitybinės formos dominuoja. Tai – svarbi informacija pirmiausia lingvistams: jau minėta, kad, aprašant vartosenos modelius, vienas iš žingsnių buvo peržiūrėti kaitybines žodžio formas. Svarbios formos atsispindi ir modeliuose (nes būtent dėl jų tam tikri modeliai ir susiformuoja), paprastai tai yra dažniausios kaitybinės formos. Nors lietuvių kalbai kaityba labai svarbi, tai nereiškia, kad reikia mokytis visų tam tikro žodžio formų. Todėl informacija apie dažniausias kaitybines formas labai naudinga ir negimtakalbiams: juk jam taip pat aktualu mokėti tas, kurios yra svarbiausios konkrečiam žodžiui, nes yra tipinės jo vartosenos dalis. Turint tekstyno duomenis galima pateikti šią svarbią morfologinę informaciją.

---

<sup>9</sup> Tekstynas morfologiškai anotuotas automatiškai ir nors automatinio anotavimo tikslumas tikrai geras (Kapočiūtė-Dzikienė et al., 2017), vis dėlto gali būti likusių anotavimo klaidų, pvz., kai sutampa bendratis ir neveikiamosios rūšies būtojo laiko vyriškosios giminės dalyvio daugiskaitos vardininkas; moteriškosios giminės vardažodžių vienaskaitos kilmininkas ir daugiskaitos vardininkas ir kt. Morfologinė informacija pateikta pagal Leipcigo glosavimo taisyklės (<https://kalbu.vdu.lt/wp-content/uploads/2022/02/Morfologines-pazymos.pdf>), skliaustuose nurodytas dažnumas, t. y. kiek tam tikros formos pavartojimo atvejų yra tekstyne; kaitybinės formos sukirčiuotos.

Rengiant žodyną ar leksinę bazę tekstyno pagrindu, antraštynas formuojamas remiantis dažniniais žodžių sąrašais (apie juos rašoma toliau). Tačiau analizuojant atsirinktų žodžių vartoseną, konkordanse dažnai išryškėja papildoma informacija, dėl kurios galima ne tik parengti tikslesnį analizuojamo žodžio aprašą, bet ir nustatyti daugiau tokių žodžių ar junginių, kuriais vėliau papildomas antraštynas. Rengiant leksikoną, analizė padėjo tekстыne nustatyti tokius vartosenos faktus, kurie neišryškėjo iš paprasto dažninio žodžių sąrašo:

- a) į kokius pastoviuosius žodžių junginius (frazologizmus, posakius, sudėtinius pavadinimus) įeina analizuojamas žodis, plg.: *akis* įeina į šiuos junginius: *badyti akis, akis už akį, akies vokas*;
- b) kokie žodžiai turi būti pateikiami keliais homonimais, pvz., veiksmažodis *derėti* buvo labai dažnas, bet dažniau vartotas kaip homonimas reikšme „tikti“, o ne kaip homonimas reikšme „duoti vaisių“. Pasitaikė nemažai atvejų, kai leksikone homonimams priskirti žodžiai nėra laikomi homonimais kituose aiškinamuosiuose žodynuose (*Lietuvių kalbos žodyne*, DLKŽ, BLKŽ), pvz., *drausti, rengti(s), tekėti*. Daug klausimų dėl homonimijos kėlė veiksmažodžių vediniai, pvz., *pašokti, nustatyti, išdarinėti, išgyventi* ir kt. DLKŽ kai kurių vedinių nėra, tačiau neiškėlus vedinių į antraštiną matyti ne visa homonimijos įvairovė.

Atrenkant antraštinius žodžius pagal dažninį sąrašą pasitaikė daug atvejų, kai automatiškai nustatyta lema paslėpdavo tikrąjį leksinį vienetą: pvz., a) kai kurie subūdvardėję ar sudaiktavardėję dalyviai buvo randami tik tarp atitinkamų veiksmažodžių formų (pvz., *įtariamasis, suaugęs*); b) suprieveiksmėję daiktavardžių ir būdvardžių linksniai buvo nustatyti tik analizuojant konkordansą (pvz., *laikas* ir *laiku, trumpas* ir *trumpam*). Žinoma, atsižvelgiant į mūsų antraštiną, tokių atvejų nebuvo itin daug, bet jų įvairovė atskleidė, kad, rengiant lietuvių kalbos leksikografinį aprašą, labai svarbu nesusikoncentruoti tik į lemą, kuri kartais gali būti net netinkamai nustatyta (plg. *duomu* ir *duomenys* – automatiškai buvo nustatyta vienaskaitinė lema, nors tekстыne visais atvejais vartojamos daugiskaitos formos).

**Leksikono antraštynas yra sąlygotas tekstyno ir pasirinktos atrankos metodikos.** Leksikone siekta pateikti besimokantiesiems aktualią leksiką. *Lingvodidaktikos terminų žodyne* (Ramonienė at al., 2012), aprašant pagrindinį žodyną (angl. *basic vocabulary*), nurodoma, kad tai yra dažniausiai vartojamų ir būtinausių tikslinės kalbos žodžių visuma, leksinis minimumas; svetimosis kalbos vadovėlių ir mokymo priemonių leksinis pagrindas – žodžiai, skirti aktyviajam išmokimui. Vadinasi, viena svarbių aktualios leksikos ypatybių yra jos dažnumas tikslinėje kalboje. Iš lietuvių kalbos išteklių turime DLKT pagrindu parengtą dažninį žodžių (formų) sąrašą<sup>10</sup>, tačiau, kaip rodo kitų kalbų, pvz., anglų kalbos, tyrimai, siekiant nustatyti bendrosios leksikos sluoksnį, svarbu atsižvelgti ne tik į tam tikrų leksinių vienetų dažnumą, bet ir į dispersiją (Brezina et al., 2015). Taigi sudarant leksikoną minėtasis dažninis žodžių (formų) sąrašas nebuvo panaudotas.

Iš dalies pagrindiniam žodynui priskirtinos leksikos galima rasti lietuvių kalbos besimokantiesiems skirtuose vadovėliuose, metodinėse priemonėse (pvz., Narbutas et al., 2002; Pribušauskaitė et al., 2000; Ramonienė et al., 2006): juose pateikiamos tam tikram kalbos lygiui aktualios temos, su jomis susiję žodžiai ir žodžių junginiai. Tačiau, kiek mums žinoma, vieno pagrindinį žodyną atspindinčio lietuvių kalbos leksikos sąrašo, kuriuo būtume galėję remtis, nėra, todėl **aktualią leksiką** atsirinkome iš *Mokomojo tekstyno*.

Leksikono antraštyno formavimo metodika yra susieta su naudotu tekstynu, todėl ją nulėmė tekstyno sandara ir apimtis. Kaip parodė tekstyno analizė, tekstų klasifikacija į A1, A2, B1 ir B2 kalbos lygius nėra visiškai tiksli: vadovėlinių tekstų kalbos lygis dažniausiai buvo aiškus, o nevadovėliniai tekstai buvo automatiškai suklasifikuoti panaudojus mašininio mokymo modelius, tačiau šių tekstų kalbos lygis nustatytas tik 60 proc. tikslumu (Grigonytė et al., 2018; Boizou et al., 2020). Todėl nutarta antraštyną suformuoti tokiu būdu, kuris padėtų sumažinti atskirų lygių klasifikavimo netolygumus. Dėl šios priežasties, deja, nebuvo galima antraštyno leksikos suskaidyti pagal lygius, kaip tai padaryta kitų kalbų mokymo ištekliuose (François et al., 2014), pateikti informacijos apie leksinio vieneto dažnumą kiekviename lygyje.

---

<sup>10</sup> <https://clarin.vdu.lt/xmlui/handle/20.500.11821/8>

Leksikono antraštynas suformuotas dviem etapais. Per pirmą etapą atskyrėme **bendrąją leksiką**, kurią sudarė dvi grupės:

- 1) visiems lygiams bendra leksika<sup>11</sup>;
- 2) bendra aukštesnių lygių (A2–B2 ir B1–B2) leksika.

Iš viso bendrosios leksikos sąrašą sudarė apie 7 700 lemu, t. y. 20 proc. visų tekstyno lemu (čia įėjo ir labai dažni, ir retesni, ir rečiau (bent 2 kartus pavartoti) žodžiai).

Antrame leksikono sudarymo etape iš 7 700 lemu atsirinkome tuos žodžius, kurių dažnumas yra 100 ar daugiau kartų (apie 770 žodžių, kurie atėjo iš visiems lygiams bendros arba aukštesniuose lygiuose (A2–B2) pavartotos leksikos). Ši 770 žodžių grupė laikyta **leksikono antraštyno branduoliu, baziniais žodžiais**. Antraštyno branduolio žodžiai buvo ištirti tekстыne taikant vartosenos modelių analizę. Parengti vadinamieji ilgieji šių žodžių aprašai.

100 ar daugiau kartų dažnumo riba pasirinkta todėl, kad vartosenos modeliams nustatyti svarbus reprezentatyvus duomenų kiekis. Reikia paminėti, kad nebuvo aprašyti kai kurie pagalbiniai ir pusiau pagalbiniai (modaliniai) veiksmažodžiai *būti, galėti, reikėti*, nors ir atitiko dažnumo kriterijų. Šių veiksmažodžių vartoseną atspindi kitų žodžių modeliuose, kuriuose parodomi sudėtiniai tariniai. Dalis pagalbinių ir pusiau pagalbinių veiksmažodžių (pvz., faziniai veiksmažodžiai) yra aprašyti: *darytis, likti, tapti ir pradėti, imti, baigti, liautis*. Jų modeliuose aiškiai atsiskiria reikšmės ir nuo jų priklausantys skirtingi modeliai, kurie rodo, kokia leksinė ir gramatinė aplinka būdinga savarankiškai vartojamam veiksmažodžiui.

Aprašant bazinius žodžius, buvo nurodomi su jais ar jų atskiriomis reikšmėmis susiję dariniai ir pamatiniai žodžiai, taip pat pastovieji žodžių junginiai. Su baziniais žodžiais susiję leksiniai vienetai (pamatiniai žodžiai, dariniai, žodžių junginiai) buvo renkami iš bendrosios leksikos, iš viso su baziniais žodžiais susieta apie 2 300 žodžių ir apie 700 žodžių junginių. Buvo sudaromi su leksikono baziniais žodžiais susijusių kitų žodžių ir pastoviųjų

---

<sup>11</sup> Kai kurie bendros leksikos žodžiai nebuvo atmesti, jeigu nebuvo paminėti kuriame nors lygyje (pvz., A2 arba B1), nes naudotas tekstynas yra gana mažas ir jo dalys pagal kalbos lygius nėra vienodos.

---

žodžių junginių trumpieji aprašai.

Aprašant susijusius žodžius, netaikyta dažnumo riba: siekiant parodyti darybinius ryšius tarp žodžių, buvo analizuojami ir dažni, ir reti bendrosios leksikos žodžiai. Šiuos retesnius žodžius laikome mokiniui aktualia leksika todėl, kad jie parodo leksinius arba darybinius ryšius: pavyzdžiui, žodis *augimvietė* nėra dažnas tekstyne, bet į leksikono antraštiną įtrauktas, nes atskleidžia darybinį ryšį su bazinės leksikos žodžiu *vieta*. Kaip nurodoma metodinėse priemonėse (pvz., Ramonienė et al., 2006), darybos mokymasis yra svarbi aukštesnių kalbos lygių užduotis.

Kitų tekstynais paremtų mokomųjų žodynų rengėjai (plg. Kallas et al., 2014) antraštynui formuoti naudojo didesnę tekstyną: estų mokomajam žodynui naudotas bendrasis 250 mln. žodžių tekstynas *Estonian Reference Corpus*, iš jo gautas 5 000 leksinių vienetų antraštynas. Mes rėmėmės tik specialiuoju, ne bendruoju, tekstynu, be to, gana nedideliu. Dėl šios priežasties vartotojai leksikone gali nerasti gana įprastų dažnų žodžių, kurie pateikti mokomuosiuose žodynuose.

Lietuvių kalbos kaip svetimšios mokymo priemonėse nurodytos temos buvo orientyras, kokius nevadovėlinius tekstus atrinkti į tekstyną (plačiau apie atrankos principus žr. Kovalevskaitė et al., 2020b), todėl dalis besimokantiems aktualios leksikos iš tekstyno perėjo ir į leksikono antraštiną. Tačiau dėl pasirinktos leksikono antraštyno formavimo metodikos dalis aktualios leksikos, nors ir randama tekstyne, neatsispindi leksikono antraštyne: pavyzdžiui, *Praktinėje gramatikoje* (Ramonienė et al., 2003) tarp dažniausių veiksmažodžių paminėtas žodis *ragauti* tekstyne pavartotas 25 kartus (kartu su *paragauti* ir *neragauti* įeina ir į bendrosios leksikos sąrašą), bet leksikone neaprašytas, nes darybiškai ar semantiškai nesusijęs su jokių baziniu žodžiu, o jo dažnumas per mažas, kad būtų aprašytas su ilguoju aprašu. Tiesa, *ragauti* vartojimo pavyzdžių galima rasti kitų antraštinių žodžių *stebuklingas*, *ledai* aprašuose, pvz.: *Pavaikštinėsite gražiausio Lietuvos kurorto gatvelėmis, ragausite **stebuklingo** vandens... Būdamas Lietuvoje jis pirmą kartą ragavo alaus **ledų**, valgė labai skaniai paruoštą silkę Trakuose, o šakotį pavadino ypatingu kepinu...*

Kai kurie retesni žodžiai ar žodžių junginiai laikomi aktualiais dėl to, kad atspindi istorinę kultūrinę informaciją, pvz., *valsčius*, *piliakalnis*, *sovietinis*,

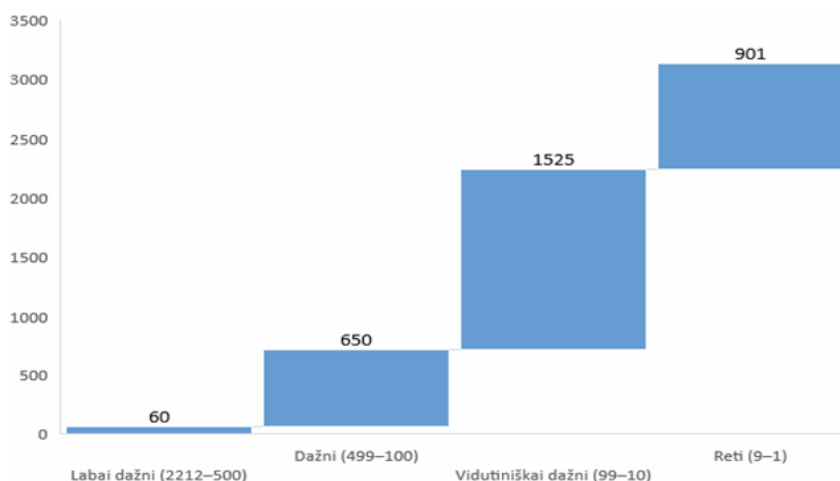
*aguonų pienas, Baltijos kelias, didysis kunigaikštis*. Tokie vienažodžiai ir keliažodžiai leksiniai vienetai yra arba istorizmai, arba kultūrinių realijų pavadinimai ir įtraukti į leksikono antraštiną, nes siejasi su kitais aprašytais žodžiais. Tekstyne yra tekstų iš vadovėlių apie Lietuvos istoriją, kultūrą, gamtą ir pan., todėl nemažai šios leksikos atsispindėjo ir bendrosios leksikos sąrašuose.

Atsirenkant bendrąją tekstyno leksiką, tikrinių žodžių leksikone atsisakyta (todėl nėra dažnų tikrinių daiktavardžių, pvz., *Lietuva, Vilnius, Kaunas, Nemunas*), į antraštiną vėliau pateko tik tokie atvejai (dažniausiai sudėtiniai pavadinimai), kurie susieti su aprašytais žodžiais, pvz., *Kaliningrado sritis* ar *Viduržemio jūra*. Kita vertus, tokius vardus, kaip *Lietuva, Nemunas*, mokiniai paprastai sužino pačioje mokymosi pradžioje (A1 lygyje), o leksikono tikslinė grupė yra A2 ir aukštesnių kalbos lygių mokiniai.

Visą leksikono antraštiną sudaro apie 3 700 vienetų (apie 3 000 vienažodžių ir apie 700 pastoviųjų žodžių junginių). Apie 55 proc. vienažodžių antraštyno vienetų atspindi visiems kalbos lygiams bendrą leksiką, apie 45 proc. – aukštesnių kalbos lygių leksiką (A2–B2, B1 ir B2 lygiams bendrą leksiką).

## 2 paveikslas

*Leksikono antraštinių žodžių skaičius pagal dažnumą*





Leksikone pateiktų žodžių dažnumas tekstyne matyti 2 paveiksle<sup>12</sup>. Paveikslo horizontaliojoje ašyje pažymėtos dažnumo grupės su nurodytais intervalais (vertikaliojoje ašyje nurodytas žodžių skaičius). Patys dažniausi (daugiau nei 2 000 pavartojimo atvejų) žodžiai yra, pavyzdžiui, *žmogus, metai, daug*. Tokių labai dažnų žodžių leksikone gana nedaug – 60, o didžiąją leksikono antraštyno dalį sudaro žodžiai, priskirti į dažnų ir vidutiniškai dažnų žodžių grupes – iš viso 2 175 žodžiai. Labai dažni ir dažni žodžiai leksikone pateikti su ilgaisiais aprašais, o vidutiniškai dažni ir reti žodžiai – su trumpaisiais aprašais. Retų žodžių grupę daugiausia sudaro dariniai. Jie pateikti tam, kad būtų parodyti žodžių ryšiai. Kadangi naudotas tekstynas gana mažas, šiame tekstyne ir reti žodžiai gali būti dažniausios leksikos, arba pagrindinio lietuvių kalbos žodyno, dalis. Bet tam nustatyti, žinoma, reikia atskiro tyrimo.

**Leksikone nėra reikšmių apibrėžčių – informaciją apie žodžio reikšmes siekiama parodyti vartosenos modeliais ir pavyzdžiais iš tekstyno.** Iš vartosenos modelių ir pavyzdžių prie kiekvieno modelio galima susidaryti vaizdą, kas būdinga aprašomo žodžio vartosenai. Kaip parodyta skyriuje „Leksikono sudarymo principai ir metodika“, reikšmių konteksto skirtumai gali išryškėti jau gramatiniame lygmenyje (plg. *reikšti* skirtingų reikšmių vartosenos modelius). Tačiau dažniausiai tam, kad būtų galima suprasti, kaip vartojamas žodis viena ar kita reikšme, svarbūs ir kiti lygmenys – t. y. semantinės grupės ir kolokatai (plg. žodį *skambinti*).

Mokomojoje leksikografijoje vartosenos dėsningumą pateikimas modeliais gali parodyti, kad ta pačia kalbos dalimi laikomi žodžiai pasižymi labai nevienoda vartoseną, pvz., tarp būdvardžių *pilnas, vertas* vartosenos modelių bus visai nebūdingų tokiems būdvardžiams, kaip *senas, naujas*, o, pavyzdžiui, modeliai, kur būdvardis vartojamas bevarde gimine kaip sudurtinio tarinio vardinė dalis, būdingi dviem būdvardžių semantinėms grupėms: fiziniams (*karštas, šaltas, pilnas, gausus*) ir vertinamiesiems (*malonus, ramus, svarbus, lengvas*) (Bielinskienė et al., 2021).

Reikia pripažinti, kad dalis modelių gali būti per sudėtingi dėl juos sudarančių komponentų skaičiaus, tačiau prie kiekvieno modelio pateikiama pavyzdžių, iliustruojančių modelyje aprašomus dėsningumus. Tiesa, pavyzdžių

---

<sup>12</sup> Leksikone yra pastovieji žodžių junginiai, bet prie jų dažnumas nenurodomas, todėl jie nepateikti 2 paveiksle.

skaičius priklauso nuo konkretaus žodžio vartosenos įvairovės, nes pavyzdžiai atrinkti pagal modeliuose fiksuojamą informaciją, kad juose atsispindėtų ir gramatinės, ir semantinės tam tikro modelio ypatybės. Taigi pavyzdžių skaičius priklauso nuo komponentų, kurie užfiksuoti modelyje. Jeigu aprašomas retesnis pastovusis žodžių junginys ir nėra gero pavyzdžio tekstyne, tada pavyzdžiai atrinkti iš DLKT.

Informacija apie pavyzdžių skaičių<sup>13</sup> pateikta 4 lentelėje. Žinoma, daugiausia pavyzdžių (nuo 10 iki 163) teikiama dažniausiems, baziniams žodžiams, aprašytais ilguoju aprašu su modeliais, pvz.: *metai* (163), *laikas* (134), *darbas* (100), *padaryti* (100), *kalbėti* (93), *žmogus* (91), *senas* (91), *diena* (78), *vieta* (77), *mėgti* (76), *laukti* (75), *kalba* (75), *mažas* (73), *eiti* (72). Mažesnis pavyzdžių skaičius (nuo 9 iki 1) pateikiamas vidutiniškai dažniems ir retiems žodžiams (žr. 2 pav.), pastoviesiems žodžių junginiams. Iš viso leksikone pateikta 28 550 pavyzdžių.

#### 4 lentelė

*Pavyzdžių skaičiaus pasiskirstymas*

Pavyzdžių skaičius	163–51	50–10	9–5	4–1
Žodžių skaičius	60	633	424	2 572

Atrinkdami pavyzdžius konkrečiam vartosenos modeliui, taikėme kelis **gerų pavyzdžių atrankos principus**. Pagal tipišumo principą atrinkdavome tokius pavyzdžius, kuriuose atsispindėtų vartosenos modeliuose užfiksuoti elementai: gramatinės formos, semantinės grupės, kolokatai ir pan. Tipišumui užtikrinti svarbu parinkti ir tipiškas (dažnas) morfologines formas, kuriomis žodis realizuojamas konkrečiame modelyje.

Aiškumas ir informatyvumas yra papildomi gerų pavyzdžių atrankos kriterijai: kadangi naudojome tekstyną, kurio didžioji dalis tekstų – nevadovėliniai, sakiniuose galėjo pasitaikyti nesuprantamų žodžių. Taikydami aiškumo kriterijų ir atrinkdami pavyzdžius, vengėme sakinių, kuriuose yra retų, stilistiškai žymėtų (*tapšnoti, nunovijo, galabijo, iš padilbų, šnairuosite,*

---

<sup>13</sup> Čia įskaičiuoti ir pasikartojantys pavyzdžiai, t. y. tokie atvejai, kai tas pats pavyzdys pateikiamas prie skirtingų modelių.

*briauna...*), perkeltinės reikšmės (pvz., *Neformalaus ugdymo centras atgyja rudeni*) žodžių, kokios nors specialiosios srities terminų.

Kad modelio pavyzdžiai atitiktų informatyvumo kriterijų, svarbu parinkti pavyzdžius, suprantamus be platesnio konteksto, tarkim, jeigu sakinyje vartojami anaforiniai vienetai, vartotojui gali pritrūkti informacijos, plg.: *Jos labai praverčia, kai svečių yra daugiau negu 10* (neaišku, ką žymi jos). Tokiais atvejais į pavyzdį būdavo įterpiama ankstesniame kontekste paminėta informacija.

Šiame leksikone remtasi Philippe'o Humble'o (2001) ir Anos Frankenberg-Garcia'os (2015) darbuose aprašoma dviejų tipų pavyzdžių samprata: pavyzdžiai supratimui (angl. *decoding examples*) ir pavyzdžiai produkavimui (angl. *encoding examples*). Leksikone siekiama iširti atrinktų žodžių vartosenos dėsningumus tekstyne ir pateikti reprezentatyvių produkavimui skirtų pavyzdžių. Tiesa, tarp leksikone pateikiamų pavyzdžių yra ir skirtų supratimui; tokio tipo pavyzdžiai dažniausiai pasirinkti tada, kai reikėjo iliustruoti perkeltinės reikšmės junginius, pvz.: *...rasti su Miesto švente susijusį suvenyrą buvo tas pats, kas ieškoti adatos šieno kupetoje; Daug perskaitytų knygų išugdo šeštą jausmą – nuojautą, padedančią atskleisti žodžių prasmę*. Pasak Frankenberg-Garcia'os (2015, p. 295), įvairių supratimui skirtų pavyzdžių nauda priylgsta geroms apibrėžtims.

Prie pavyzdžių nėra žymima, iš kokių tekstų jie paimti. Jeigu tekstai neutralūs, tai nėra reikalinga, bet pateikus ištraukas, tarkim, iš humoristinių tekstų be metainformacijos vartotojui gali būti neaišku. Pavyzdžiui, prie vieno iš *patarti* modelio, kur šis veiksmažodis vartojamas forma *patartina*, *patariama*, vartotojas kartu su tokiais pavyzdžiais, kaip *Rytą patartina tepti odą drėkinamuoju kremu*, ras ir štai tokį: *Važiuojant 150 km per valandą greičiu patariama giedoti „Dieve, arčiau Tavęs“*. Tai yra humoristinio pobūdžio tekstas, bet prie jo kol kas nėra pateikiama paaiškinimų dėl žanro ar stiliaus. Šią informaciją galima rasti tik atliekant paiešką *Mokomajame tekstyne*.

Taikant modelių metodą, buvo svarbu pasirinkti tam tikrą vartosenos modelių užrašymo būdą. Kaip jau minėta dalyje „Leksikono sudarymo principai ir metodika“, kai kurie susitarimai buvo labai sąlygiški. Tarkim, buvo galima kiekvienam objektui rašyti po atskirą modelį, bet pasirinkome pateikti vieną objekcinį modelį dėl glaustesnės informacijos. Kita vertus, tada viename

modelyje dažnai nurodomas keliais skirtingais linksniais reiškiamas objektas, o skirtingą linksnį lemia skirtingi veiksmažodžiai, esantys predikato pozicijoje. Vadinasi, vartotojas tik iš pavyzdžių galės suprasti, su koku veiksmažodžiu koks objekto linksnis bus tinkamas. Pavyzdžiui, antroji *centro* reikšmė yra realizuojama kalbant apie įvairių rūšių centrus – įstaigas, pvz., *prekybos centras, kino centras, paslaugų centras* ir pan. Iš viso ši reikšmė aprašoma 9 vartosenos modeliais (daiktavardžio *centras* dažnumas tekstyne – 410 pavartojimo atvejų). Vartosenos modelyje, kur eina objektu, žodis *centras* gali būti vartojamas galininko, kilmininko arba naudininko linksniu: gramatiniame lygmenyje modelis užrašytas taip:

[Pred] [CENTRAS\_acc][CENTRAS\_gen][CENTRAS\_dat].

Tik iš pavyzdžių bus matyti, kad su *statyti* reikės galininko, o su *priklausyti* – naudininko: *jie statė šalia esančius viešbučius bei prekybos centrus; prieš palikdami automobilį mokamoje aikštelėje, išsiaiškinkite, iki kada ji dirba ir ar nepriklauso prekybos centrui.*

Dažnai semantinės informacijos eilutėje nurodoma semantinė grupė, bet kolokatų eilutėje neįrašytas nė vienas tą grupę iliustruojantis kolokatas. Jeigu tokių atvejų viename modelyje yra keli, vartotojas pats turi susiorientuoti, kurie žodžiai pavyzdžiuose iliustruoja kurią semantinę grupę. Taip pat vartotojui gali būti sudėtingi tokie atvejai, kai modelyje semantinėje eilutėje rodoma grupė be pavadinimo, tada taip pat tik iš pavyzdžių, lyginant kitų grupių žodžius, galima susiorientuoti, kokie žodžiai atspindi tą neįvardytą grupę. Semantinių grupių be pavadinimo atsirado dėl pernelyg margo konteksto, kurio negalima apibendrinti kokia nors viena semantine grupe. Galima kritikuoti, kad *grupės be pavadinimo* kategoriją galbūt būtų buvę geriau rinktis tik tuo atveju, jeigu leksiniame lygmenyje būtų pateikiami kolokatai, susieti su ta grupe.

Leksikone vartotojas ras ir užfiksuoto variantiškumo: pvz., jeigu frazeologizmui būdingas leksinis variavimas, tai parodome atskirais antraštiniais žodžiais (*būti dėmesio centre, atsidurti dėmesio centre*), o morfologinius variantus (*sukti galvą, nesukti galvos*) pateikiame prie pavyzdžių. Kaip matyti iš antroje dalyje paaiškinto vartosenos modelių

aprašymo, modeliuose taip pat dažnai prireikdavo žymėti varijuojančius komponentus (jie atskiriami statmenu brūkšniu). Pavyzdžiui, minėtasis modelis [Pred] [CENTRAS\_acc][CENTRAS\_gen][CENTRAS\_dat] kaip tik rodo, kad predikatas gali būti vartojamas su objektu, kuris gali būti išreikštas galininku, kilmininku arba naudininku.

Leksikone dabar nepažymėta kai kurių vartosenos ypatybių, susijusių su kalbos atmainomis ar pragmatiniais dalykais<sup>14</sup>. Pavyzdžiui, atskirais antraštiniais vienetais pateikti žodžiai *senai* ir *seniai*. Iš dažnumo informacijos galima matyti, kad *seniai* rašytinėje kalboje yra daug dažnesnis, bet leksikone nėra pateikiama informacijos, kad *senai* neteiktinas bendrinėje kalboje reikšme „seniai“. Žodžio *nepergyventi* reikšmė „nesijaudinti“ nevertotina bendrinėje kalboje, tačiau tai irgi nepažymėta. Yra nedaug atvejų (pvz., dalelytė *tipo*), būdingų slengui, tačiau atskirų tai rodančių pažymų leksikone nėra. Tiesa, antraštyne tokių žodžių reta, o skirtingoms kalbos atmainoms, stiliams priklausantys žodžiai gali dažniau pasitaikyti pavyzdžiuose: tarkim, vieno iš homonimų *pašokti* pavyzdžiuose yra laisvesnio, šnekamojo stiliaus raiškos, pvz., *baliavoti* aprašomojo žodžio *pašokti* pavyzdžiuose (*Susirenkam tai pas vieną, tai pas kitą, tai pabaliavoji, tai pašoki.*), o slengo žodis *žiauriai* – prieveiksmio *daug* pavyzdžiuose (*Taip, normaliomis aplinkybėmis, kai nereikia žiauriai daug mokyti, kasdien po valandą kankinuosi.*).

## 5 lentelė

*Būdvardžio paskutinis vartosenos modelis*

Lygmuo	Vartosenos modelis
Gramatinis lygmuo	[PASKUTINIS] [Mod]
Semantinis lygmuo	[AtrA] [asmuo]
Leksinis lygmuo	[PASKUTINIS] [ ]
Pavyzdžiai	<i>Jeigu nejudėsite į priekį, jus gali apimti jausmas, kad esate niekam nereikalingi, kad esate paskutiniai netikėliai ir nevykėliai. ...į tokį skelbimą atsilieptų tik paskutinės ištvirkėlės. Jie ten sėdi visi, išsišiepę it paskutiniai idiotai.</i>

<sup>14</sup> Lingvistai informaciją apie pragmatines ypatybes žymėjo leksikono duomenų bazėje, tačiau vartotojams šie duomenys kol kas nėra pateikiami.

Vartosenos modelyje nėra žymimas neigiamas vertinimas, pvz., modelyje [PASKUTINIS] [Mod] komponentas [Mod] realizuojamas tik neigiamą konotaciją turinčiais asmenų įvardijimais: *nevykėliai, ištvirkėlės, idiotai*. Ši informacija matyti tik pavyzdžiuose, ji nėra fiksuojama modelyje (žr. 5 lentelę).

**Leksikonas skirtas produkciniams gebėjimams ugdyti, todėl čia siekiama parodyti žodžių ryšius, išryškėjusius tekстыne.** Kaip teigia užsienio kalbų mokymo praktikai, mokant leksikos, ypač pažengusius mokinius, labai svarbūs žodžių ryšiai (Lewis, 2001). Ugdant leksinę kompetenciją, svarbu mokyti ne izoliuotų žodžių, bet analizuoti jų ryšius: darybinius, leksinius, semantinius. Kaupiant informaciją apie žodžių vartoseną ir ryšius leksikone, siekta, kad mokiniai ir jų mokytojai turėtų išteklių, naudingą aktyviam leksikos mokymui(si).

Leksikone vartotojas gali rasti informacijos apie žodžių **semantinius (paradigminius) ryšius** vartosenos modeliuose, kur nurodomos semantinės grupės ir su jomis susieti konkretūs kolokatai, pvz., minėtame *skambinti* reikšmės „groti“ modelyje objektas reiškiamas semantinės grupės „muzikos instrumentas“ žodžiais, o nurodytas kolokatas yra *pianinas*. Vartotojui būtų aktualu parodyti ir su kiekviena žodžio reikšme susietas semantines grupes: tarkim, prie aprašomo žodžio *sausis* nurodyti ir kitus mėnesių pavadinimus; taip pat aktualūs sinoniminiai, antoniminiai ir hiperoniminiai ryšiai. Kol kas leksikone tai nepadaryta, bet informaciją apie susijusius ir priešingos reikšmės žodžius mokinys gali rasti jau esamuose žodynuose, pvz., *Mokomajame žodyne* prie *anksti* nurodyta *vėlai, vėlu*; prie *spyna* – *užraktas*; žodžiui *siauras* nurodomas priešingos reikšmės žodis *platus*; prie *stalas* pateikta nuoroda į *rašomasis stalas*; prie *ragelis* – nuoroda į *telefonas*.

Leksikone vartotojai galės studijuoti žodžių **leksinį junglumą**: kolokatai nurodomi atskirame vartosenos modelio lygmenyje, plg. aprašomo daiktavardžio *amžius* modelyje [AtrA AMŽIUS\_sg.gen] [Mod] leksiniame lygmenyje nurodomus kolokatus, kurie užima derinamojo pažyminio vietą: *įvairaus, senyvo, panašaus, garbaus, garbingo, vyresnio, kokio, tokio, mokyklinio, pensinio, vidutinio, darbingo, x metų amžiaus* + [asmuo]. Nors leksikono antraštyne nepateikti būdvardžiai *senyvas, garbus, garbingas, vyresnis, pensinis*, šiuos būdvardžius vartotojas ras prie daiktavardžio *amžius*.

Vadinasi, modeliuose fiksuojant žodžiui būdingą leksinę aplinką, vartotojui suteikiama informacija apima daugiau žodžių, nei yra antraštyne. Kad šie ryšiai būtų labiau matomi ir lengviau randami, leksikono vartotojo sąsajoje veikia kolokatų paieška: vartotojas, įvedęs dominantį žodį, galės rasti visus su juo sietinus žodžius, pvz., įvedus *senas*, bus galima matyti, kad šis būdvardis yra dažnas kolokatas žodžiams *tradicija*, *draugas*, *žmogus* ir kt. Kadangi dėl pasirinkto leksikono antraštyno formavimo principo palyginti nemažai dažnos leksikos galėjo likti neaprašytos, naudinga tai, kad su ja vartotojas galės susipažinti analizuodamas žodžių ryšius.

Svarbu priminti, kad kolokatai nurodyti tik baziniams žodžiams, kurie aprašyti ilguoju aprašu (t. y. su modeliais). Kolokatų skaičius susijęs su modelių skaičiumi, konkretaus modelio gramatiniais ir semantiniais komponentais (kolokatų gali būti ir nepateikta dėl per mažo dažnumo). Daugiausia kolokatų nurodyta prie tų žodžių, kurie yra labai dažni: *darbas* (115 kolokatų), *laikas* (99), *metai* (97), *eiti* (89), *gerai* (86), *gyvenimas* (82), *žmogus* (78), *diena* (73), *kalbėti* (70). Taigi, didesnis žodžio dažnumas ir tarp pavyzdžių, ir tarp kolokatų tekстыne lėmė, kad apie juos sukaupta daugiau informacijos.

Baziniai žodžiai pagal kolokatų skaičių pasiskirsto taip:

- a) nuo 115 iki 60 kolokatų – 12 žodžių;
- b) nuo 59 iki 40 kolokatų – 28 žodžiai;
- c) nuo 39 iki 20 kolokatų – 137 žodžiai;
- d) nuo 19 iki 10 kolokatų – 262 žodžiai;
- e) nuo 9 iki 1 kolokato – 273 žodžiai.

Leksikone vartotojai ras daug informacijos ir apie **darybinius ryšius**. Žodžių darybą suprantame plačiai: ir kaip morfologinę (kai pridedami priešdėliai, priesagos, suduriami žodžiai, keičiamos morfemos), ir kaip funkcinę (sintaksinę) (kai žodis ar tam tikra jo žodžio forma pakeičia funkciją, virsta kita kalbos dalimi, pvz., *trumpam*, *ilgam* kaiprieveiksmiai siejami su pamatiniais žodžiais *trumpas*, *ilgas*). Toliau trumpai aptarsime leksikone taikomus darinių pateikimo principus: 1) dariniai pateikiami atskirais antraštiniais žodžiais; 2) dariniai nurodomi kiekvienai antraštinio žodžio reikšmei atskirai.

Leksikone kai kurie reguliarieji dariniai (pvz., žodžiai, sudaryti su neiginiais (pvz., *nenorėti, neseniai, negeras*), kitaip nei seniau rengtuose aiškinamuosiuose žodynuose, pateikiami atskirais įrašais, o ne prie pamatinio žodžio. Panaši praktika taikyta ir BLKŽ, argumentuojant, kad įtrauktinis vedinio pateikimas nepaaiškina, neatskleidžia jo semantikos (ar nepakankamai, netiksliai ją paaiškina) (Murmulaitytė, 2012, p. 76).

Kadangi leksikone dariniai pateikiami prie konkrečios aprašomo žodžio reikšmės, taip išvengiama dažnos lietuvių kalbos aiškinamųjų žodynų problemos, kurią kritikuoja Jolanta Vaskelienė (2006, p. 57), pvz., apie *-ėjas ir -ikas* vedinius DLKŽ ji rašo: „Kadangi DLKŽ priesagų *-ėjas ir -ikas* vediniai rašomi visų (ir vienareikšmių, ir daugiareikšmių) veiksmažodžių straipsnyje, atrodo, kad jie gali remtis bet kuria pamatinio veiksmažodžio reikšme. <...> tačiau dažniausiai (dėl sintaksinio ir semantinio junglumo) aptariamų priesagų vediniai turi mažiau reikšmių negu jų pamatiniai veiksmažodžiai.“

Kol kas leksikone nenurodyta, kokia (-iomis) pamatinio žodžio reikšme (-ėmis) remiasi vedinys, jei pamatinis žodis turi daugiau negu vieną reikšmę. Kitaip sakant, kai nurodome vedinius prie konkrečių reikšmių, nenurodome konkrečios reikšmės, kuria tas vedinys yra darybiškai susijęs su aprašomo žodžio reikšme. Pavyzdžiui, prie *nelaikyti* yra dvi reikšmės: 1) ***nelaikykite prietaiso ant stalo***; 2) ***nelaikykite to praradimu***. Prie kiekvienos reikšmės pamatinis žodis yra veiksmažodis *laikyti*, bet pirmajai *nelaikyti* reikšmei *laikyti* bus pamatinis reikšme „turėti“, o antrajai reikšmei – pamatinis reikšme „vertinti“, nors tai nenurodyta. Vaskelienė (2006: 67) teikia rekomendaciją: jeigu vedinys remiasi pamatiniu žodžiu, kuris yra daugiareikšmis, prie vedinio svarbu nurodyti, kuria pamatinio žodžio reikšme jis remiasi. Pavyzdžiui, BLKŽ gryniesiems abstraktai (veiksmų ar ypatybių pavadinimai) pateikiami nurodant pamatinį veiksmažodį ar vardažodį bei atitinkamą jų reikšmės numerį (Murmulaitytė, 2009, p. 77). Tokia praktika galėtų būti ir viena iš leksikono tobulinimo krypčių, nes ši informacija aktualiai negimtakalbiams mokiniams.

Kita negimtakalbiams problemų kelianti ypatybė leksikone jau fiksuojama: leksikone darybiškai susiejami, bet semantiškai su aprašomu žodžiu nesusiję žodžiai (pvz., *sakyti ir užsakyti*) pažymimi specialiu ženklau. Su pamatiniu žodžiu susiejame tik tuos vedinius, kurie gali būti suprantami



sinchroniškai, pvz., *eiti* ir *nueiti*. Jei sąsaja nustatoma tik diachroniškai, pvz., *tikti* ir *tikslus*, *tikslas*; *gyti* ir *gyvas*; *skiesti* ir *skaidrus*, tokių atvejų nelaikome vediniais, nenurodome pamatinio žodžio (išsamiau apie veiksmažodžių darybinius ryšius žr. Jancaitė, 2020).

Leksikone yra „nutrūkusių“ darybinių ryšių, pvz., randami žodžiai *vykti*, *nejvykti*, bet nėra *įvykti*, iš kurio padarytas į leksikoną įtrauktas žodis *nejvykti*. Tokiais atvejais pamatinis žodis nurodomas, tačiau prie jos prirašomas specialus simbolis norint parodyti, kad tekstyne, iš kurio nustatyti darybiniai ryšiai, tokio žodžio nėra. Taip pat pasitaiko, kad leksikone nurodytų vedinių pamatiniai žodžiai yra, bet jie nepavartoti ta reikšme, kuri susijusi su vedinio reikšme: tarkim, prie vedinio *nepergyventi* reikšmės „nesijaudinti“<sup>15</sup> nurodytas pamatinis žodis *pergyventi*, bet tarp aprašytų šio žodžio reikšmių reikšmės „jaudintis“ nėra, nes tokia reikšmė *pergyventi* nepavartotas tekstyne.

Dariniai yra įvairaus dažnumo: pavyzdžiui, baziniam žodžiui *vieta* leksikone nurodyti šie dariniai (skliaustuose nurodytas dažnumas): prie pirmos reikšmės: *dvarvietė* (3), *gyvenvietė* (30), *turgavietė* (6), *prekyvietė* (3), *augavietė* (3), *vietinis* (74), *vietovė* (58), *vietelė* (9), *vietomis* (13); prie antros reikšmės: *dvivietis* (21); prie ketvirtos reikšmės: *darbovietė* (16).

## 6 lentelė

*Antraštinių žodžių skaičius pagal jų darinių skaičių*

<b>Antraštinių žodžių skaičius</b>	33	16	13	18	27	31	62	106	188	434
<b>Darinių skaičius</b>	Daugiau nei 10	9	8	7	6	5	4	3	2	1

Leksikono antraštyne (atmetus pastoviuosius žodžių junginius) didžioji dalis vienažodžių leksinių vienetų yra dariniai (vediniai arba dūriniai), iš viso – 2 488 žodžiai (tai sudaro 80 proc. vienažodžių leksinių vienetų). Tarp bazinės leksikos žodžių darinių, žinoma, daug mažiau (jie sudaro apie 37 proc.). Prie 940 antraštinių žodžių (ar jų reikšmių) pateikta nuo 1 iki 22 darinių, iš viso –

<sup>15</sup> Kaip minėta anksčiau, leksikone nenurodoma informacija apie norminę ar nenorminę žodžių vartoseną.

2 542 dariniai (vidutiniškai 2,7 darinio). Yra 33 žodžiai (žr. 6 lentelę), kurių apraše parodytų darybinių ryšių skaičius nuo 22 iki 10 darinių (pvz., *diena* – 22 dariniai, *eiti* – 19 darinių, *vesti* – 17 darinių; *mažas, kelti* – po 15, *gyventi, rašyti, bėgti* – po 14, *važiuoti, augti* – po 13, *darbas, ranka* – po 12, *daryti, vieta, statyti, pusė* – po 11, *laikas, žiūrėti* – po 10). 6 lentelėje matyti, kad gana nedaug žodžių susieti su labai daug darybinių ryšių (turi nuo 9 iki 4 darinių), o mažai darinių (nuo 3 iki 1) turi apie 700 antraštinių žodžių (tai sudaro apie 70 proc. visų žodžių, turinčių darinius). Tokie rezultatai nestebina, turint omenyje tekstyno dydį ir leksinę įvairovę.

Viena vertus, leksikone aprašytos leksikos vaizdas nėra toks išsamus, kokį gautume panaudoję didelio tekstyno duomenis: kaip jau rašyta, leksikone kai kurių žodžių reikšmių nėra arba yra trūkinėjančių darybinių ryšių. Kita vertus, mūsų darbo rezultatas panašus į negimtakalbio galvoje formuojamą žodyno vaizdą (plg. Lewis, 2001), kai ryšiai tarp žodžių yra neišsamūs, nes ne viskas žinoma, ne viskas įsiminta, kartais trūksta jungčių, nes vyksta mokymasis, kai informacija atnaujinama, perdėliojama, pamirštama ir išmokstama pakartotinai.

Leksinė kompetencija susideda iš kelių lygmenų ir tos skirtingos informacijos mokymasis vyksta nevienodai, žodžio mokymasis yra laipsniškas procesas (Barclay et al., 2019, p. 806), be to, skirtingiems žodžiams gali tikti skirtingos mokymo(si) strategijos (žr. Schmitt, 2010, p. 35). Tačiau tyrimai apie mokomųjų žodynų medžiagos naudą rodo (Frankenberg-Garcia, 2015, p. 95), kad suprasti nežinomo žodžio reikšmę besimokančiajam vienodai padeda ir kai pateikiamas žodžio aiškinimas, ir kai pateikiamas supratimo pavyzdys (tiesa, jų reikia daugiau negu vieno). Be to, yra daug tyrimų (Bogaards, van der Kloot (2002), Chan (2011), Dziemianko (2012), apžvalgą žr. Frankenberg-Garcia, 2015), kurių išvados patvirtina mintį, kad pavyzdžiai padeda kalbos produkavimui.

### **Apibendrinamosios pastabos**

Šiame straipsnyje pristatytą leksikoną galima laikyti aktyviojo mokomojo žodyno prototipu. Leksikonas vertintinas kaip duomenynas, papildantis jau esamus (ypač aukštesniųjų lygių) lietuvių kalbos mokymo resursus.

Ši leksinė bazė yra pirmasis bandymas pasiūlyti tekstyno pagrindu parengtą lietuvių kalbos aprašą, kuris galėtų būti naudingas mokytojams ir dėstytojams, norintiems savo praktikoje dažniau taikyti kalbos vartosenos analize paremtus metodus, ypač organizuojant mokymo veiklas pažengusiems studentams. Kadangi leksikono medžiaga skirta kalbos produkavimo (rašymo, kalbėjimo) gebėjimams lavinti, recepcijos (skaitymo, klausymo) poreikiams patenkinti galima naudoti ne tik *Mokomąjį lietuvių kalbos žodyną*, bet ir atitinkamus dvikalbius žodynus: atitikmenys savoje kalboje padėtų suvokti reikšmę, o leksikone aprašyti žodžių vartosenos modeliai ir pavyzdžiai atskleistų, kaip ta reikšmė realizuojama lietuvių kalboje.

Leksikonas prieinamas portale <https://kalbu.vdu.lt/mokymosi-priemones/leksikonas>. Čia yra paaiškinamos leksinės bazės paieškos galimybės. Leksikono naudotojams jau parengta ir mokymo priemonė (Kovalevskaitė et al. 2022), kaip šį išteklių galima pritaikyti lietuvių kalbos mokymui(si).

## Literatūra

- Barclay, S., & Schmitt, N. (2019). Current Perspectives on Vocabulary Teaching and Learning. In X. Gao (Ed.), *Second Handbook on English Language Teaching* (pp. 799–819). Springer.
- Bielinskienė, A., Kovalevskaitė, J., & Rimkutė, E. (2021). *Grammatical Patterns in the Corpus-Driven Lexical Database of Lithuanian*. Spausdinama: *Language: Meaning and Form 12 (Valoda: nozīme un forma 12)*.
- Bielinskienė, A., Boizou, L., Bumbulienė, I., Kovalevskaitė, J., Krilavičius, T., Mandravickaitė, J., Rimkutė, E., & Vilkaitė-Lozdienė, L. (sud.). (2019). *Lietuvių kalbos kolokacijų žodynas*. Vytauto Didžiojo universitetas. <http://mwe.lt/wp-content/uploads/2019/02/zodynas.pdf>
- Boizou, L., Kovalevskaitė, J., & Rimkutė, E. (2020). Lithuanian Pedagogic Corpus: Correlations between Linguistic Features and Text Complexity. The Ninth Conference Human Language Technologies – The Baltic Perspective. In A. Utkā, J. Vaičėnienė, J. Kovalevskaitė, D. Kalinauskaitė (Eds.), *Frontiers in Artificial Intelligence and Applications* 328 (pp. 233–240). IOS Press.

<https://doi.org/10.3233/FAIA200628>

- Brezina, V., & Gablasova, D. (2015). Is There a Core Vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36(1), 1–22.
- Dargis, R., Auziņa, I., Levāne-Petrova, K., & Kaija, I. (2020). Quality Focused Approach to a Learner Corpus Development. *Proceedings of the 12<sup>th</sup> International Conference on Language Resources and Evaluation (LREC 2020)*. European Language Resources Association (ELR). <http://lava.korpuss.lv/publicatoins/LREC2020-Dargis.pdf>
- Frankenberg-Garcia, A. (2015). Dictionaries and encoding examples to support language production. *International Journal of Lexicography*, 24(4), 490–512.
- François, Th., Gala, N., Watrin, P., & Fairon, C. (2014). FLELex: a graded lexical resource for French foreign learners. *Proceedings of International conference on Language Resources and Evaluation (LREC 2014)*, 91–102. [http://www.lrec-conf.org/proceedings/lrec2014/pdf/1108\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/1108_Paper.pdf)
- Grigonytė, G., Kovalevskaitė, J., & Rimkutė, E. (2018). Linguistically-Motivated Automatic Classification of Lithuanian Texts for Didactic Purposes. *Proceedings of the Eighth International Conference Baltic HLT 2018*. K. Muischnek, & K. Müürisep (Eds.), *Frontiers in Artificial Intelligence and Applications* 307 (pp. 38–46). IOS Press. <https://doi.org/10.3233/978-1-61499-912-6-38>
- Hanks, P. (2004). Corpus Pattern Analysis. *Euralex 2004 Proceedings* (pp. 87–99). Université Bretagne Sud. [http://www.euralex.org/elx\\_proceedings/Euralex2004/009\\_2004\\_V1\\_Patrick%20HANKS\\_Corpus%20pattern%20analysis.pdf](http://www.euralex.org/elx_proceedings/Euralex2004/009_2004_V1_Patrick%20HANKS_Corpus%20pattern%20analysis.pdf)
- Hanks, P. (2012). How People Use Words to Make Meanings: Semantic Types Meet Valencies. In A. Boulton, & Th. James (Eds.), *Input, Process and Product: Developments in Teaching and Language corpora* (pp. 52–67). Masaryk University Press.
- Hill, J. (2000). Revising priorities: from grammatical failure to collocational success. *Teaching Collocations: Further Development in the Lexical Approach*. Language Teaching Publication, 47–69.
- Humblé, Ph. (2001). *Dictionaries and Language Learners*. Haag & Herchen.

- Jakaitienė, E. (2005). *Leksikografija*. Mokslo ir enciklopedijų leidybos institutas.
- Jancaitė, L. (2020). Priešdėlinių lietuvių kalbos veiksmažodžių leksikalizacija. *Bendrinė kalba*, 93, 1–27. <http://journals.iki.lt/bendrinekalba/article/view/2037/2130>
- Jasionytė-Mikučionienė, E. (2016). Imperatyvinės kilmės diskurso markeriai lietuvių kalboje: *klausyk* ir *žiūrėk* atvejais. *Kalbotyra*, 68, 23–41. <https://doi.org/10.15388/KIbt.2016.10317>
- Jezek, E., & Hanks, P. (2010). What lexical sets tell us about conceptual categories. *Lexis. Journal in English Lexicology*, 4, 1–15.
- Kallas, J., Tuulik, M., & Langemets, M. (2014). The Basic Estonian Dictionary: the first Monolingual L2 learner's Dictionary of Estonian. *Proceedings of the 16<sup>th</sup> EURALEX International Congress* (pp. 1109–1117). EURAC research.
- Kamandulytė-Merfeldienė, L., & Balčiūnienė, I. (2016). Apie pažymimuosius pasakymus sakininėje lietuvių kalboje. *Taikomoji kalbotyra*, 8, 55–71. <https://www.zurnalai.vu.lt/taikomojikalbotyra/article/view/17506/16675>
- Kapočiūtė-Dzikienė, J., Rimkutė, E., & Boizou, L. (2017). A Comparison of Lithuanian Morphological Analyzers. In K. Ekšteina, V. Matoušek (Eds.), *Lecture Notes in Computer Science, vol 10415* (pp. 47–56). Springer, Cham. [https://doi.org/10.1007/978-3-319-64206-2\\_6](https://doi.org/10.1007/978-3-319-64206-2_6)
- Kovalevskaitė, J., & Jancaitė, L. (2019). Vartosenos modelių analizė mokomojoje leksikografijoje: žvalgomas tyrimas lietuvių kalbos veiksmažodžių pavyzdžiu. *Taikomoji kalbotyra*, 12, 124–154. <https://www.zurnalai.vu.lt/taikomojikalbotyra/article/view/17235/16403>
- Kovalevskaitė, J., Bielinskienė, A., Boizou, L., Jancaitė, L., & Rimkutė, E. (2020a). The First Corpus-driven Lexical Database of Lithuanian as L2. The Ninth Conference Human Language Technologies – The Baltic Perspective. In A. Utkā, J. Vaičėnienė, J. Kovalevskaitė, D. Kalinauskaitė (Eds.), *Frontiers in Artificial Intelligence and Applications* 328 (pp. 245–252). IOS Press. <https://doi.org/10.3233/FAIA200630>

- Kovalevskaitė, J., & Rimkutė, E. (2020b). Mokomasis lietuvių kalbos tekstynas: naujas išteklius lietuvių kalbos besimokantiejiems. *Darnioji daugiakalbystė*, 17, 201–233. <https://doi.org/10.2478/sm-2020-0019>
- Kovalevskaitė J., Rimkutė E., Ruzaitė J. (2022). *Tekstynai ir jų išvestiniai produktai lietuvių kalbos mokymui(si) bei tyrimams*. Vytauto Didžiojo universitetas. <https://doi.org/10.7220/9786094675089>
- Labutis, V. (2002). *Lietuvių kalbos sintaksė*. Vilniaus universiteto leidykla.
- Lewis, M. (2001). *Teaching Collocations: Further Development in the Lexical Approach*. Language Teaching Publication.
- Marcinkevičienė, R. (2000). Tekstynų lingvistika: teorija ir praktika. *Darbai ir Dienos*, 24, 7–64.
- Murmulaitytė, D. (2012). Vedinių pateikimo sistemiškumas *Bendrinės lietuvių kalbos žodyne*. *Kalbos kultūra*, 85, 71–91. [http://www.bendrinekalba.lt/Straipsniai/85/Murmulaityte\\_KK\\_85\\_straipnis.pdf](http://www.bendrinekalba.lt/Straipsniai/85/Murmulaityte_KK_85_straipnis.pdf)
- Murmulaitytė, D. (2009). Bendrinės lietuvių leksikos duomenų bazė – ne elektroninis Bendrinės lietuvių kalbos žodynas. *Kalbų studijos*, 15, 77–84. <https://etalpykla.lituanistikadb.lt/fedora/objects/LT-LDB-0001:J.04~2009~1367168037459/datastreams/DS.002.0.01.ARTIC/content>
- Narbutas, E., Pribušauskaitė, J., Ramonienė, M., Skapienė, S., & Vilkienė, L. (2002). *Slenkstis*. Council of Europe Press.
- Norkaitienė, M., Šepetytė, R., & Šimėnaitė, Z. (2010). *Mokomasis lietuvių kalbos žodynas* (antras pataisytas leidimas). Baltos lankos.
- Pribušauskaitė, J., Ramonienė, M., Skapienė, S., & Vilkienė, L. (2000). *Aukštuma*. Council of Europe Press.
- Ramonienė, M., & Pribušauskaitė, J. (2003). *Praktinė lietuvių kalbos gramatika*. Baltos lankos.
- Ramonienė, M., Pribušauskaitė, J., & Vilkienė, L. (2006). *Pusiaukelė*. Europos Taryba.
- Ramonienė, M., Brazauskienė, J., Burneikaitė, N., Daugmaudytė, J., Kontutytė, E., & Pribušauskaitė, J. (2012). *Lingvodidaktikos terminų žodynas*. Vilniaus universiteto leidykla.
- Rimkutė, E., Bielinškienė, A., & Kovalevskaitė, J. (sud.) (2012). *Lietuvių kalbos*

- daiktavardinių frazių žodynas*. Vytauto Didžiojo universitetas.  
[http://donelaitis.vdu.lt/lkk/pdf/daikt\\_fr.pdf](http://donelaitis.vdu.lt/lkk/pdf/daikt_fr.pdf)
- Ruzaitė, J., Dereškevičiūtė, S., Kavaliauskaitė, V., & Krivickaitė, E., (2020). Error tagging in the Lithuanian learner corpus. The Ninth Conference Human Language Technologies – The Baltic Perspective. In A. Utkā, J. Vaičėnonienė, J. Kovalevskaitė, D. Kalinauskaitė (Eds.), *Frontiers in Artificial Intelligence and Applications* 328 (pp. 253–260). IOS Press.  
<https://doi.org/10.3233/FAIA200631>
- Schmitt, N. (2010). Key Issues in Teaching and Learning Vocabulary. *Insights into Non-native Vocabulary Teaching and Learning*. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Blue Ridge Summit* (pp. 28–40). Multilingual Matters.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sinclair, J. (1996). The Search for Units of Meaning. *Textus IX, 1*, 75–106.
- Vaskelienė, J. (2006). Dėl kai kurių darinių pateikimo Bendrinės lietuvių kalbos žodyne. In R. Petrokienė (Sud.), *Leksikografija ir leksikologija. Aiškinamųjų bendrinės kalbos žodynų aktualijos* (pp. 55–71). Lietuvių kalbos institutas.

**Jolanta Kovalevskaitė**

Vytautas Magnus University, Lithuania  
jolanta.kovalevskaite@vdu.lt

**Erika Rimkutė**

Vytautas Magnus University, Lithuania  
erika.rimkute@vdu.lt

**A NEW CORPUS-DRIVEN LEXICAL DATABASE FOR LITHUANIAN  
AS A FOREIGN LANGUAGE**

**Summary.** In this paper, we describe a new lexicographic resource for advanced learners of Lithuanian, the *Lexical Database of Lithuanian Language Usage*, which is the first attempt in Lithuanian lexicography to prepare a description of vocabulary based on the word usage analysis in the particular corpus. The written subpart of the Lithuanian Pedagogic Corpus (approx. 620,000 tokens) was used to develop headword lists and collect word usage information in the form of corpus patterns. In the database, there are 3,700 lexical items, words and multi-word units (compounds, idioms or sayings). For the appr. 700 most frequent words from a shared vocabulary (they appear in texts assigned to A1, A2, B1 and B2 levels, and their frequency in the whole corpus is 100 occurrences and above), we prepared a full-record entry: it includes sense-related corpus patterns with grammatical, semantic and lexical information and the examples illustrating all pattern components. The short-record entry (no patterns, only examples) is prepared for the less frequent words from the shared vocabulary, which are derivationally related to the most frequent headwords. The users are provided with 2,542 derivatives, which are linked to 940 headwords. In the database, 28,550 encoding examples are manually selected for all 3,000 headwords and 700 phrases. We discuss the features of the database, and, particularly, the adopted semi-automated procedure of Corpus Pattern Analysis, which was used for the description of word usage. We evaluate the approach applied, and discuss its advantages for users as well as provide the suggestions for the future improvements of the resource, which can be used as an additional resource in the classroom of Lithuanian as a foreign language, and, together with the available corpora, fill in a gap of usage information in the existing (learner) dictionaries.

**Keywords:** Lexical database; Corpus Pattern Analysis; corpus; corpus linguistics; learner lexicography; Lithuanian language.