

Knygos nobažnystės (1653)
morfologijos tyrimo galimybės
naudojant reliacinę duomenų bazę

DALIA JAKULYTĖ

Klaipėdos universitetas

The article presents the relational database created for the investigation of the morphology of old Lithuanian writings. A logical data model was developed and implemented within the Microsoft Access database management system (DBMS). All the words of two parts of the Knyga Nobažnystės were entered into the database. An engine was created for morphological (inflectional) data selection, so that the words can now be selected by any assigned property or set of properties. Such a database serves as a tool for data sorting, selection and correction, as well as for text and index media storage. It allows the researcher to deal with various linguistic problems, such as noun stem alternation, authorship of texts etc. in an efficient, quick and accurate manner.

0. ĮVADAS

Viena didžiausių problemų, su kuria susiduria kalbininkai, tyrinėjantys senuosius raštus, tai medžiagos atranka. Pagrindiniai medžiagos šaltiniai yra patys tekstai (ppr. knygų fotokopijos), žodynai, žodžių registrai. Iš jų duomenys yra perrašinėjami (ranka ar kompiuteriu), sudaromos popierinės kartotekos, jos rūšiuojamos ir rikiuojamos tam tikra tvarka. Medžiagos atranka iš tokios kartotekos – ilgas ir gana sudėtingas darbas: peržiūrima visa kartoteka, atrenkami reikiami lapeliai, jie suskirstomi ir t. t.; pabaigus darbą atrinkti lapeliai turi būti gražinami į jų vietas bendroje kartotekoje, priešingu atveju ji nebus tinkama naudoti.

Tokias problemas padeda spręsti kompiuterinės technologijos. Leksikografiniams tyrimams sudaromi kompiuteriniai tekstynai, senųjų raštų duomenų bazės, etc.

Vienas esminių kompiuterinių tekstynų požymių – galimybė atlikti automatinę teksto analizę, pvz., automatiškai daryti teksto konkordansus, t. y. pateikti ieškomą žodį minimaliame kontekste. Automatinei kalbos analizei tekstai turi būti sužymėti, kad konkordansų programa ieškotų kalbos vienetų ne tik pagal formalius, bet ir pagal gramatinius požymius. (Marcinkevičienė 1997). Tačiau tik norminei kalbai galima kurti ir taikyti automatines kalbos gramatinių formų ar pan. atpažinimo programas, o senųjų raštų gramatiniai požymiai turi būti sužymėti rankiniu būdu. Pvz., Vytauto Didžiojo universiteto Humanitarinių mokslų fakulteto Kompiuterinės lingvistikos centre parengtą 100 milijonų žodžių apimties lietuvių kalbos tekstyną (donelaitis.vdu.lt/tekstynas) sudaro nepriklausomybės laikotarpio leidiniai, parinkti taip, kad kuo ge-

riau atspindėtų dabartinę rašytinę lietuvių kalbą. Tekstyno paieškos sistema leidžia atlikti žodžio paiešką, sužinoti pasirinkto žodžio statistinę informaciją, ieškoti pasirinkto žodžio konkrečia gramatine forma (pvz., *neprisiklausomybės*), arba visų žodžio formų naudojant simbolių „*“ (*neprisiklausomyb**). Latvijos universiteto Matematikos ir informatikos institutas internete pateikė ir senųjų XVI–XVIII a. rašto paminklų tekstus su panašiomis paieškos galimybėmis (galima ieškoti konkretaus žodžio ar jo dalies su simboliu „%“) bei archyvuotas žodžių formų dažnumų žodyno ir indekso rinkmenas (www.ailab.lv/senie). Panašus darbas yra atliekamas ir Lietuvių kalbos institute – kompiuteriu renkami ar nuskaitomi senųjų raštų tekstai ir sudaromi žodžių formų indeksai. Jie sudaromi pusiau rankiniu būdu, t. y. kompiuteris pagal specialią programą tik išdėsto žodžių formas abėcėlės tvarka, nurodydamas jų vietą tekste ir pateikdamas kontekstą. Tekstologinė informacija bei gramatinė ir leksinė konkretos žodžio formos reikšmė įrašoma ranka (Ambrasas 1998).

Senųjų raštų morfologijai tirti tradiciškai naudojamos tokios pat kartotekos kaip ir žodynams, bet skirtingas jų panaudojimo mechanizmas, pavyzdžiui, rūšiuojama ne tik pagal leksemas, bet ir pagal kitus požymius (kalbos dalis, morfemas, darybą, formą ir t. t.). Tad atitinkamai turi būti pritaikyta ir kompiuterinės duomenų bazės struktūra bei medžiagos atrankos galimybės.

Senųjų raštų morfologijos duomenų bazės kūrimo idėja kilo tęsiant Antano Jakulio pradėtus *Knygos nobažnystės* (1653)¹ tyrimus. Iki 2004 m., kai pasirodė faksimilinis *Knygos Nobažnystės* leidimas (Pociūtė, reng., 2004), Lietuvos kalbininkams buvo prieinamas tik 1968 m. darytas mikrofilmas (saugomas Mokslų akademijos Centrinėje bibliotekoje). Klaipėdos universiteto Baltistikos centras turėjo tais pačiais metais A. Jakulio darytas *Knygos Nobažnystės* (*Sumos evangelijų* ir *Maldu Krikščioniškų*) fotokopijas, Menų fakulteto biblioteka – giesmyno kserokopijas. Iš jų A. Jakulis sudarė *Knygos Nobažnystės* leksikos registrą, kurio publikuota tik viena dalis – *Sumos evangelijų* leksika (Jakulis 1995). Šiam leksikos registru buvo sudaryta popierinė visos leksikos kartoteka, t. y. į atskirus lapelius surašyti žodžiai, pažymint jų pagrindines formas, reikšmę, gramatinę formą (1 pav.), tie lapeliai sugrupuoti pagal pagrindines formas ir sudėti abėcėlės tvarka. Tačiau sudarytojas toliau tirdamas *Knygos Nobažnystės* kalbą pats naudojosi kartoteka ir lapelius perrūšiojo jam vienam žinoma tvarka. Iš esmės ją reikėjo sudaryti iš naujo, bet norėjosi padaryti tinkamą daug kartų naudoti. Tad buvo nutarta kurti elektroninę kartoteką – duomenų bazę, kurioje būtų kaupiami, rūšiuojami ar filtruojami įvairūs morfologiniai, leksiniai, semantiniai ar kt. duomenys ir kuri tarnautų kaip senųjų raštų morfologijos tyrėjo įrankis.

Tokios duomenų bazės kūrimą sudaro du pagrindiniai etapai – projektavimo ir realizacijos. Projektuojant sprendžiama, kokio pobūdžio duomenys reikalingi senųjų raštų morfologijos tyrėjui, apibrėžiama tų duomenų specifika, požymiai, tarpusavio ryšiai, numatoma, į kokius klausimus turi atsakyti kuriama duomenų bazė. Šio etapo rezultatas – loginė duomenų struktūra. Ji aprašyta šio straipsnio 2.1. skyriuje. Antra-

¹ Plačiau apie *Knygą Nobažnystės* ir jos istoriją žr. Dainoros Pociūtės (2004), Ingės Lukšaitės (2001), Zigmo Zinkevičiaus (1988: 213–223), Juozo Tumelio (1967, 1968) ir kt. darbuose.

sis etapas – sudarytos duomenų struktūros realizacija vienoje iš duomenų bazių valdymo sistemų – aprašytas 2.2. skyriuje.

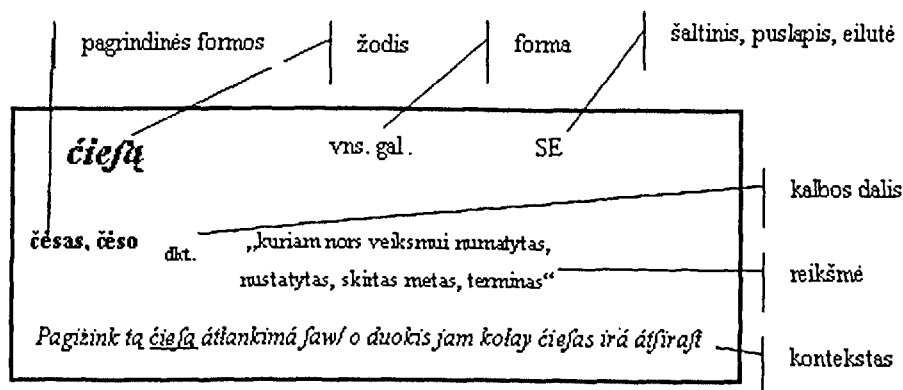
Sukurtą ir realizuotą senųjų raštų morfologijos duomenų bazę jau galima naudoti tyrimams, t. y., įkėlus duomenis ir sužymėjus reikiamus jų požymius, tuos duomenis atrinkti, tvarkyti ir rūšiuoti norima tvarka. Antrajame straipsnio skyriuje aprašomas sukurtos senųjų raštų morfologijos duomenų bazės taikymas konkretaus šaltinio – *Knygos nobažnystės* – morfologijos tyrimams.

1. SENŪJŲ RAŠTŲ MORFOLOGIJOS DUOMENŲ BAZĖS KŪRIMAS

1.1. Loginė duomenų struktūra

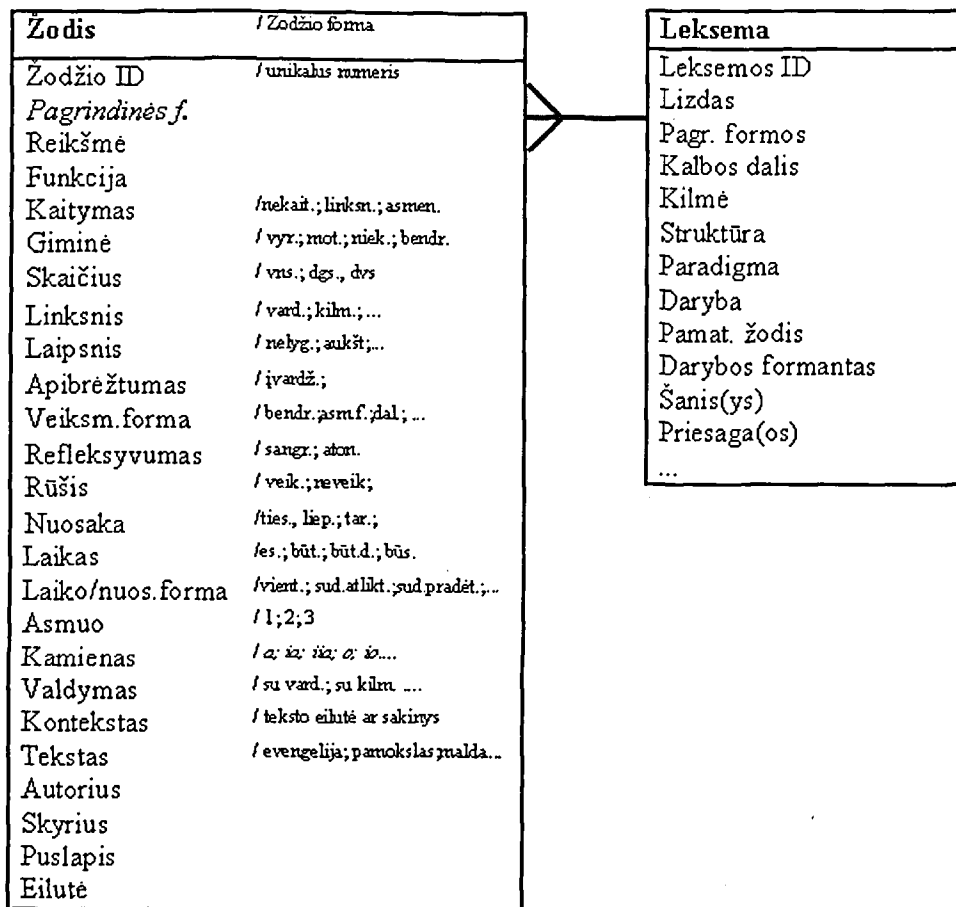
Pagrindiniai leksikos registro kartotekos elementai yra žodžių, pavartotų tekste, formos (čia ir toliau kiekviena žodžio forma, t. y. kiekvienas bet kokios leksemos pavartojimo atvejis bus vadinamas „žodžiu“). Viena kortelė – vienas žodis ir jo požymiai (atributai) (žr. 1 pav.).

1 PAV. KNYGOS NOBAŽNYSTĖS LEKSIKOS KARTOTEKOS KORTELĖ



Kiekvienas žodis yra unikalus. Net jei tas pats žodis tokia pat forma ir reikšme pavartotas kelis kartus, kiekvieną kartą jis bus kitoje vietoje, kitame kontekste; jei kortelės būtų sunumeruotos, tokios pačios žodžių formos turėtų skirtingus numerus.

Tad ir pagrindinis duomenų bazės elementas turi būti **žodis**, turintis unikalų numerį. Jo atributai – forma, šaltinis, pagrindinės formos, reikšmė etc. Kai kurie atributai (pvz., kalbos dalis, paradigma) yra bendri visai leksemai. Leksemos atributams apibrėžti kuriamas elementas „Leksema“, susietas su elementu „Žodis“ santykiu „vienas su daugeliu“, t. y. vieną „Leksemos“ įrašą gali atitikti keli „Žodžio“ įrašai. Apibendrinta loginė senųjų raštų morfologijos duomenų struktūra pavaizduota 2 pav.

2 PAV. LOGINĖ DUOMENŲ STRUKTŪRA²

Plačiau apie morfologijos duomenų specifiką, numatytus elementus, jų atributus ir tarpusavio ryšius aprašyta straipsnyje apie *Knygos nobažnystės* morfologijos duomenų bazės kūrimą (Jakulytė 2001).

1.2. Morfologijos duomenų bazės realizacija

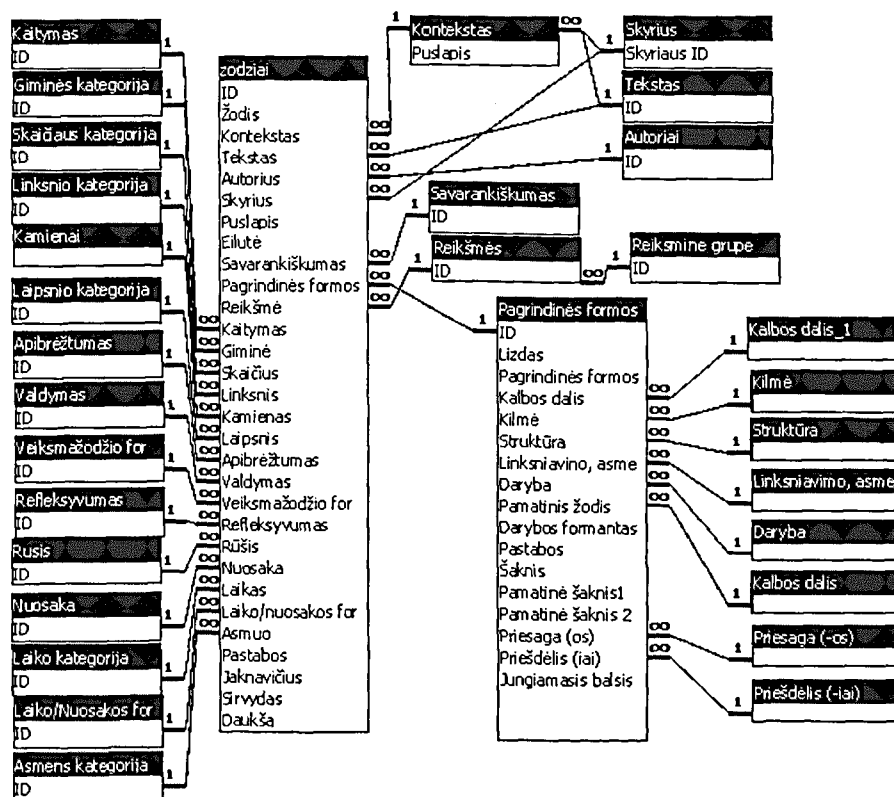
Elektroninis kartotekos atitikmuo – duomenų bazė. Kaupiamų duomenų apdorojimui, peržiūrai, atrankai, spausdinimui bei naujų duomenų įvedimui naudojama specializuota programinė įranga – duomenų bazių valdymo sistemos (DBVS). Dabartiniu metu plačiausiai taikomos reliacinės duomenų bazių valdymo sistemos (pvz., *Visual*

² Duomenų bazėje saugomos informacijos elementas – esybė – loginėje duomenų struktūroje žymimas stačiakampiu. Viršuje nurodomas esybės pavadinimas, apatinėje dalyje išvardijamas esybės savybės (atributai). Ryšys „vienas su daugeliu“ vaizduojamas išsišakojančia tiese.

FoxPro, dBase, Paradox ir kt). Į Microsoft Office programų paketą įtraukta DBVS MS Access pasižymi visomis reliacinio tipo DBVS būdingomis savybėmis ir gera integracija su kitomis šio paketo programomis. Joje duomenys saugomi lentelėse. Kiekviena pagrindinės lentelės cilutė – atskiras įrašas (pagrindinis elementas ir jo atributai). Pasikartojančių elementų (atributų) sąrašai saugomi pagalbinėse lentelėse. Į pagrindinę lentelę elementai įrašomi, o atributai parenkami iš pagalbinių lentelių. Susiejus lenteles galima daryti įvairias užklausas, leidžiančias atrinkti informaciją iš vienos ar kelių lentelių (t. y. pasirinkti reikiamas korteles iš kartotekos ir jas susidėti norima tvarka).

Remiantis sudaryta logine duomenų struktūra senųjų raštų morfologijos duomenų bazė buvo realizuota naudojant Microsoft Access 2003 DBVS. Čia sukurtos dvi pagrindinės lentelės: „Žodžiai“ (žodžių formoms) ir „Pagrindinės formos“ (leksemoms)³. Pagalbinėse lentelėse – atributų sąrašai. Visos pagalbinės lentelės su pagrindinėmis susietos ryšiu „vienas su daugeliu“. Senųjų raštų morfologijos duomenų bazės lentelės ir ryšiai tarp jų pavaizduoti 3 pav.

3 PAV. DUOMENŲ BAZĖS LENTELIŲ IR RYŠIŲ TARP JŲ DBVS MS ACCESS SCHEMA



³ Čia kol kas palikti pirminiai lentelių pavadinimai. Jie, ko gero, keičiami nebus – pavadinimo keitimas gana sudėtingas, o vartotojas jo paprastai nemato.

Vienos pagalbinės lentelės jau užpildytos, kitos – pritaikomos kiekvienam rašto paminklui. Užpildytos yra gramatinių kategorijų lentelės (kurios pavaizduotos 3 pav. kairėje). Kiekvieną jų sudaro kelios grafos: identifikatorius, atributo pavadinimas (-ai) ir pastabos. Į pagrindinę lentelę iš pagalbinės įrašomas tik identifikatorius, bet duomenų atrankoje galima naudoti bet kurios grafos duomenis, todėl vartotojų patogumui kai kurie atributai apibrėžiami tiek lietuviškais, tiek tarptautiniais terminais (plg. linksnio kategorijos sąrašą 1 lentelėje). „Konteksto“, „Skyriaus“, „Teksto“ ir „Autorių“ lentelės pritaikomos kiekvienam rašto paminklui. „Pagrindinių formų“ (leksemų) ir „Reikšmių“ lentelės pradžioje yra tuščios, bet duomenys gali būti įkelti iš kitam rašto paminklui sukurtos duomenų bazės.

1 LENTELĖ. LINKSNIO KATEGORIJOS ATRIBUTŲ SĄRAŠAS

Linksnio kategorija			
ID	Linksnis	Lotyniškai	Pastabos
0			-
1	vard.	N.	Vardininkas
2	kilm.	G.	Kilmininkas
3	naud.	D.	Naudininkas
4	gal.	A.	Galininkas
5	įnag.	I.	Įnagininkas
6	ines.	In.	Inesyvas
7	iliat.	Il.	Iliatyvas
8	adcs.	Ad.	Adesyvas
9	aliat.	Al.	Aliatyvas
10	šauksm.	V.	Šauksmininkas
21	sav. kilm.	G. posses.	Savybinis kilmininkas

2. DUOMENŲ BAZĖS TAIKYMAS KNYGOS NOBAŽNYSTĖS MORFOLOGIJOS TYRIMAMS

2.1. Duomenų įvedimas

Prieš įvedant konkretaus rašto paminklo morfologijos duomenis, duomenų bazė pritaikoma. Pavyzdžiui, tiriant *KN* kalbą reikia atsižvelgti į teksto struktūrą ir autorystę (plačiau žr. Jakulis 1982, 1984). Skirtingos knygos dalys parengtos ar verstos skirtingų autorių, tad lentelėje „Tekstas“ surašomi teksto tipai (evangelija, pamokslas, malda, giesmė...), lentelėje „Autoriai“ – surašomi žinomi *KN* autoriai. Užpildoma ir lentelė „Skyrius“ – surašomi visi *KN* skyriai (žr. 6 pav.).

Duomenų įvedimas prasideda nuo teksto rinkimo. Surinktas tekstas eilutėmis įkeltas į lentelę, eilutėms priskirti identifikacijos numeriai. Tada tekstas išskaidytas į žodžius, jiems priskirti identifikacijos numeriai, sužymėti eilučių ID, teksto ID, skyriaus ID ir originalo puslapio ir eilučių numeriai. Visi šie veiksmai atlikti automatiškai autorės sukurtomis teksto redaktoriaus *Microsoft Word* makrokomandomis. Kadangi šiame tekste dažnai žodžių ribos nesutampa su tarpais tarp žodžių, sudarytoji lentelė peržiūrima ir pataisoma. Peržiūrėtos lentelės įkeliamos į atitinkamas morfologijos duomenų bazės lenteles „Kontekstas“ (ji pervadinta į „KNtekstas“) ir „Žodžiai“.

Kiti požymių (giminės, skaičiaus, linksnio, laiko etc.) ID įrašomi rankiniu būdu (neautomatiškai) duomenų bazės lentelėje „Žodžiai“. Įvedimo patogumui ID yra skaičiai, atitinkamos kategorijos eilės numeriai, tad, pvz., žodžio *Kryštaus* formai (vyr. vns. kilm.) nurodyti į atitinkamas grafas tereikia įrašyti „1“, „1“, „2“ (plg. 4 pav.).

4 PAV. LENTELĖS „ŽODŽIAI“ FRAGMENTAS. FORMOS NURODYMAS

ID	Žodis	Kontekstas	Tekst	Autor	Skynus	Puslap	Eilutė	Savar	Pagrindinės fo	Reikšmė	Kaitų	Gimn	Skaid	Link	Komen	Leit
20030701	ateityfiandis	200307	2	1	S01	3	7	0	eti (at)	ateiti 2	1	2	1	2	3	0
20030702	Ponš	200307	2	1	S01	3	7	0	ponas	ponas la	1	1	1	2	1	0
20030703	Kryštaus	200307	2	1	S01	3	7	0	Kristus	102/1	1	1	1	2	10	0
20030704	ėnt	200307	2	1	S01	3	7	0	ant	804/1	0	0	0	0	0	0
20030705	fuda	200307	2	1	S01	3	7	0	sūdas 1	sūdas	1	1	1	2	1	0

Prieš nurodant pagrindines formas, atitinkamas įrašas sukuriamas lentelėje „Pagrindinės formos“, o lentelės „Žodžiai“ grafoje „Pagrindinės formos“ jis įrašomas arba parenkamas išskleidus sąrašą (5 pav.).

5 PAV. LENTELĖS „ŽODŽIAI“ FRAGMENTAS. PAGRINDINIŲ FORMŲ
NURODYMAS

ID	Žodis	Kontekstas	Tekst	Autor	Sklymas	Puslapis	Eilutė	Savara	Pagrindinės formos	Reikšmė	Kaityn	Gimn	Skaid	Link	Kamien	La
20030701	ateityšiančia	200307	2	1	S01	3	7	0	eiti (sti)	ateiti 2	1	2	1	2	3	0
20030702	Ponas	200307	2	1	S01	3	7	0	ponas	ponas la	1	1	1	2	1	0
20030703	Kryžtaus	200307	2	1	S01	3	7	0	Kryžtas	- 102/1	1	1	1	2	10	0
20030704	šnt	200307	2	1	S01	3	7	0	Kryžtas	Kryžtu Kryžtas, Kryžtaus				1	0	0
20030705	šuda	200307	2	1	S01	3	7	0	Kryžtas	Kryžtu Kryžtas, Kryžtaus				1	1	0
20030706	rašytas	200307	2	1	S01	3	7	0	krivida	krividi krivida, krividos	1	1	0	1	0	0
20030707	šventas	200307	2	1	S01	3	7-8	0	krividijimas	krividi krividijimas, krividijimo	1	1	0	1	0	0
20030801	prineans	200308	2	1	S01	3	8	0	kryžavoti (nu kryžia- nukryžavoti, --ojo		5	1	0	1	0	0
20030802	Kéyp	200308	2	1	S01	3	8	0	kryžtas	kryžiu kryžtas, kryžtas	1	0	0	1	0	0
20030803	taley	200308	2	1	S01	3	8	0	krosyti	krosyti krosyti*	5	0	0	0	0	0
20030804	ira	200308	2	1	S01	3	8	0	krūpauti	krūpauti krūpauti	5	0	0	1	0	0

Šiuo metu į duomenų bazę suvesti *SE*, *Pas*, *MKr* ir *K* žodžiai, sužymėti tie jų požymiai, kuriems nereikia papildomų tyrimų arba tyrimai jau atlikti. Todėl, pavyzdžiui, pagal Antano Jakulio hipotezę (Jakulis 1982) sužymėta *SE* dalių autorystė, o lentelėje „Pagrindinės formos“ įvestas tik identifikatorius, lizdas, numanomos pagrindinės formos bei kalbos dalies ID. Pagrindinės formos gali būti koreguojamos ir kiti požymiai nurodomi vėliau, peržvelgus visas žodžio formas (žr. 3.2. sk.).

Ne visi požymiai nurodomi dėl keleto priežasčių:

– medžiagos atrankai morfologijos tyrimams nurodytų požymių (vietos, formos bei pagrindinių formų) pakanka; daug požymių numatyta kitiems, pvz., leksikologijos ar semantikos tyrimams;

– kai kurie požymiai reikalingi tik vienam kitam tyrėjui (pvz., daryba, pamatinis žodis ir kt. rūpės tik žodžių darybos specialistams), kai kuriuos ir turėtų nurodyti atitinkamas specialistas (pvz., reikšmės – leksikologas), etc.;

– daug požymių yra ginčytini ar abejotini, kai kurie priklauso nuo tyrimų atskaitos taško: pvz., žodis *kalnas* laikytinas šakniniu ar priesaginiu, priesagos vediniu, t. y., ar jis kalbėtojo sąmonėje siejamas su *kilti*, *kelti* ar ne? O *delnas*? Kaip nurodyti šaknį – tokia, kokia yra (*kal-*), ar su pamatiniu balsių kaitos laipsniu (*kel-*) – priklauso ir nuo tyrėjo poreikių, pvz., ar jis norės išrinkti žodžius su *a* šaknyje, ar visus tam tikros šaknies vedinius⁴.

– dažnai sunku apibrėžti ir linksniavimo tipą (jis nurodomas lentelėje „Pagrindinės formos“), kadangi *Knygoje Nobažnystės* gausu įvairuojančių formų, pvz.: *bjauras / bjaurus, garba / garbė* ir t. t. Kiekvienu tokiu atveju reikia iš pradžių atlikti išsamius tyrimus, nustatyti, nuo ko tas įvairavimas priklauso (gal jį lemia semantika, gal skirtingas formas vartoja skirtingi autoriai, etc.);

- regis, kiek mažiau problemų turėtų kilti nurodant konkrečios formos kamieną (lentelėje „Žodžiai“), bet ir čia tenka suabejoti: jei *ausis* – *i* kamieno, *ausiai* – *io* kamieno, *ausų* – priebalsinio, tai *ausį* – ?

Todėl tokie požymiai paliekami suvesti pačiam tyrėjui. Tam patogiu naudoti skirtingus duomenų bazės valdymo sistemos medžiagos peržiūros režimus.

⁴ Tiesa, tam numatytos dvi grafos: „Šaknis“ ir „Pamatinė šaknis“.

2.2. Medžiagos peržiūra, rūšiuojimas ir koregavimas

Medžiagą galima peržiūrėti tiesiogiai su pagrindine susietose pagalbinėse lentelėse. Lentelėje „Skyriai“ galime peržvelgti visą kiekvieno skyriaus tekstą (6 pav.).

6 PAV. LENTELĖS „SKYRIUS“ FRAGMENTAS. 13 SKYRIAUS TURINIO

PERŽIŪRA

■ Skyrius - Table					
	Skyriaus ID	Knyga	Skyriaus pavadinimas	Evangelija	Pradžios pu:
+	S12	1	Ant penktos nedelos po trims kar	Mato 13, 24:30	37
-	S13	1	Ant nedelos senu uzugaveniu	Mato 20, 1:16	4C
	ID	Eilutės			Tekstas
+	204005	ANT NEDELÓS			8
+	204006	SENU UZUGAVENIU.			8
+	204007	EWANGELIA Matt. 20.			8
+	204008	PRiliginta ir karalife dangaus žmoguj kurfay			1
+	204009	išeio tábay ankti/ famdít dárbinikus ing			1
+	204010	winnicią fawą o kad fudereia dárbinikus iz gra-			1
+	204011	fia ánt dienos/ nuliunte juos winnición fawa.			1
+	204012	Ir iBaies éna edina trečia imvada kitose fawin			1

Peržvelgdami tekstą, galime matyti ir redaguoti visus kiekvienos eilutės žodžius bei jų požymius (7 pav.).

7 PAV. DUOMENŲ PERŽIŪRA LENTELĖJE „KN TEKSTAS“

ID	Eilutės										Tekstas	Pustapis	Eilutė			
+	207005 uzulaykis/ ne rágaus smierties ánt ámziú/ kál.										1					
-	207006 beia tada iam židėy. Dabar pažind išug welnia										1					
ID	Žodis	Tekst	Autor	Skyrius	Pustap	Eilutė	Savar	Pagrindinės for	Reikšmė	Kaitų	Gimin	Skaid	Link	Kamien	Leips	Ap
	20700601	tada	1	S20	70	6	0	tada	0	0	0	0	0	0	0	0
	20700602	iam	1	S20	70	6	0	jis	0	1	2	1	3	0	0	0
	20700603	židėy	1	S20	70	6	0	žydas	0	1	1	3	1	1	0	0
	20700604	Dabar	1	S20	70	6	0	dabar	dabar 1/	0	0	0	0	0	0	0
	20700605	pažind	1	S20	70	6	0	žinti (pa)	0	2	0	3	0	0	0	0
	20700606	išug	1	S20	70	6	0	įgok	0	0	0	0	0	0	0	0
	20700607	welnia	1	S20	70	6	0	welnias	0	1	1	1	4	2	0	0
*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
+	207007 turi. Abrahamės numire ir pránašey/ o tu										1					
+	207008 kabiey kas zodi mano uzulaykis/ ne rágaus										1					

Peržvelgdami pagrindines formas, matome visus kiekvieno žodžio pavartojimo atvejus (8 pav.). Pagrindines formas koreguoti bei kitus leksemos požymius įvesti patogiau šiaime režime, nes čia matomos visos kiekvieno žodžio formos, tekstas ir skyriai, kur jos pavartotos etc. Tad tik čia galima bandyti nustatyti, pvz., kiekvienos leksemos kaitybės tipą.

8 PAV. DUOMENŲ PERŽIŪRA LENTELĖJE „PAGRINDINĖS FORMOS“

Pagrindinės formos - Table																	
ID	Lindas	Pagrindinės formos			Kalbos d	Kilmė	Strukt	Linksnis	Daryba	Pamatinis kodis	Darybos formantas			P			
* meilauti (pa)	meilauti	pameilauti, -			5	0	0	0	0	0	0	0	0				
- meile	meile	meilė, meilės // meila, meilos			1	0	1	8	0	0	0	0	0				
ID	Žodis	Kontekstas	Tekst	Autor	Slymus	Puslapj	Eihad	Savar	Reikšmė	Kaity	Gimin	Skaid	Link	Kamien	Laips	Apibrė	Va
20792301	meyles	207923	2	2	S22	79	23	0	0	1	3	1	2	8	0	0	
20792707	meyles	207927	2	2	S22	79	27	0	0	1	3	1	2	8	0	0	
20920207	meylās	209202	2	2	S24	92	2	0	0	1	3	1	2	8	0	0	
21111608	meylā	211116	2	2	S29	111	16	0	0	1	3	1	4	8	0	0	
21240304	meyleia	212403	2	2	S32	124	3	0	0	1	3	1	6	8	0	0	
21242809	meyles	212428	2	2	S32	124	28	0	0	1	3	1	2	8	0	0	
21262404	meylā	212624	2	2	S33	126	24	0	0	1	3	1	4	8	0	0	
21270204	meylā	212702	2	2	S33	127	2	0	0	1	3	1	4	8	0	0	
21270905	meyla	212709	2	2	S33	127	9	0	0	1	3	1	1	8	0	0	
21271307	meyles	212713	2	2	S33	127	13	0	0	1	3	1	2	8	0	0	
21272706	meyles	212727	2	2	S33	127	27	0	0	1	3	1	2	8	0	0	

Be to, visus duomenis galima rūšiuoti pagal bet kurį požymį didėjančia ar mažėjančia tvarka, galima jų paieška ir filtravimas.

2.3. Duomenų atranka

Medžiagos atrankai *Microsoft Access* DVBS numatytos specialios užklauskos (*Query*). Čia medžiaga (pvz., žodžiai) atrenkama pagal bet kurį požymį ar požymių grupę. Vartotojo patogumui kuriamos formos, kuriose įrašomi pageidaujami atrankos kriterijai. Tokia forma autorės yra sukurta morfologijos duomenų (kaitybos) atrankai (9 pav.), kuriamos formos leksikos ir žodžių darybos atrankai ir kt.

9 PAV. KAITYBOS DUOMENŲ ATRANKOS UŽKLAUSOS FORMA

kaityba

Pageidaujamas reikšmes pasirinkite iš sąrašų ir įrašykite (teksto laukelius vietoje *.
 * - bet kokia reikšmė.
 Nurodydami žodį, pagrindines formas ar reikšmę galite naudoti simbolius * ir ?, pvz.:
 Žodis: k*la - žodžiai prasidedantys k ir besibaigiantys la.
 Žodis: k?la - žodžiai "kala", "kyla" etc.

Žodis: Pagrindinės formos: Reikšmė:

Kalbos dalis:

Kaitymas: Savarankiškumas:

Kaitybos tipas: Veiksmažodžio forma:

Kamienas: Refleksyvumas:

Giminė: Rūšis:

Skaičius: Nuosaka:

Linksnis: Laikas:

Laipsnis: Laiko/nuos. forma:

Apibrėžtumas: Asmuo:

Valdymas:

SARAŠAI:

Kaitymas <input type="text"/>	Savarankišk <input type="text"/>	Pagr. formos <input type="text"/>
Giminės <input type="text"/>	Veiksmažodžio formos <input type="text"/>	Kalbos dalys <input type="text"/>
Skaičiai <input type="text"/>	Refleksyvumas <input type="text"/>	Kaitybos tipai <input type="text"/>
Linksniai <input type="text"/>	Rūšys <input type="text"/>	
Kamienai <input type="text"/>	Nuosakos <input type="text"/>	
Laipsniai <input type="text"/>	Laikai <input type="text"/>	
Apibrėžtumas <input type="text"/>	Laiko/nuos. formos <input type="text"/>	Tekstai <input type="text"/>
Valdymas <input type="text"/>	Asmenys <input type="text"/>	Autoriai <input type="text"/>

Record: 14 of 1

Į sukurtą formą įrašomi reikiami parametrai (jų sąrašai pateikti apatinėje formos dalyje). Pavyzdžiui, norėdami išrinkti *mokyti, moko, mokė* tipo veiksmažodžių esamojo laiko formas, į lauką „Kaitybos tipas“ įrašome „511“, į „Laikas“ – „1“. Bakstelėjus mygtuką „Išrinkti“, bus pateiktas toks rezultatas:

2 LENTELE. MOKYTI, MOKO, MOKĖ TIPO VEIKSMAŽODŽIŲ ESAMOJO LAIKO FORMŲ ATRANKOS REZULTATAS

Lizdas	Pagrindinės formos	Žodis	Forma	Kontekstas	Autorius	Skyrius
pildyti	išpildyti, -pildo/-pildžia, -pildė	išpildžia	tics.es.3	204623	1	S14
laikyti	užlaikyti, -laiko, -laikė	vžfiláyka	sangr.tics.es.3	204224	1	S13
valgyti	valgyti, valgo, valgė	wałgi	tics.es.3	205703	1	S17
sakyti	sakyti, sako, sakė	fakote	tics.es.dgs.2	206213	1	S18
klausyti	klausyti, klauso, klausė	kłaufa	tics.es.3	204727	1	S14
mokyti	mokyti, moko, mokė	mokia	tics.es.3	200821	1	S02
.....						

Pastebėję, kad esamojo laiko formas vartojamos tiek *o*, tiek ir *ia* kamieno, atranką patiksliname, nurodydami išrinkti tik *ia* kamieno formas (kamienas: 2). Pateikiamas toks rezultatas:

3 LENTELE. MOKYTI, MOKO/MOKIA, MOKĖ TIPO VEIKSMAŽODŽIŲ IA KAMIENO FORMŲ ATRANKOS REZULTATAS

Lizdas	Pagrindinės formos	Žodis	Forma	Kontekstas	Autorius	Skyrius
rodyti	rodyti, rodo/rodžia, rodė	rodziáme	tics.es.dgs.1	223129	1	S57
pildyti	išpildyti, -pildo/-pildžia, -pildė	išpildžia	tics.es.3	204623	1	S14
rodyti	rodyti, rodo/rodžia, rodė	rodžia	tics.es.3	201219	1	S03
mokyti	mokyti, moko, mokė	mokia	tics.es.3	200821	1	S02
mokyti	mokyti, moko, mokė	mokia	tics.es.3	200219	1	S01
mokyti	mokyti, moko, mokė	mokia	tics.es.3	221313	2	S53

Galima ir toliau tikslinti atranką, pavyzdžiui, išrinkti kiekvieno šių veiksmažodžių visas esamojo laiko formas:

4 LENTELĖ. VEIKSMAŽODŽIO *IŠPILDYTI* ESAMOJO LAIKO FORMŲ
 ATRANKOS REZULTATAS

Lizdas	Pagrindinės formos	Žodis	Forma	Kontekstas	Autorius	Skyrius
pildyti	išpildyti, -pildo/-pildžia, -pildė	išpildžia	ties.cs.3	204623	1	S14
pildyti	išpildyti, -pildo/-pildžia, -pildė	išpilda	ties.es.3	304612	2	M36

Nuosekliai atliekant atranką, siaurinant duomenų diapazoną ir tikslinant kriterijus galima rinkti medžiagą bet kokio konkretaus klausimo sprendimui, pvz., jau iš 3 lentelėje pateiktų duomenų galima spręsti, kad *ia* kamieno šio tipo veiksmožodžių formas labiau linkę vartoti pirmasis, t. y. pirmųjų ir paskutiniųjų pamokslų, autorius.

IŠVADOS

Tinkamas informacinių technologijų priemonių naudojimas leidžia gerokai sumažinti senųjų raštų tyrėjų laiko sąnaudas. Autorės sukurta **senųjų raštų morfologijos duomenų bazė** leidžia greitai ir tiksliai spręsti įvairius kalbotyros uždavinius (pvz., daiktavardžių kamienų mišimo, teksto autorystės ir pan.). Laiko užima tik duomenų įvedimas, o jų atranka ir rūšiavimas atliekamas per kelias sekundes, patys duomenys išsaugomi, išlieka nepakitę. DBVS paieškos, atrankos, rūšiavimo, koregavimo ir ataskaitų generavimo priemonės sudaro naują morfologijos tyrinėtojo metodikos arsenalą, o sudarytas loginis modelis leidžia kalbą tirti kompleksiskai ir sistemiskai. Ne mažiau svarbu ir tai, kad duomenys į sukurta morfologijos duomenų bazę gali būti įkeliami iš kitų šaltinių, taip pat nesunku juos konvertuoti ir į kitus formatus (pvz., Lietuvių kalbos institute kuriamą Istorinį lietuvių kalbos žodyną).

Morfologijos duomenų bazė gali būti naudojama ne tik morfologijai tirti, bet ir duomenims saugoti, jiems pateikti įvairiomis formomis įvairių poreikių vartotojams. Pavyzdžiui, teologams ar literatūrologams galima sukurti formą, kurioje būtų skaitomas tekstas ir paaiškinami nesuprantami žodžiai, numatomos vietos komentarams rašyti.

Sudaryta konkretaus šaltinio – *Knygos nobažnystės morfologijos duomenų bazė* yra tik įrankis, tarpinis žingsnis, kuriame bus kaupiami esamų ir būsimų šios knygos tyrimų rezultatai. Tokia duomenų bazė gali būti publikuojama tik suvedus visus duomenis (ar dalį būtinausių duomenų), tačiau tai ne vieno žmogaus darbas. Kaip įrankis *Knygos nobažnystės* morfologijai tirti ji gali būti naudojama ir šiuo metu (ji pateikiama internete adresu www.ik.ku.lt/~jakulytelapie.htm).

SUTRUMPINIMAI

DBVS – Duomenų bazių valdymo sistema.

ID – Identifikatorius, unikalus duomenų bazės elemento pavadinimas.

K – katekizmas, paskutinysis MKr skyrius („KATECHIZMAS ABA TRVMPAS PAMOKSLAS Wieros Krikščioniškios del waykialu mažu.“).

KN – *Knyga nobažnystės* („KNIGA Nobazniftes Krikščionifzkos/ Ant gárbos Diewuy TRAYCEY Szwentoy Wienatijam: Ant wartoimá Báznicioms dides Kunigiftes Lietuwos išduotá. KIEDAYNISE, DRVKAWOIA, IOCHIMAS IVRGIS RHETAS, Meatu Poná, 1653.“).

MKr – *Maldos Krikščioniškos* („MALDOS KRIKSCIONISZKOS, Wifokiam meatuy ir reykalams bendriems prigulinčios: Ant garbos Diewa Traycey Szwentoy wienu Tėwá, Sunaus, ir Dwašios S. O del naudos Bažnicios io, ant fwieta ifzlayftos <...> KIEDAYNISE, SPAUDE, JOCHIMAS JURGIS RHETAS, Meatu Poná, 1653.“). Trečioji KN dalis, 58 maldos ir katekizmas (žr. *K*).

Pas – Pasija, paskutinysis SE skyrius („PASSYA ABA HISTORYA Apie kánčią fmeri, ir pákafimá Poná mufu IEZVSA Chryftufá, iB wifu kiáturiu Ewánjelistu pátogiy furinkta“).

SE – *Suma Evangelijų* („SUMMA Abá Trumpas ifzguldimas EWANIELIV SZWENTU Per Wifus meatus/ dienomis Nedelos/ Bážniciofe Krikščioniškofe fkyptomu/ O ánt gárbos Wienam DIEWUY, fu wala Wiriafniuju del pážitka Wierniems, dides Kunigiftes Lietuwos, ir del naminiku pabažnu ánt užiwomá namifzka ifzduotá. KIEDAYNISE, DRVKAWOIA, IOCHIMAS IVRGIS RHETAS, Meatu Poná, 1653.“). Antroji KN dalis, 60 evangelijų ir pamokslų bei Pasija (žr. *Pas*).

LITERATŪRA

AMBRAZAS, S., ALEKNAVIČIENĖ, O., ZINKEVIČIUS V. 1998: Istorinis lietuvių kalbos žodynas ir senųjų raštų kompiuterizavimas. *Lietuvių kalbotyros klausimai* 39, 192–210.

BIRŽIŠKA, V. 1990: *Aleksandrynas I*. Antrasis fotografuotinis leidimas, Vilnius: Lietuvos kultūros fondas.

JAKULIS, A. 1982: „Sumos evangelijų“ ir „Maldų krikščioniškų“ autorystės klausimu. *Baltistica* 18 (2), 173–178.

JAKULIS, A. 1984: „Sumos evangelijų“ ir „Maldų krikščioniškų“ ryšiai su ankstesniais lietuviškais raštais. *Kalbotyra* 35 (1), 31–41.

JAKULIS, A. 1995: „*Knygos nobažnystės*“ leksika, Klaipėda: Sauliaus Jokužio leidykla-spaustuvė.

JAKULYTĖ, D. 2001: Informacinių technologijų panaudojimo galimybės senųjų raštų kalbos tyrimams („*Knygos nobažnystės*“ morfologijos duomenų bazė). *Vārds un tā pētīšanas aspekti. Rakstu krājums* 5, 152–168.

LUKŠAITĖ, I. 2001: *Knygos nobažnystės* (1653) parengimo kultūrinė aplinka. *Knyga nobažnystės krikščioniškos (1653) – XVII a. Lietuvos kultūros paminklas*, Kėdainiai: Kėdainių krašto muziejus, 4–29.

MARCINKEVIČIENĖ, R. 1997: Tekstynų lingvistika ir lietuvių kalbos tekstynas. *Lituanistica* 1 (29), 58–78.

POCIŪTĖ, D. 2001: *Knygos nobažnystės* giesmių šaltiniai ir kalvinistinio pamaldumo slinkty: nuo tikėjimo tikro prie tobulo. *Knyga nobažnystės krikščioniškos (1653) – XVII a. Lietuvos kultūros paminklas*, Kėdainiai: Kėdainių krašto muziejus, 41–54.

POCIŪTĖ, D., reng., 2004: *Knyga nobažnystės krikščioniškos 1653*, Vilnius: Lietuvių literatūros ir tautosakos institutas.

POCIŪTĖ, D. 2004: Lietuviškosios bažnyčios knyga. *Knyga nobažnystės krikščioniškos 1653*, Vilnius: Lietuvių literatūros ir tautosakos institutas, xiii–lxv.

TUMELIS J. 1967: *Knygos nobažnystės krikščioniškos* (1653) pirmojo leidimo klausimu. *Bibliotekininkystės ir bibliografijos klausimai* 6, 132.

TUMELIS J. 1968: Dar kartą apie „*Knygą nobažnystės krikščioniškos*“ (1653). *Surastas pilnas pirmojo leidimo egzempliorius, Literatūra ir menas* 44 (1146), 12.

ZINKEVIČIUS, Z. 1988: *Lietuvių kalbos istorija III. Senųjų raštų kalba*, Vilnius: Mokslas.

Dalia Jakulytė
Klaipėdos universitetas
Baltų kalbotyros ir etnologijos katedra
H. Manto g. 84, 5808 Klaipėda

Gauta 2004 11 12