

Leksikografiniai tarmių duomenys lietuvių kalbos išteklių informacinėje sistemoje „E. kalba“

Lexicographic Data on Dialects in the Lithuanian Language Resources
Information System E-Language

Jolita Urbanavičienė
kalbos institutas, Lithuania
jolita.urbanaviciene@lki.lt

 <https://orcid.org/0000-0002-5514-2263>

Lietuvių kalba núm. 17 109 121 2022

Vilniaus Universitetas
Lituania

Recepción: 24 Octubre 2022
Aprobación: 28 Noviembre 2022

Summary: Straipsnyje pristatoma lietuvių kalbos išteklių informacinė sistema „E. kalba“, ypač tie ištekliai, kuriuose randami leksikografiniai tarmių duomenys: „Lietuvių kalbos žodynas“ ir trys jo kartotekos: LKŽ Pagrindinė kartoteka, LKŽ Papildymų kartoteka ir LKŽ Tarmių kartoteka. Straipsnyje analizuojama modernizuota šių leksikografinių išteklių duomenų struktūra, paieškos sistema, aptariamų sistemos kūrėjams iškylusios problemos ir jų sprendimo būdai. Daroma išvada, kad „E. kalbos“ sukūrimas kiekybiškai ir kokybiškai pagerino ankstesniąją, 2015 m. pristatytą Lietuvių kalbos išteklių informacinę sistemą (LKIS): ji buvo modernizuota tiek kiekybiniu lygmeniu (į informacinę sistemą integruoti 3 žodynai, 1 kartoteka ir sukurtos 7 paslaugos), tiek kokybiniu lygmeniu (pagerintas lietuvių kalbos išteklių prieinamumas, patogumas ir tarpusavio integralumas). „E. kalbos“ kūrimo procesas pagreitino ir kitų elektroninių produktų bei paslaugų plėtrą: prisidėjo sprendžiant sistemines lietuviško šrifto „Palemonas“ problemas, paskatino klaviatūros tvarkyklės „Kopenhagos transkripcija“ atsiradimą ir kt. Vienas iš tolesnių informacinės sistemos „E. kalba“ plėtros uždavinių galėtų būti integruotų išteklių kiekybinis pildymas, nes tik nedidelė dalis „Lietuvių kalbos žodyno“ kartotekų yra suskaitmenintos.

Keywords: Lietuvių kalbos išteklių informacinė sistema E kalba, Lietuvių kalbos žodynas, Lietuvių kalbos žodyno Pagrindinė kartoteka, Lietuvių kalbos žodyno Papildymų kartoteka, Lietuvių kalbos žodyno Tarmių kartoteka, Palemono šriftas.

Abstract: The article presents the Lithuanian Language Resources Information System *E-Language*, with a focus on those resources which include the lexicographic data on dialects: *The Dictionary of the Lithuanian Language* (DLL) and its three catalogues, namely *The Main Catalogue*, *The Catalogue of New Additions*, and *The Catalogue of Dialects*. The article addresses the modernized structure of the data from these lexicographic resources and its search system and discusses the problems faced by the developers of the system and their solutions. It is concluded that the development of *E-Language* marked a quantitative and qualitative improvement of its predecessor, the Lithuanian Language Resources Information System (LLRIS), which came to life in 2015: it was modernized on both a quantitative level (by integrating 3 dictionaries and 1 catalogue into the information system and creating 7 services) and a qualitative level (the availability, usability and integrity of Lithuanian language resources were improved). *E-Language* development process also accelerated the advancement of other electronic products and services: it contributed to solving the systemic problems of the Lithuanian font *Palemonas*, encouraged the appearance of the keyboard layout

known as the Copenhagen Transcription, etc. One of E-Language could be the quantitative addition of integrated resources, as only a small part of the catalogues of The Dictionary of the Lithuanian Language is digitalized.

Keywords: Lithuanian Language Resources Information System E-Language, The Dictionary of the Lithuanian Language, The Main Catalogue, The Catalogue of New Additions, The Catalogue of Dialects, the Lithuanian font Palemonas.

„E. kalba“ – modernizuota lietuvių kalbos išteklių informacinė sistema

2021 m. plačiai visuomenei buvo pristatyta Lietuvių kalbos institute sukurta lietuvių kalbos išteklių informacinė sistema „E. kalba“ (<https://ekalba.lt/>). Ji sudaryta iš trijų stambių dalių: „E. žodynai“, „E. kartotekos ir duomenų bazės“ ir „E. paslaugos“. Sistema pradėta kurti 2012–2015 m., kai Lietuvių kalbos institutas kartu su partneriais – Lietuvių literatūros ir tautosakos institutu, Vilniaus universitetu ir tuometiu Lietuvos edukologijos universitetu – vykdė projektą „IRT sprendimų bei turinio, padedančių išsaugoti lietuvių kalbą viešojoje erdvėje, kūrimas bei galimybių jais naudotis sudarymas“^[1]. Projekto metu buvo sukurta Lietuvių kalbos išteklių informacinė sistema (LKIIS), kurią sudarė 11 suskaitmenintų vienakalbių ir dvikalbių žodynų, 5 kartotekos, elektroninės šių išteklių duomenų bazės ir jų valdymo įrankiai. LKIIS'u visuomenė galėjo naudotis nuo 2015 m.

Pasibaigus pirmajam projektui iškart pradėta ruošti LKIIS'o modernizavimui, kad kalbiniai ištekliai tarpusavyje būtų geriau integruoti, susieti prasminiais ryšiais, geriau atitiktų vartotojų poreikius, be to, buvo siekiama į LKIIS'o struktūrą įtraukti dar daugiau reprezentatyvių lietuvių kalbos išteklių. O pagrindinis projekto tikslas buvo didinti lietuvių kalbos išteklių prieinamumą elektroninėje erdvėje ir jų panaudojimą išvestiniams produktams bei elektroninėms paslaugoms teikti. Tad 2018–2021 m. vykdyta LKIIS'o plėtra^[2] vyko dviem kryptimis: kiekybiškai, t. y. integruojant naujus išteklius, ir kokybiškai, t. y. plečiant sistemos funkcionalumus. 2021 m. visuomenei pristatyta modernizuota lietuvių kalbos išteklių informacinė sistema „E. kalba“, kuri papildyta (žr. 1 pav.):

1. trimis žodynais: įkeltas „Lietuvių kalbos žodynas“ (LKŽ, <https://ekalba.lt/lietuviu-kalbos-zodynas/>), „Bendrinės lietuvių kalbos žodynas“ bei jo antraštytas (BLKŽ, <https://ekalba.lt/bendrinės-lietuviu-kalbos-zodynas/>) ir „Lietuvių kalbos naujažodžių duomenynas“ (LKND, <https://ekalba.lt/naujazodziai/>);
2. viena kartoteka: įkelta Kazimiero Būgos etimologinio žodyno kartoteka (BEŽK, <https://ekalba.lt/BugosEtimologinioZodynoKartoteka/>);
3. septyniomis elektroninėmis paslaugomis, suskirstytomis į tris grupes: sukurtas „Žodžių prasmų tinklas“, kuris susideda iš e. paslaugų „Paieška Žodžių prasmų tinkle“ (ŽPT, <https://ekalba.lt/zodziu-prasmiu-tinklas/?p=1>) ir „E. sąvokos“ (ES,

<https://ekalba.lt/esavokos/>). Pastarąją paslaugą sudaro „Ekonomikos sąvokų ontologijos“, „Kompiuterinės technikos ir jos dalių sąvokų ontologijos“ ir „Žmogaus anatomijos sąvokų ontologijos“. Paslauga „E. rinkodara“ susideda iš „E. pavadinimo“ (EP, <https://ekalba.lt/epavadinimas/>), „Naudotojų užklausų analizės“ (NUA, <https://ekalba.lt/naudotoju-uzklausu-analize/>) ir „Nuomonių analizės“ (NA, <https://ekalba.lt/nuomoniui-analize/>). Trečioji paslaugų grupė – „E. patarimai“, kuri susideda iš e. paslaugų „Žodžių darybos vedlys“ (ŽDV, <https://ekalba.lt/zodziudarybos-vedlys/>) ir „Kalbos patarimai“ (KP, <https://ekalba.lt/kalbos-patarimai/>).

Naujoji sistema ne tik papildyta žodynais bei kartotekomis, bet ir sukurta naujos elektroninės paslaugos, padidintas lietuvių kalbos elektroninių išteklių prieinamumas, tarpusavio integralumas, užtikrintas didesnis jų patogumas.

Straipsnyje siekiama pristatyti „E. kalbos“ išteklius, kuriuose randami leksikografiniai tarmių duomenys, aktualūs įvairių sričių kalbos tyrėjams ir plačiai visuomenei. Straipsnio **objektas** – vienakalbiai žodynai ir kartotekos, konkrečiai – „Lietuvių kalbos žodynas“ ir trys jo kartotekos[3]: LKŽ Pagrindinė kartoteka (LKŽ PK, <https://ekalba.lt/lkz-pagrindine-kartoteka/>), *LKŽ Papildymų kartoteka* (LKŽ PpK, <https://ekalba.lt/lkz-papildymu-kartoteka/>) ir LKŽ Tarmių kartoteka (LKŽ TK, <https://ekalba.lt/lkz-tarmiu-kartoteka/>) (plačiau apie kartotekų pobūdį ir jų struktūrą žr. Urbanavičienė 2021, 38–45; 2022, 29–53). LKŽ PK ir LKŽ PpK kartotekos jau buvo įtrauktos į LKIIS'o sudėtį, o LKŽ TK buvo sukurta LKIIS'o rengimo metu 2014 m., tačiau kuriant „E. kalbą“ jų paieškos sistema peržiūrėta, sukurta sąsajos su „Lietuvių kalbos žodynu“ bei kitais ištekliais. Straipsnio **tikslas** – pristatyti modernizuotą šių leksikografinių išteklių struktūrą, naują paieškos sistemą, taip pat aptarti sistemos kūrėjams iškilusias problemas ir jų sprendimo būdus.

E. žodynai	E. kartotekos ir duomenų bazės	E. paslaugos
Vienakalbiai žodynai <u>Bendrinės lietuvių kalbos žodynas</u> Dabartinės lietuvių kalbos žodynas <u>Lietuvių kalbos žodynas</u> <u>Lietuvių kalbos naujažodžių duomenynas</u> Sinonimų žodynas Antonimų žodynas Frazeologijos žodynas Palyginimų žodynas Sisteminis lietuvių kalbos žodynas	Žodynų kartotekos „Lietuvių kalbos žodyno“ Pagrindinė kartoteka „Lietuvių kalbos žodyno“ Papildymų kartoteka „Lietuvių kalbos žodyno“ Tarmių kartoteka <u>Kazimiero Būgos etimologinio žodyno kartoteka</u> Tautosakos paveldo kartotekos Pokario partizanų dainų kartoteka Mįslių kartoteka Liaudies tikėjimų kartoteka Duomenų bazės Lietuvos vietovardžių geoinformacinė duomenų bazė Istorinių vietovardžių duomenų bazė Pavardžių duomenų bazė Tarmių archyvas	Žodžių tinklai <u>Paieška žodžių prasmių tinkle</u> <u>E. sąvokos</u> E. rinkodara <u>E. pavadinimas</u> <u>Naudotojų užklausų analizė</u> <u>Nuomonių analizė</u> E. patarimai <u>Žodžių darybos vedlys</u> <u>Kalbos patarimai</u> E. mokymai ir kalbos žaidimai Nuotolinis lietuvių kalbos mokymas Mokomieji kalbos žaidimai

1 pav.

Nauji ištekliai ir paslaugos (pabraukti) lietuvių kalbos išteklių informacinėje sistemoje „E. kalba“

Lietuvių kalbos elektroninių išteklių prieinamumas sistemoje „E. kalba“

Kartu su naujai integruotais ištekliais šiuo metu „E. kalbos“ sistemą sudaro 33 ištekliai. „E. kalboje“ realizuota greitoji ir detalioji paieška tiek visoje sistemoje bendrai, tiek atskiruose ištekluose. Greitoji paieška galima visuose 33 sistemos ištekluose arba, paspaudus mygtuką „Rinktis išteklis“, galima paiešką susiaurinti iki vieno ar kelių aktualių šaltinių (žr. 2 pav.).

Vienakalbiai žodynai	Žodinių kartotekos	Žodžių tinklai
☐ Bendrinės lietuvių kalbos žodynas	☒ „Lietuvių kalbos žodyno“ Papildymų kartoteka	☐ Ontologijų sąvokos
☐ Dabartinės lietuvių kalbos žodynas	☒ „Lietuvių kalbos žodyno“ Tarmių kartoteka	E. patarimai
☒ Lietuvių kalbos žodynas	☒ „Lietuvių kalbos žodyno“ Pagrindinė kartoteka	☐ Kalbos patarimai
☐ Lietuvių kalbos naujažodžių duomenynas	☐ Kazimiero Būgos etimologinio žodyno kartoteka	☐ Žodžių darybos vedlys
☐ Sinonimų žodynas		
☐ Antonimų žodynas		
☐ Frazeologijos žodynas		
☐ Palyginimų žodynas		
☐ Sisteminis žodynas		
Dvikalbiai žodynai	Tautosakos paveldo kartotekos	
☐ Lietuvių–anglų kalbų žodynas	☐ Mįslių kartoteka	
☐ Anglų–lietuvių kalbų žodynas	☐ Liaudies tikėjimų kartoteka	
☐ Lietuvių–latvių kalbų žodynas	☐ Pokario partizanų dainų kartoteka	
☐ Latvių–lietuvių kalbų žodynas		
☐ Lietuvių–lenkų kalbų žodynas	Duomenų bazės	
☐ Lenkų–lietuvių kalbų žodynas	☐ Lietuvos vietovardžių geoinformacinė duomenų bazė	
☐ Lietuvių–vokiečių kalbų žodynas	☐ Istorinių vietovardžių duomenų bazė	
☐ Vokiečių–lietuvių kalbų žodynas	☐ Pavardžių duomenų bazė	
☐ Lotynų–lietuvių kalbų žodynas	☐ Tarmių archyvas	
☐ Senosios graikų–lietuvių kalbų žodynas		

Rinktis išteklius

2 pav.

„E. kalbos“ paieška: išteklių pasirinkimas

Galima ieškoti į paieškos laukelį įrašius visą žodį arba tik jo dalį – tuomet reikia naudoti žvaigždutę (*) arba klaustuką (?). Žvaigždutė fragmento pradžioje, viduryje arba pabaigoje reiškia, kad toje vietoje galima bet kokia arba jokia raidžių seka, pvz.: įvedus „šalt*is“ sistema išfiltruos žodžius *šaltanosis, šaltenis, šaltininis, šaltis, šaltmėtis...*, įvedus „*nešti“ – *atnešti, atsineštinukas, išnešti, neštinis, suneštinis* ir t. t. Klaustukas naudojamas, kai jo vietoje gali būti praleista bet kokia viena raidė, pvz.: surinkus „?aras“ randami žodžiai *baras, garas, karas, maras, šaras...*, surinkus „??žioti“ – *ažioti, apžioti, édžioti, išžioti, vežioti* ir kt.

Atsidarius konkretų išteklių paieška taip pat galima naudojant žvaigždutę (*) arba klaustuką (?). Galima ieškoti ir pagal pirmąją žodžio raidę, kurią reikia pasirinkti iš abėcėlės juostos – tuomet išfiltruojamas pasirinktos raidės žodžių sąrašas, surikiuotas pagal abėcėlę (žr. 3 pav.).

The screenshot shows a web interface for searching Lithuanian words. At the top, there is a navigation bar with letters A through Ž, with 'O' highlighted in a red box. Below this is a search bar labeled 'Paieška' with a search button 'ieškoti'. A checkbox option 'ieškoti visuose duomenyse' is present. Below the search bar, it says 'Paieškos rezultatai: 1-50 iš 876'. On the left, a list of words starting with 'o' is shown, with 'obaliauti' highlighted in a red box. On the right, the word 'obaliauti' is displayed with its inflected forms and a definition.

Paieška

ieškoti visuose duomenyse
+ Detalioji paieška

Paieškos rezultatai: 1-50 iš 876

o¹
o²
o³
o⁴
o žip
oa
oazė
obalas
obalė
obalėti
obaliauti
obalienė
obalynas

< > **obaliauti**

obaliauti: prisiobaliauti

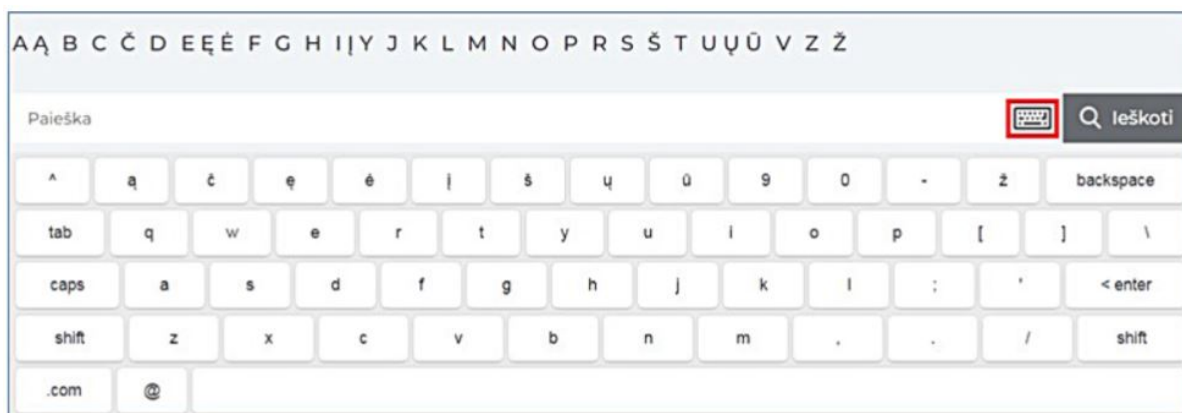
obaliáuti, -iáuja, -iávo žr. **obuoliauti** 1: Jis eina į girią obaliauti, t. y. rinkti obalius J.
^ Neik su velniu obaliáut, o su ponu – riešutaut Mžž.

prisiobaliáuti *prisirinkti obalių*: Prisiobaliávo vaikas pilnas kišenes obalių ir išėjo iš sodo Ėr.

3 pav.

Paieška pagal pirmąją antraštinio žodžio raidę ir abėcėlinis „O“ raidės žodžių sąrašas „Lietuvių kalbos žodyne“.

Kuriant paieškas buvo atsižvelgta į lietuvių kalbos vartotojų, gyvenančių užsienyje, poreikius bei numatyta galimybė raides su lietuviškais ar kitų kalbų diakritikais surinkti naudojant „E. kalboje“ įdiegtas klaviatūras, esančias šalia paieškos laukelių (žr. 4 pav.).



4 pav.

Klaviatūra su lietuviškomis raidėmis, išsiskleidžianti paspaudus klaviatūros piktogramą (pažymėta), skirta vartotojams, neturintiems lietuviškos klaviatūros savo kompiuteriuose.

Detalioji paieška buvo kuriama atsižvelgiant į kiekvieno ištekliaus duomenų savitumą, todėl skirtinguose žodynuose ir kartotekose realizuoti skirtingi paieškos kriterijai. Paspaudus mygtuką „Detalioji paieška“ galima kompleksinė paieška pagal kelis kriterijus. *LKŽ Tarmių kartotekoje* galima ieškoti žodžių pagal gramatines pažymas, kirčiuotę, kilmę, užrašymo vietą, metus ir užrašytoją, pvz.: galima surasti Adutiškio apylinkėse (Švenčionių r.) užrašytus 2 kirčiuotės slavizmus (žr. 5 pav.). Taip pat įmanoma paieška pagal

5 pav.

Detalioji paieška LKŽ Tarmių kartotekoje.

antraštinio žodžio kilmę, t. y. galima išfiltruoti kartotekoje saugomus slavų, germanų, baltarusių, lenkų, rusų ir kt. kalbų žodžius ar hibridus. Galima pasirinkti ieškoti pagal užrašytoją, laikotarpį, pvz.: *LKŽ Papildymų kartotekoje* galima surasti, kad kalbininkė Birutė Jasiūnaitė per 1990–2000 metus užrašė 5966 vnt. kortelių, iš jų 885 – iš Leliūnų (Utenos r.), 1168 – iš Endriejavo (Klaipėdos r.) ir t. t.

Kai kuriuose detaliosios paieškos laukeliuose įdiegtas daugybinis žymėjimas (žr. 6 pav.), t. y. vartotojas vienu metu gali atlikti paiešką pagal kelis kriterijus, pvz.: „Lietuvių kalbos žodyno“ paieškos laukelyje „Gramatinė pažyma“ galima vienu metu vykdyti paiešką pagal kalbos dalį, giminę, skaičių..., laukelyje „Stilistinė pažyma“ – pagal kelis stilistinius atspalvius (pvz., juokaujama ar ironišką), laukelyje „Tarmybė“ – ieškoti žodžių iš karto keliose patarmėse (pvz.: rytų, pietų ir vakarų aukštaičių plote).

Antraštinis/paantraštinis žodis 	Gramatinė pažyma 	Nenorminis ☐
Visa žodžio forma 	Reikšmės aiškinimas 	Homonimo numeris
Žodžio formos galūnė 	Stilistinė pažyma 	Nuorodos tipas
Žodžio variantas 	Tarmybė 	Susijęs įrašas

6 pav.

Daugybinio žymėjimo langeliai „Lietuvių kalbos žodyno“ detaliojoje paieškoje.

Detalioji paieška taip pat galima pagal reikšmės aiškinime pavartotus žodžius, pvz.: „Lietuvių kalbos žodyno“ detaliosios paieškos langelyje „Reikšmės aiškinimas“ įrašius žodį „patiekalas“ sistema išfiltruoja visus žodžius, savo definicijose turinčius žodį „patiekalas“: *apkepas, dubenėliai, muštinis, piršteliai, silkienė, troškinys, užkandis* ir t. t. Tokią paiešką galima dar labiau susiaurinti ir ieškoti tik tam tikros patiekalo rūšies, pvz.: „muštos mėsos patiekalas“, „patiekalas iš supjaustytų daržovių“, „saldus patiekalas“, „silkų patiekalas“, „skystas patiekalas“, „šaltas, sausas patiekalas“ ir t. t. Tokios „E. kalbos“ galimybės itin pravers leksikografams, siekiantiems sistemiškai apibūdinti tam tikrą žodžių grupę, ar tyrėjams, renkantiems empirinius duomenis.

Lietuvių kalbos elektroninių išteklių patogumas: duomenų peržiūra, atsisiuntimas ir dalinimasis

Kuriant svetainę „E. kalba“ siekta ją padaryti kuo intuityvesnę, patogesnę naudoti, atitinkančią tiek Lietuvoje, tiek už jos ribų esančių vartotojų lūkesčius. Visi „E. kalboje“ esantys lietuvių kalbos ištekliai pritaikyti naudoti tiek kompiuteriuose, tiek ir mobiliuosiuose įrenginiuose – išmaniuosiuose telefonuose, planšetiniuose kompiuteriuose. Yra sukurta galimybė popierinių kartotekų korteles padidinti (paspaudus lupos ikonėlę) per visą įrenginio ekraną, kad būtų patogiau perskaityti (žr. 7 pav.), o skaitmeninių kortelių duomenis galima kopijuoti. Kad nebūtų iškraipytas fonetine transkripcija surinktas tekstas, būtina naudoti lietuvišką šriftą *Palemonas*

(Aleknavičienė et al. 2005), kurį galima rasti ir „E. kalboje“: <https://ekalba.lt/EKALBA%20Palemonas>.

< >

akis (31)

f

d

y

Kartotekos kortelės vaizdai:

Kortelės numeris:

31

Antraštė

Žodis/Žodžių junginys:

akis

Paantraštė

Žodis/Žodžių junginys:

atsikrapšyti akis

Metrika

Vietovė kortelėje:

Antz

Vietovė:

Antazavė, Zarasų r.

Užrašytojas kortelėje:

Z. Šimėnaitė

Užrašytojas:

Z. Šimėnaitė

Užrašymo metai:

1996

7 pav.

LKŽ Pagrindinės ir LKŽ Papildymų kartotekų popierinių kortelių peržiūra

„Lietuvių kalbos žodynas“ yra parašytas lizdiniu būdu: visi priešdėliniai vediniai pateikiami pamatinių žodžių lizde, t. y. nepriešdėlinių veiksmažodžių straipsniuose pateikiami priešdėliniai veiksmažodžiai, būdvardžių straipsniuose – iš jų padarytirieveiksmiai ir t. t. Vieno LKŽ straipsnio apimtis svyruoja nuo vienos eilutės iki keliasdešimties ar beveik 100 puslapių, pvz.: veiksmažodžio *leisti* straipsnio apimtis – daugiau nei 5 aut. l., jame aprašytos 94 nepriešdėlinio veiksmažodžio reikšmės, taip pat straipsnyje pateikta keliasdešimt priešdėlinių vedinių postraipsnių: *antleisti, apleisti, atleisti, daleisti, įleisti, išleisti, nuleisti, paleisti, panuleisti, papaleisti, parleisti, perleisti, praleisti, prileisti, razleisti, suleisti, užleisti*. Vartotojų patogumui „E. kalboje“ ieškomi paantraštiniai žodžiai ir jų postraipsniai yra paryškinti, kad būtų nesunkiai randami ilgame žodyno straipsnyje. Ieškomos teksto dalies paryškinimas palengvina ir pagreitina naudojimąsi LKŽ duomenimis.

Rašant „Lietuvių kalbos žodyną“ buvo panaudota medžiaga iš daugiau kaip 1 tūkst. vietovių ir šaltinių, kurie žodyno tekste turi savo santrumpas, pvz.: *Br* – *Barzdai* (*Šakių r.*), *Dv* – *Dieveniškės* (*Šalčininkų r.*), *Mž* – *Mažvydas*, *S.Dauk* – *Simonas Daukantas* ir t. t.

Santrumpomis taip pat žymimos gramatinės ir stilistinės pažymos, pvz.: *adv.* – *prieveiksmis*, *sm.* – *vyrishkosios giminės daiktavardis*, *dem.* – *mažybinis* ir kt. Kad santrumpų ir specifinių ženklų gausa netrukdytų skaityti žodyno, kiekvieną santrumpą galima išskleisti užvedus ant jos kompiuterio pelę. Tokiu būdu paaiškinami ir skirtingais šriftais išskirti atskiri teksto segmentai, pvz.: „antraštinis žodis“, „nuoroda į kitą žodį“, „homonimo numeris“, „žodžio reikšmės numeris“ ir kt. (žr. 8 pav.).



8 pav.

„Lietuvių kalbos žodyno“ išsiskleidžiančios santrumpos



9 pav.

Galimybė atsisiųsti LKŽ Tarmių kartotekos duomenis PDF, XML formatais ar dalintis jais Facebook tinkle (pažymėtos atitinkamos piktogramos).

Vartotojų patogumui žodynų ir kartotekų duomenis iš „E. kalbos“ galima atsisiųsti PDF, XML formatais, formatais, taip pat yra

numatyta galimybė jais dalintis socialiniame tinkle *Facebook* (žr. 9 pav.).

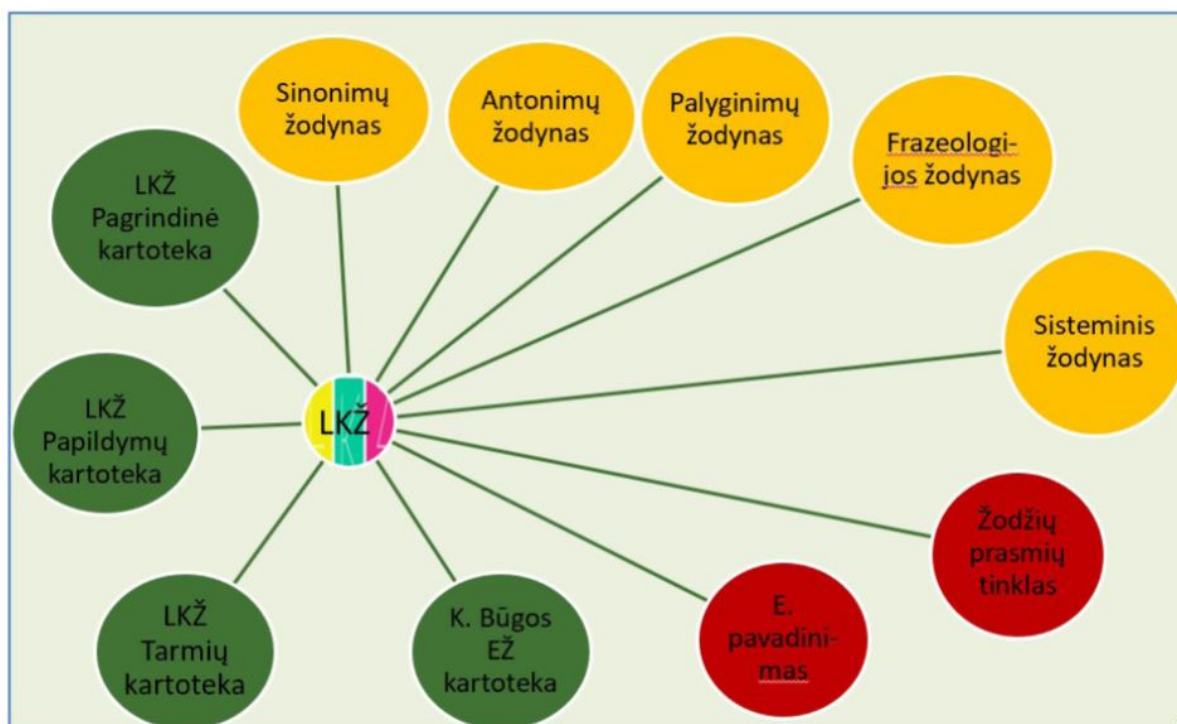
Lietuvių kalbos elektroninių išteklių integralumas

I bendrą „E. kalbos“ struktūrą integravus didįjį *Lietuvių kalbos žodyną*, jo išteklius galima susieti su LKŽ kartotekų duomenimis, taigi lietuvių kalbos leksikos ištekliai randami vienoje vietoje. Pvz., LKŽ žodžio *abėcė* („abėcėlė, skaitymas“) straipsnį papildė *LKŽ Papildymų kartotekoje* saugomas kitas kirčiavimo variantas *ābēcė* (1 kirčiuotė) bei *LKŽ Tarmių kartotekoje* užfiksuotas naujas frazeologizmas *kaip abėcė* („labai gerai, lengvai. (ppr. apie mokymąsi)“) (žr. 1 lentelę).

1 lentelė.

„Lietuvių kalbos žodyno“ straipsnių papildymas LKŽ kartotekų medžiaga: žodžio *abėcė* atvejis

Lietuvių kalbos žodynas	LKŽ Papildymų kartoteka	LKŽ Tarmių kartoteka
abėcė sf. ppr. ind. KI48, Gs, ābēcė (1) <i>abėcėlė, skaitymas</i> : Da abėcė neišmoko, o jau nosį riečia Lkč. Neišmokė ābēcės, būk visą amžių akla Lnkv. ◇ nei abėcė <i>nieko (nemokėti)</i> : Jis nei abėcė nemoka Trg.	ābēcė <i>abėcėlė</i> : Jau išmokau visą ābēcę RdN (Nemunėlio Radviliškis, Biržų r.).	abėcė ◇ kaip abėcė <i>labai gerai, lengvai (ppr. apie mokymąsi)</i> : man_ta_vo.kiečū· kalbà_tai ė·jū kaip_a.bė·cė· Pgg (Pagė- giai, Šilutės r.).



10 pav.

„Lietuvių kalbos žodyno“ sąsajos su kitais žodynais, kartotekomis ir e. paslaugomis informacinėje sistemoje „E. kalba“

Tokiu moderniu būdu „E. kalbos“ sistemoje išspręstas „Lietuvių kalbos žodyno“ papildymų rengimo klausimas, lietuvių leksikografų keltas XXI a. pradžioje (Naktinienė, Černiauskas 2000, 68–69). Informacinėje sistemoje „E. kalba“ realizuota ir „Lietuvių kalbos žodyno“ sąsaja su kitais žodynais („Sinonimų žodynu“, „Antonimų žodynu“, „Frazeologijos žodynu“, „Palyginimų žodynu“, „Sisteminis lietuvių kalbos žodynas“), taip pat su „Lietuvių kalbos žodyno“ kartotekomis (*LKŽ Pagrindinė*, *LKŽ Papildymų* ir *LKŽ Tarmių kartotekomis*) bei e. paslaugomis – *Žodžių prasmų tinklu* ir *E. pavadinimu* (žr. 10 pav.).

Tokiu būdu sumažinta lietuvių kalbos integruotų informacinių išteklių stoka, nepakankamas išteklių tarpusavio integralumas, skaitmeninių lietuvių kalbos išteklių fragmentacija, tarpusavyje integruotų lietuvių kalbos išteklių trūkumas. Vartotojai gali paprastai ir greitai rasti dominančius lietuvių kalbos išteklius vienoje vietoje bei pasinaudoti jų pagrindu sukurtais e. paslaugomis.

Informacinės sistemos „E. kalba“ kūrimo iššūkiai ir jų sprendimo būdai

„E. kalba“, kaip ir Lietuvių kalbos išteklių informacinė sistema (LKIIS), buvo kurta pagal duomenų bazės principus, tačiau dalį išteklių galima priskirti elektroniniams duomenų rinkiniams, kuriuos

integruojant į „E. kalbą“ kilo duomenų atvaizdavimo problemų. Ypač tai pasakytina apie transkribuotų iliustracinių sakinių perkėlimą į informacinę sistemą. Keletą probleminių atvejų ir jų sprendimo būdų aptarsime žemiau.

1. Nekorektiškas balsių ilgumo atvaizdavimas išryškino lietuviško šrifto „Palemonas“ sisteminį trūkumą, kai ilgumą žymintis taškas balsės dešinėje vaizduojamas ne ties viršutine raidės riba, bet ties raidės viduriu (žr. 11 pav.).

išmerá.va žė.mi | bjáu.r^ṑ bú.ua | kulī.nė | akmeni^ī.ne bú.ua |
ėškeř.ta té.te | amžená.tėîš ||

11 pav.

Nekorektiškas balsių ilgumo atvaizdavimas.

Šiai problemai pašalinti buvo inicijuota diskusija su „Palemono“ kūrėjais ir informacinių technologijų specialistais: Vladu Tumasoniu, Petru Skirmantu, Mindaugu Strockiu. 2020 m. lituanistikos mokslo bendruomenei buvo pristatyta atnaujinta „Palemono“ šrifto versija 3.2.05. Tobulinant „Palemono“ ženklų aibę atlikti du svarbūs pakeitimai: 1) atskirtos kompozicinės sekos nuo ligatūrų; 2) „Open Type“ keitiniai papildyti alternatyvomis (žr.: <http://www.vlkk.lt/palemonas>, 2020-06-02). Vartotojams siūloma įsodiegti naująją šrifto versiją, kuri balsių ilgumus atvaizduoja korektiškai, plg.: išmerFI.va žĖAi.mi | bjā.rṑ b.a | kulĀ.n | akmenĀ.ne b.a | šķēḂ.ta tĀi.te | amženFI.tė || (žodis „akmenynė“, LKŽ TK, [https://ekalba.lt/lkz-tarmiu-kartoteka/akmenyn%C4%97%20\(1\)?paieska=akmenyn%C4%97&i=c164a643-0bef-4a47-b291-521254c7d55a](https://ekalba.lt/lkz-tarmiu-kartoteka/akmenyn%C4%97%20(1)?paieska=akmenyn%C4%97&i=c164a643-0bef-4a47-b291-521254c7d55a)).

2. Nekorektiškas balsių redukcijos atvaizdavimas susijęs su netinkamu ženklų pasirinkimu iš „Palemono“ repertuaro: redukuotiems balsiams pavaizduoti vartotojai renka ne diakritinius ženklus, o naudoja programos Microsoft Word viršutinio indekso funkciją, kuri duomenų bazės terpėje neišlieka, t. y. redukcija neatspindima. Plg.: vietoj sakinio a_{ku}ḡk graž tFI.vuḡ. vuḡ.k.l. | kap_{uv}Ā-dḡu vaicævā r.rištā aš_{tv}art. | nāt_{nu}vejḡ. padabúoā || gali būti matomas sakiny: a_{ku}ḡk graž tFI.vuḡ. vuḡ.k.l. | kap_{uv}Ā-dḡu vaicævā r.rištā aš_{tv}art. | nāt_{nu}vejḡ. padabúoā || (žodis „ožkelė“, LKŽ TK, [https://ekalba.lt/lkz-tarmiu-kartoteka/o%C5%BElkel%C4%97%20\(2\)?paieska=o%C5%BElkel%C4%97&i=8f3af7fb-3395-44cd-ac1f-bc99ebf0dad5](https://ekalba.lt/lkz-tarmiu-kartoteka/o%C5%BElkel%C4%97%20(2)?paieska=o%C5%BElkel%C4%97&i=8f3af7fb-3395-44cd-ac1f-bc99ebf0dad5)). Ši problema išryškino poreikį turėti klaviatūros tvarkyklę, kuri palengvintų ir pagreitintų naudojimąsi „Palemono“ šriftu, kuriame yra apie 4 600 rašmenų[4] (Aleknavičienė et al. 2005,

16). 2020 m. doc. dr. Mindaugas Strockis sukūrė fonetinę klaviatūrą „Kopenhagos transkripcija“, kuri leidžia tyrėjams rinkti transkribuotus lietuvių dialektologijos tekstus naudojant „Palemono“ simbolius ir įprastą klaviatūrą (Strockis 2020). Taigi „E. kalbos“ rengimo darbai ir jų metu iškilusios problemos paskatino kitų elektroninių produktų ir paslaugų plėtrą.

Baigiamosios pastabos

„E. kalbos“ sukūrimas kiekybiškai ir kokybiškai pagerino 2015 m. sukurta Lietuvių kalbos išteklių informacinę sistemą (LKIIS):

1. kiekybinis lygmuo – į informacinę sistemą buvo integruoti 3 žodynai, 1 kartoteka ir sukurtos 7 paslaugos;
2. kokybinis lygmuo - pagerino lietuvių kalbos išteklių prieinamumą, patogumą ir tarpusavio integralumą;
3. pagreitino kitų elektroninių produktų ir paslaugų plėtrą: prisidėjo sprendžiant sisteminės lietuviško šrifto „Palemonas“ problemas, paskatino klaviatūros tvarkyklės „Kopenhagos transkripcija“ atsiradimą ir kt.

Ateities uždaviniai susiję ne tik su naujų išteklių integracija bei inovatyvių el. produktų ar paslaugų kūrimu, bet ir į „E. kalbos“ sistemą jau integruotų išteklių pildymu. Pvz., tik nedidelė dalis „Lietuvių kalbos žodyno“ kartotekų yra suskaitmenintos, kita jų dalis saugoma Lietuvių kalbos instituto archyvuose popieriniu pavidalu^[5]. Suskaitmenintos LKŽ kartotekos taptų pasiekiamos akademinei bendruomenei bei visiems lietuvių kalbos vartotojams, jų skaitmeninimas palengvintų naudojimąsi lietuvių tarmių leksikografiniais ištekliais. Taigi ateityje siekiant padidinti lietuvių kalbos plėtrą skaitmeninėje terpėje bei patenkinti kalbos vartotojų poreikius lietuvių kalbos išteklių informacinės sistemos kūrėjų laukia nauji iššūkiai.

Literatūra ir šaltiniai

- Aleknavičienė, Ona, Laima Grumadienė, Albertas Gurskas, Petras Skirmantas, Mindaugas Strockis, Vladas Tumasonis. 2005. *Lituanistinis šriftas* Palemonas. Vilnius: Tautinių bendrijų namai.
- BEŽK – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“: Kazimiero Būgos etimologinio žodyno kartoteka*. Prieiga internetu: <https://ekalba.lt/BugosEtimologinioZodynoKartoteka/>.
- BLKŽ – Liutkevičienė, Danutė, vyr. red. *Bendrines lietuvių kalbos žodynas*. Prieiga internetu: <https://ekalba.lt/bendrines-lietuviu-kalbos-zodynas/>.
- E. kalba – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“*. Prieiga internetu: <https://ekalba.lt/>.
- EP – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“: E. pavadinimas*. Prieiga internetu: <https://ekalba.lt/epavadinimas/>.
- ES – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“: E. sąvokos*. Prieiga internetu: <https://ekalba.lt/esavokos/>.
- KP – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“: Kalbos patarimai*. Prieiga internetu: <https://ekalba.lt/kalbos-patarimai/>.
- LKIIS – *Lietuvių kalbos išteklų informacinė sistema*. Prieiga internetu: <https://ekalba.lt/EKALBA%20Apie%20projekt%C4%85>.
- LKND – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“: Lietuvių kalbos naujažodžių duomenynas*. Prieiga internetu: <https://ekalba.lt/naujazodziai/>.
- LKŽ – Naktinienė, Gertrūda, vyr. red. 2018. *Lietuvių kalbos žodynas (1–20, 1941–2002)*. Elektroninis variantas, atnaujinta versija. Vilnius: Lietuvių kalbos institutas. Prieiga internetu: <https://ekalba.lt/lietuviu-kalbos-zodynas/>.
- LKŽ PK – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“: Lietuvių kalbos žodyno Pagrindinė kartoteka*. Prieiga internetu: <https://ekalba.lt/lkz-pagrindine-kartoteka/>.
- LKŽ PpK – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“: Lietuvių kalbos žodyno Papildymų kartoteka*. Prieiga internetu: <https://ekalba.lt/lkz-papildymu-kartoteka/>.
- LKŽ TK – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“: Lietuvių kalbos žodyno Tarmių kartoteka*. Prieiga internetu: <https://ekalba.lt/lkz-tarmiu-kartoteka/>.
- NA – *Lietuvių kalbos išteklų informacinė sistema „E. kalba“: Nuomonių analizė*. Prieiga internetu: <https://ekalba.lt/nuomoniui-analize/>.

- Naktinienė, Gertrūda, Valentinas Černiauskas. 2000. Leksikografinės duomenų bazės kūrimas: patirtis ir perspektyvos. *Tarmės: tekstų ir žodynų rengimo problemos*. Šiauliai: Šiaulių universitetas, 66–77.
- NUA – *Lietuvių kalbos išteklių informacinė sistema „E. kalba“: Naudotojų užklausų analizė*. Prieiga internetu: <https://ekalba.lt/naudotoju-uzklausu-analize/>.
- Palemonas *Lietuvių kalbos išteklių informacinė sistema „E. kalba“: Palemonas*. Prieiga internetu: <https://ekalba.lt/EKALBA%20Palemonas>.
- Strockis, Mindaugas. 2020. *Fonetinė klaviatūra „Kopenhagos transkripcija“*. Vilnius: Lietuvių kalbos institutas. Prieiga internetu: <http://lki.lt/fonetine-klaviatura/>.
- Urbanavičienė, Jolita. 2021. *Tarmių leksikos rinkimas Lietuvių kalbos žodyno kartotekoms. Visos tarmės gražiausios... Tarmių medžiagos rinkimo instrukcija. 2-asis pataisytas ir papildytas leidimas*. Vilnius: Mokslo ir enciklopedijų leidybos centras, 38–45.
- Urbanavičienė, Jolita. 2022. *Lietuvių kalbos žodyno Tarmių kartoteka – elektroninis duomenynas lietuvių tarmių ir kitų kalbų kontaktams tirti. Vakarų ir pietų aukštaičiai: tarmių ir kitų kalbų sąveika*. Vilnius: Lietuvių kalbos institutas, 29–53.
- VLKK – *Valstybinė lietuvių kalbos komisija*. Prieiga internetu: <https://www.vlkk.lt/>.
- ŽDV – *Lietuvių kalbos išteklių informacinė sistema „E. kalba“: Žodžių darybos vedlys*. Prieiga internetu: <https://ekalba.lt/zodziu-darybos-vedlys/>.
- ŽPT – *Lietuvių kalbos išteklių informacinė sistema „E. kalba“: Žodžių prasmų tinklas*. Prieiga internetu: <https://ekalba.lt/zodziu-prasmiu-tinklas/?p=1>.

undefined

- 1 2012–2015 m. projektas finansuotas pagal Ekonomikos augimo veiksmų programos 3 prioriteto „Informacinė visuomenė visiems“ įgyvendinimo priemonę Nr. VP2-3.1-IVPK-12-K „Lietuvių kalba informacinėje visuomenėje“ (kodas NR. VP2-3.1-IVPK-12-K-01-005).
- 2 2018–2021 m. projektas „Lietuvių kalbos išteklių informacinės sistemos plėtra (E. kalba)“ (Nr. 02.3.1-CVPAV-527-01-0006) finansuotas iš Europos Sąjungos struktūrinių fondų lėšų.
- 3 Minėtų išteklių integravimo, adaptavimo ir patikros darbus projekte atliko Lietuvių kalbos instituto mokslininkės dr. Aurelija Gritėnienė ir dr. Jolita Urbanavičienė bei vyriausiasis programuotojas Evaldas

Ožeraitis, bendradarbiaudami su pažangiųjų informacinių technologijų kūrėjais – UAB „Proit“.

- 4 „Palemono“ ženklų aibė susideda iš kelių poaibių: 1) įprastinių raidžių; 2) lietuviškų kirčiuotų raidžių; 3) lituanistinių rašmenų; 4) graikų abėcėlės; 5) kirilikos; 6) skaitmenų; 7) simbolių. Didžiausią aibės dalį sudaro lietuvių filologijoje vartojami specialieji rašmenys: 1) tarmių fonetinės transkripcijos; 2) senųjų raštų; 3) baltistikos ir indoeuropeistikos“ (Aleksavičienė et al. 2005, 15-16).
- 5 Šiuo metu „E. kalboje“ skelbiami LKŽ Pagrindinės kartotekos dalies A ir G raidžių kortelių vaizdai ir jų metaduomenys (žr. <https://ekalba.lt/lkz-pagrindine-kartoteka/apie/Pagr.%20kart.%20Apie%20kartotek%C4%85>), taip pat didžioji dalis LKŽ Papildymų kartotekos (išskyrus O, Š ir Ž raides) (žr. <https://ekalba.lt/lkz-papildymu-kartoteka/apie/Papild.%20kart.%20Apie%20kartotek%C4%85>). Likusi dalis – apie 4 mln. kortelių - dar laukia skaitmeninimo.



Disponible en:

<https://www.redalyc.org/articulo.oa?id=692574304009>

Cómo citar el artículo

Número completo

Más información del artículo

Página de la revista en redalyc.org

Sistema de Información Científica Redalyc
Red de revistas científicas de Acceso Abierto diamante
Infraestructura abierta no comercial propiedad de la
academia

Jolita Urbanavičienė

**Leksikografiniai tarmių duomenys lietuvių kalbos išteklių
informacinėje sistemoje „E. kalba“**

Lexicographic Data on Dialects in the Lithuanian Language
Resources Information System E-Language

Lietuvių kalba

núm. 17, p. 109 - 121, 2022

Vilniaus Universitetas, Lituania

irena.smetoniene@flf.vu.lt

/ ISSN-E: 1822-525X

DOI: <https://doi.org/10.15388/LK.2022.8>



CC BY 4.0 LEGAL CODE

Licencia Creative Commons Atribución 4.0 Internacional.