



DARBAI IR DIENOS

58

Redakcijos kolegija

Egidijus Aleksandravičius (Vytauto Didžiojo universitetas, istorija) – vyriausiasis redaktorius

Ineta Dabašinskiė (Vytauto Didžiojo universitetas, kalbotyra)

Milda Danytė (Vytauto Didžiojo universitetas, literatūra)

Leonidas Donskis (Vytauto Didžiojo universitetas, filosofija)

Bronislovas Genzelis (Vytauto Didžiojo universitetas, filosofija)

Jan Jurkiewicz (Poznanės Adomo Mickevičiaus universitetas, istorija)

Violeta Kelertienė (Ilojaus universitetas, literatūra)

Jūratė Kiaupienė (Lietuvos istorijos institutas, istorija)

Antanas Kulakauskas (Vytauto Didžiojo universitetas, istorija)

Darius Kuolys (Lietuvių literatūros institutas, literatūra)

Rūta Marcinkevičienė (Vytauto Didžiojo universitetas, kalbotyra)

Gintautas Mažeikis (Vytauto Didžiojo universitetas, kultūrinė antropologija)

Rein Raud (Talino universitetas, filosofija)

Leonardas Sauka (Lietuvių literatūros ir tautosakos institutas, folkloristika)

Alfred Erich Senn (Viskonsino universitetas, istorija)

Viktorija Skrupskelytė (Oberlino kolegija, JAV, Vytauto Didžiojo universitetas, literatūra)

Giedrius Subačius (Ilojaus universitetas, kalbotyra)

Joachim Tauber (Liuneburgo Šiaurės Rytų Europos institutas, istorija)

Rasa Žukienė (Vytauto Didžiojo universitetas, menotyra)

Mūsų svetainė internete

<http://biblioteka.vdu.lt/ml-dd.htm>

UNIVERSITAS VYTAUTI MAGNI



DARBAI IR DIENOS

58

Eina nuo 1930 metų



Vytauto Didžiojo universitetas

2012

TURINYS – CONTENTS

DARBAI IR DIENOS T. 58 – DEEDS AND DAYS VOL. 58

NUO REDAKCIJOS – FROM THE EDITOR

AKADEMINIO SUPRATIMO SUNKUMAI IR DAR DAUGIAU	7
Difficulties of Academic Understanding and More	

STRAIPSNIAI – ARTICLES

Kastytis Antanaitis	SOVIETINĖ KULTŪROS REGLAMENTAVIMO SISTEMA LIETUVOJE XX A. II PUSĖJE	9
	The System of Soviet Culture Regulation in Lithuania during the Second Half of the 20 th Century	27
Justina Petrulionytė	ŽANRINIŲ APIBRĖŽČIŲ SUMAIŠTIS: NEFIKCINIS AR ŽURNALISTINIS ROMANAS?	29
	The Confusion of a Genre's Definitions: Non-Fiction or Journalistic Novel?	49
Jogilė Teresa Ramonaitė	GIMTOJI IR PIRMOJI KALBA. AR VISADA JOS SUTAMPA?	51
	Native and First Language: Do They Always Coincide?	61

MONOGRAFINĖ DALIS – MONOGRAPHS

TEKSTYNŲ LINGVISTIKA: PRIEIGOS IR IŽVALGOS LINGUISTICS OF TEXTUAL CORPUSES: APPROACHES AND INSIGHTS

Inesa Šeškauskienė	METAFORŲ TYRIMO KLAUSIMU, ARBA KAIP NUSTATYTI METAFORAS	65
	On Metaphor Research Methodology, or How to Identify Metaphors	77
Jurgis Pakerys, Erika Rimkutė, Andrius Utka	LIETUVIŲ KALBOS DENOMINATYVINIAI VEIKSMAŽODŽIAI IR GRETA JŲ VARTOJAMI PAMATINIAI ŽODŽIAI (TEKSTYNO TYRIMAS)	81
	Co-Occurrence of Denominal and Deadjectival Verbs with Their Base Words in Modern Lithuanian (A Corpus-Based Study)	91
Viltė Drūlienė	APIBENDRINAMIEJI DAIKTAVARDŽIAI	93
	Generalizing Nouns	114
Jurgita Vaičenonienė	NEĮPRASTŲ KOLOKACIJŲ VERTIMAS GROŽINĖS LITERATŪROS KŪRINIULOSE: TEKSTYNU PAREMTAS TYRIMAS	117
	Translation of Unusual Collocations in Literary Texts: A Corpus-Based Approach	135
Jonė Grigaliūnienė, Rita Juknevičienė	CORPUS-BASED LEARNER LANGUAGE RESEARCH: CONTRASTING SPEECH AND WRITING	137
	Tekstiniais pagrįsti mokinių kalbos tyrimai: sakytinės ir rašytinės kalbos gretinimas	150

Jolanta Šinkūnienė, Daniël Van Olmen	MODAL VERBS OF NECESSITY IN ACADEMIC ENGLISH, DUTCH, AND LITHUANIAN: EPISTEMICITY AND/OR EVIDENTIALITY? 153 Būtinybės modaliniai veiksmažodžiai anglų, olandų ir lietuvių mokslo kalboje: epistemiškumas ir (ar) evidencialumas? 179
Jūratė Ruzaitė	ATTENUATIVE MORPHOLOGICAL FORMS ACROSS VARIETIES AND REGISTERS: A CORPUS-BASED STUDY OF THE SUFFIX <i>-ISH</i> 183 Švelninamųjų morfologinių formų žanrinis ir regioninis variantiškumas: tekstynais paremta priesagos <i>-ish</i> analizė 199
Violeta Kalėdaitė, Ieva Bastienė	NEGATION STRATEGIES IN ENGLISH AND LITHUANIAN 201 Neigimo raiška anglų ir lietuvių kalbose 221
Aurelija Usonienė	KOMPLEMENTINIAI PREDIKATAI IR JŲ MULTIFUNKCIONALUMAS: LIETUVIŲ KALBOS TEKSTYNAIS PAREMTAS TYRIMAS 223 Complement-Taking Predicates and Their Multifunctionality in Lithuanian: A Corpus-Based Study 232
Danguolė Valančė	BŪTENT KAS YRA <i>BŪTENT</i> 235 What, After All, Is <i>Būtent</i> ? 255
Jolanta Gelumbeckaitė, Mindaugas Šinkūnas, Vytautas Zinkevičius	<i>SENOSIOS LIETUVIŲ KALBOS TEKSTYNAS</i> (SLIEKKAS) – NAUJA DIACHRONINIO TEKSTYNO SAMPRATA 257 <i>Old Lithuanian Reference Corpus</i> (SLIEKKAS). A New Concept of a Historical Corpus 275
Skaistė Aleksandravičiūtė, Loreta Vaicekauskienė	JUOKAI JUOKAIS, BET ŠIANDIEN JUOKIAMĖS DAUGIAU IR KITAIP: JUOKAS KAIP TV IR RADIJO DISKURSO BRUOŽAS 1960–2010 METAIS . . 283 Today We Laugh More and Differently: Laughter as a Discourse Feature in TV and Radio Broadcasts (1960-2010) 293
ATRADIMAI – DISCOVERIES	
Mindaugas Balkus, Vytautas Budvytis	LIETUVOS DIDŽIOSIOS KUNIGAIKŠTYSTĖS VAIZDAI XVIII AMŽIUIJE . . 297 Images of the Grand Duchy of Lithuania in the 18 th Century 322
RECENZIJOS – REVIEWS	
Tomas Kavaliauskas	APIE ZENONO NORKAUS TELEOLOGINĮ MĄSTYMĄ KLAUSIANT: AR VIDURIO EUROPOJE POKOMUNISTINĖ TRANSFORMACIJA JAU UŽBAIGTA? 323 On the Teleological Thought of Zenonas Norkus Wondering Whether the Post-Communist Transformation of Central Europe Is Already Complete

Rūstis Kamuntavičius	<i>LIETUVISAI IR LICVINAI: NAUJOS SENOS GUDŲ ISTORIOGRAFIJOS PAIEŠKOS</i> 335 <i>Novel Old Explorations in Belarusian Historiography</i>
Vytas Jankauskas	<i>APIE AKADEMINĮ SAŽININGUMĄ</i> 341 <i>On Academic Conscientiousness</i>
	POLEMIKOS AIDAS – POLEMICAL ECHO
Darius Staliūnas	<i>APIE RUSINIMĄ: ATSAKYMAS RECENZENTUI</i> 345 <i>On Russification: Reply to a Reviewer</i>
	APIE AUTORIUS – CONTRIBUTORS 357

SENOSIOS LIETUVIŲ KALBOS
TEKSTYNAS (SLIEKKAS) –
NAUJA DIACHRONINIO TEKSTYNO SAMPRATA



JOLANTA GELUMBECKAITĖ, MINDAUGAS ŠINKŪNAS,
VYTAUTAS ZINKEVIČIUS

ISSN 1392-0588
2012. 58

1. PRAELIMINARIA. TIKSLAI

Giliosios anotacijos referencinis „Senosios lietuvių kalbos tekstynas“ (SLIEKKAS, vok. „Referenzcorpus Altlitauisch“)¹ turi apimti visus lietuviškai parašytus tekstus nuo rišliosios raštijos pradžios apie 1520 m. iki 1800 metų (XVI–XVII–XVIII a.). Jį sudarys apie 10 mln. lietuviškų teksto žodžių. Periodo pabaigą sutartinai gali žymėti 1800 metais išleista Kristijono Gotlybo Milkaus (Christian Gottlieb Mielcke, 1732–1807) gramatika „Anfangs-Gründe einer Littauischen Sprach-Lehre“ („Lietuvių kalbos pagrindai“, Karaliaučius: Hartung)². Tekstyno tekstų pagrindas – moksliniai kritiniai leidimai, kurie visais atvejais bus gretinami su rankraštiniais ar spausdintais originalais, arba specialiai tekstynui parengti dokumentiniai paraidiniai tekstų perrašai.

Sudėtinga ir kompleksiška istorinių tekstynų rengimo metodika verčia pradiniam etape pasirinkti reprezentatyvų mažesnės apimties bandomąjį tekstyną, apimančią charakteringus senosios lietuvių kalbos laikotarpio tekstus bei jų šaltinius, kuriuo remiantis bus parengta visa tekstyno sandara ir išmėgintas jo veikimas.

SLIEKKO galutinis tikslas yra reprezentuoti šio laikotarpio kalbą kaip visumą ir pateikti išsamią informaciją apie kiekvieno teksto (ir apie visų tekstų kartu) kalbą, jos variantus, žodyną ir gramatiką. Tekstynas skirtas diachroninių lingvistinių duomenų procesų analizei ir pranoksta tradiciškai suprantamo tekstyno (kaip teksto vienetų rinkinio) ribas savo struktūra, sudaryta pagal keturis kriterijus: segmentinė struktūra (teksto sluoksnio vienetų sekvencijos: ženklai, žodžiai, frazės, sakiniai ir pan.), intra-segmentinė struktūra (anotacijų sluoksniai: tekstologinė ir lingvistinė informacija),

¹ Rengiant tekstyną nuo 2009 m. bendradarbiauja Frankfurto prie Maino Goethe's universitetas, Lietuvių kalbos institutas ir Pisos universitetas. 2010 metais projektą „Senosios lietuvių kalbos referencinis korpusas: *Wolfenbüttelio postilė* (1573)“ rėmė LR švietimo ir mokslo ministerija (Nr. S-358, Nr. S-512). Nuo 2012 m. tekstyno rengimą finansuoja Lietuvos mokslo taryba („Senosios lietuvių k. tekstyno [SLIEKKAS] lingvistinės anotacijos technologinė ir mokslinė bazė“, sutarties Nr. VAT-42/2012).

² Nustatant lietuvių kalbos senojo periodo pabaigos ribą remiamasi Jono Palionio skirstymu (Palionis ²1995, 14–15).

intersegmentinė struktūra (anotacijų susiejimas į seką vienu lygmeniu) ir hipersegmentinė struktūra (teksto ir faksimilės, versto teksto ir vertimo šaltinio, atskirų tekstų lygiagretinimas).

SLIEKKAS yra daugiapakopės (angl. *stand-off*) architektūros struktūra, kurią sudaro trys pagrindiniai sluoksniai: metalingvistinė informacija (angl. *header information*)³, patys tekstai, pateikiami įvairiais lygmenimis, ir tekstologinė bei lingvistinė anotacija, kuri savo ruožtu bus skirstoma į smulkesnius sluoksnius (1 schema).

Tekstų metaduomenys	Paieškos įrankiai
Rašytiniai tekstai (empiriniai kalbos duomenys)	
Tekstologinė ir lingvistinė tekstų anotacija	

1 schema. SLIEKKO architektūra

Tokią architektūrą įgalina sukurti XML (angl. *Extensible Markup Language*) duomenų struktūrų kalbos formatus. Kiekvieną sluoksnį galima plėsti ir pildyti nepažeidžiant kitų, tačiau visi sluoksniai yra tarpusavyje sinchronizuoti ir susieti. Šis formatas leidžia pagal įvairius kriterijus koduoti pirminius teksto duomenis ir pateikti visas reikalingas anotacijas. Visą tekстыne pateiktą informaciją bus galima tirti pagal įvairiopai pasirenkamus paieškos kriterijus, modeliuojamus pagal kiekvieną anotacijos sluoksnį ir (ar) pagal kelių sluoksnių kombinacijas. Tekstyno ypatybės yra šios:

- jis yra multimodalinis, nes lietuviški tekstai lygiagretinami (alignuojami) su originalo faksimilėmis;
- išsami informacija pateikiama pažymų (anotacijų) forma, kurios, be metainformacijos, apima lemavimą, glosavimą (dabartine lietuvių k. ir anglų k., numatoma ir vokiečių k.), gramatinę informaciją (*POS-tagging = part of speech-tagging*, morfologinė ir bazinė sintaksinė informacija), informaciją apie teksto struktūrą, paleografinę ir tekstologinę informaciją;
- anotacijų sluoksniai pateikiami daugiapakopės (angl. *stand-off*) architektūros principu;
- lietuviški tekstai su visais anotacijų sluoksniais lygiagretinami tarpusavyje ir su atitinkamai anotuotais vertimo šaltinių tekstais.

Būdingas senosios lietuvių k. raštijos bruožas yra glaudus ryšys su lotyniškais, vokiškais, lenkiškais ir, ypač Biblijos atveju, sen. graikiškais ir hebrajiškais vertimo šaltiniais. Autochtoniškų tekstų šiame periode yra mažuma. Dauguma tekstų yra pažodinių vertimų, perdirbinių bei kompiliuotų vertimų ir originalių lietuviškų tekstų sintezė. Vertimų šaltiniai (bandomojo tekstyno atveju apie 190 000 teksto žodžių; kiek įmanoma remiamasi moksliniais kritiniais leidimais) taip pat bus įtraukti į tekstyną, jiems taikoma

³ Metaaprašė pateikiama (istorinei) kalbotyrai, literatūrologijai ir kitiems mokslams (istorijai, knygotyrai, teologijai) svarbi informacija. Remiantis metaaprašo pažymomis paieška tekстыne bus galima pagal įvairius kriterijus, pvz., pagal datą, žanrą, kalbos variantą, autorių ir t. t.

ta pati architektūra, jie anotuojami pagal tuos pačius principus ir lygiagretinami su atitinkamais lietuviškais tekstais. Šis procesas būtinas siekiant dokumentuoti ir nustatyti vertimo šaltinių ir kalbų poveikio lietuviškiems vertimams laipsnį. To paties žanro lietuviški tekstai taip pat bus lygiagretinami tarpusavyje siekiant nustatyti įvairiopą interferenciją ir sekti proceso eigą viename žanre bei tarp žanrų.

SLIEKKO sandara remiasi senosios vokiečių kalbos referencinio tekstyno „Referenzkorpus Altdeutsch“⁴ sistema. SLIEKKO principai derinami su Vokietijos Heseno žemės skaitmeninės humanitarikos programoje „Digital Humanities“⁵ rengiama istorinių tekstynų sudarymo metodika. Ji apima anotacinių procedūrų kūrimą ir taikymą istoriniams tekstynams, vertimo bei vertimo šaltinio(-ių) tekstų susiejimo ir įvairiopų lygmenų lygiagretinimo technologijas. SLIEKKAS taip pat atsižvelgia į Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centre rengiamo „Dabartinės lietuvių kalbos tekstyno“⁶ pusiau automatinio lemavimo, morfologinės anotacijos ir morfosintaksinio išdviprasminimo (disambiguacijos) taikymo principus⁷.

„Senosios lietuvių kalbos tekstyno“ tikslas yra parengti XML struktūrą, kuri apimtų visus anotacinius sluoksnius. Visų pažymų generavimui šiuo metu geriausiai pritaikyta Maxo Plancko psicholingvistikos institute, Nijmegene (Olandija), sukurta leidybinė programa ELAN (angl. „Linguistic Annotator“, Max-Planck-Institute for Psycholinguistics Nijmegen, The Netherlands, <<http://www.lat-mpi.eu/tools/elan>>), reguliariai tobulinama istorinių tekstynų rengimo reikmėms. ELAN'o duomenų struktūros gali būti sudarytos tiesiogiai remiantis kalbinių duomenų tvarkymo ir analizės programa „Toolbox“ (angl. *Data management, parsing and text analysis*, <<http://www.sil.org/computing/toolbox/>>) arba jos gali būti generuojamos iš specialiai SLIEKKUI programuojamų modulių, kuriuose suvedami teksto duomenys ir susiejami su leksine, gramatine bei kita informacija. Taigi SLIEKKO pirminis uždavinys yra parengti pusiau automatinę procedūrą, kuri įgalintų skaitmeninius teksto duomenis išplėsti reikalingais pažymų sluoksniais (pirmiausia lemavimo, glosavimo ir morfologinės analizės).

⁴ URL: <<http://www.deutschdiachrondigital.de/>>. Nuo 2008 m. pabaigos Berlyno Humboldtų, Frankfurto prie Maino Goethe's ir Jenos Friedricho Schillerio universitetuose vykdomas „Referenzkorpus Altdeutsch“ yra iniciatyvos „Deutsch Diachron Digital“ (DDD) dalis. Rengiamą tekstyną sudaro senosios vokiečių aukštaičių ir senosios vokiečių žemaičių (saksų) kalbos tekstai (750–ca. 1050 m.), jų vertimo šaltinių tekstai ir daugiapakopės morfosintaksinės anotacijos.

⁵ LOEWE-Schwerpunkt „Digital Humanities – Integrierte Aufbereitung und Auswertung textbasierter Corpora“ (URL: <<http://www.dhhe.de/>>) yra nuo 2011 m. pradžios vykdoma Frankfurto prie Maino Goethe's ir Darmstadt techninio universitetų bei Laisvojo vokiškojo rėmimo fondo (*Freies Deutsches Hochstift*, Frankfurto prie Maino Goethe's muziejus) programa, siekianti humanitarinių ir informatikos mokslų sinergijos giliosios anotacijos referencinių tekstynų srityje. Atsižvelgiant į diachroninių tekstynų rengimo sudėtingumą ir dar trūkstant tokių tekstynų problematikos sprendimus 2012 m. liepos mėnesį Hamburgo universitete įsteigta vokiškai kalbančių šalių sąjunga „Digital Humanities Deutschland“ (URL: <<http://www.dig-hum.de/>>).

⁶ URL: <<http://tekstynas.vdu.lt/locale.xhtml?method=lt>, <<http://tekstynas.vdu.lt/tekstynas/>>.

⁷ Plačiau žr.: Marcinkevičienė 2010, Rimkutė 2006.

Toliau aprašoma SLIEKKO ir viso būsimo „Senosios lietuvių kalbos tekstyno“ struktūra.

2. ATRANKINIAI TEKSTAI. SKAITMENINIMAS

Bandomajam tekstynui pasirinkta 10 tekstų (apie 350 000 teksto žodžių [toliau – t. ž.]) ir jų vertimo šaltinių (apie 190 000 t. ž.):

1. **DzP** – „**Dzūkiški poteriai**“ („Pater noster“, „Ave Maria“, „Credo“), ca. 1520:

seniausias žinomas lietuviškas rišlus tekstas; rankraštis (Vilniaus u-to biblioteka, sign.: RS II–3006); 168 t. ž.; vertimas iš lotynų, lenkų ir (ar) vokiečių k.; kritinis lietuviško teksto leidimas: Lebedys, Palionis 1964.

2. **MžK 1547** – Martynas Mažvydas, „Catechismusa prasty šadei“:

spaudinys (Vilniaus u-to biblioteka, sign.: RS LR 5650, faksimilė: <<http://gluosnis.vu.lt/biblio/dperziura.pradzia?nr=32>>; Torunės u-to biblioteka, sign.: Pol.6.II 189 adl.); ca. 7 000 t. ž.; verstas ir originalus tekstas; vertimas pagal Martino Lutherio „Mažojo katekizmo“ lenkišką vertimą (Jan Seklucjan 1545, Jan Malecki-Sandecki 1546⁸) ir lotynišką 1529 m. leidimą (pirmas ir antras Johannes Sauermano [Sauerromanno] vertimas) bei lotynišką Jodoco Willichio katekizmą (1542); giesmyno vertimas pagal lenkišką Jano Seklucjano giesmyną (1547) ir atskirus vokiškus giesmynus; kritinis lietuviško teksto leidimas: Ročka 1974 (gretintinas su: Bezzenberger 1874, Ford 1971); kritiniai lenkiškų vertimo šaltinių leidimai: Rospond 1948 (gretintini su: Neumann 1941), Kalisz 2007; lotyniškų 1529 m. katekizmų leidimai: Luther 1529.

3. **MžGA 1549** – Martynas Mažvydas, „Giesme S. Ambrašeiaus“:

spaudinys (Lenkijos mokslų akademijos biblioteka, sign.: Cim.O.279); 2 500 t. ž.; verstas ir originalus tekstas; vertimas iš lotyniškų „Te Deum laudamus“, „Christus passus est pro nobis“ ir „Christe surrexisti“ tekstų remiantis lenkišku Lutherio „Mažojo katekizmo“ vertimu (Jan Seklucjan 1545, Jan Malecki-Sandecki 1546) bei lenkišku Jano Seklucjano giesmynu (1547); kritinis lietuviško teksto leidimas: Ročka 1974; kritiniai lenkiškų vertimo šaltinių leidimai: Rospond 1948 (gretintini su: Neumann 1941), Kalisz 2007.

4. **MžFK 1559** – Martynas Mažvydas, „Forma Chrikstima“:

spaudinys (Torunės u-to biblioteka, sign.: Pol.6.II.5 adl., Pol.6.II.190 adl., faksimilė: <<http://kpbc.umk.pl/dlibra/docmetadata?id=26738>>); ca. 3 200 t. ž.; verstas tekstas; vertimas pagal vokišką „Form der Tauff. Wie die in der Kirchenordnung des Hertzogthumbs Preussen vnd andern mehr verfasst“ (atskiras leidimas, Karaliaučius:

⁸ Jano Malecki-Sandeckio katekizmo, konvoliute su MžFK 1559 saugomo Torunės u-to bibliotekoje (sign.: Pol.6.II.4 adl.), faksimilė: <<http://kpbc.umk.pl/dlibra/docmetadata?id=26738>>.

Johann Daubmann, 1559; laikomas dingusiu); kritinis lietuviško teksto leidimas: Ročka 1974 (gretintinas su: Bezzenberger 1875); vokiško teksto leidimas: Hubatsch 1968.

5. **WP 1573** – „Wolfenbüttelio postilė“:

rankraštis (Hercogo Augusto biblioteka, Wolfenbüttelis, sign.: Cod. Guelf. 11. 2 Aug. 2°, faksimilė: <<http://diglib.hab.de/mss/11-2-aug-2f/start.htm>>); ca. 150 000 t. ž.; verstas ir originalus tekstas; vertimas pagal mažiausiai 9 lotyniškus Reformacijos teologų tekstus⁹, gausios klasikų, Biblijos ir Bažnyčios tėvų citatos; kritinis leidimas: Gelumbeckaitė 2008a, b.

6. **VE 1579** – Baltramiejus Vilentas, „Enchiridion“:

spaudinys (Slaptasis valstybinis Prūsijos kultūros paveldo archyvas, sign.: XX StUB Königsberg, Nr. 54 [defektinis]; Vilniaus u-to biblioteka, sign.: LR 1386 [defektinis]); ca. 12 000 t. ž.; verstas ir originalus tekstas; vertimas pagal vokišką M. Lutherio „Mažąjį katekizmą“; iš dalies MžK 1547 perdirbinys; kritinis lietuviško teksto leidimas: Ford 1969 (gretintinas su: Bechtel 1882).

7. **DK 1595** – Mikalojus Daukša, „Kathechismas“:

spaudinys (Vilniaus u-to biblioteka, sign.: LR 4165, faksimilė: <<http://gluosnis.vu.lt/biblio/dperziura.pradzia?nr=42>>); ca. 17 000 t. ž.; verstas ir originalus tekstas; vertimas pagal anoniminių lenkišką Jakobo Ledesmos (Diego de Ledesma) katekizmo vertimą „Nauka Chrzescianska. Abo Kátechizmik dla dżiatek“ (prieš 1595); kritinis lietuviško teksto leidimas ir lenkiško teksto leidimas: Jakštienė, Palionis 1995.

8. **LyK 1719** – Heinrich Johann Lysius, „Mažas Katgismas“:

rankraštis (Jogailaičių biblioteka, Krokuva, sign.: Ms. slav. quart. 4); ca. 7 000 t. ž.; vertimas pagal vokišką M. Lutherio „Mažąjį katekizmą“; kritinis lietuviško teksto leidimas: Dini 1993 (gretintinas su: Panzer 1993).

9. **EnK 1722** – Gabriel Engel, „Mažas Katgismas“:

spaudinys (Saksonijos žemės ir u-to biblioteka, sign.: 155 M 13 c.; Torunės u-to biblioteka, sign.: Pol.6.II.196, faksimilė: <<http://kpbc.umk.pl/dlibra/docmetadata?id=26447>>; Latvijos u-to akademinė biblioteka, sign.: D6 5512 [defektinis]); ca. 12 500 t. ž.; perdirbtas ir išplėstas LyK 1719 variantas, paralelus vokiškas ir lietuviškas tekstas.

10. **DM 1765 / 1775** – Kristijonas Donelaitis, „Metai“:

rankraštis ir spaudinys (dvi rankraščio dalys „Pavasario linksmybės“ ir „Vasaros darbai“ Lietuvių literatūros ir tautosakos i-to bibliotekoje, sign.: F1-5259; kitos dvi rankraščio dalys laikomos dingusiomis); ca. 130 000 t. ž.; originalus tekstas; **DMRh 1818** – Martyno Liudviko Rėzos (Ludwig Jedemin Rhesa) pirmas leidimas (faksimilė: <<http://www.epaveldas.lt/vbspi/biRecord.do?biRecordId=34061>>);

⁹ Visų nustatytų šaltinių (Gelumbeckaitė 2008b) faksimilės prieinamos pagal Wolfenbüttelio skaitmeninės bibliotekos katalogą: <<http://dbs.hab.de/katalog/?opac=wdb>>.

DMSch 1865 – Augusto Schleicherio leidimas (faksimilė: <<http://www.epaveldas.lt/vbspi/biRecord.do?biRecordId=24704>>); **DMN 1869** – Georgo Heinricho Ferdinando Nesselmanno leidimas (faksimilė: <<http://www.epaveldas.lt/vbspi/biRecord.do?biRecordId=24978>>); kritinio leidimo pagrindu laikomas: Nesselmann 1869.

Šeši tekstai (MžGA 1549, WP 1573, VE 1579, DK 1595, LyK 1719, DM 1765 / 1775; iš viso ca. 319 000 t. ž.) yra įtraukti į Frankfurto prie Maino Goethe's universiteto TITUS (vok. *Thesaurus Indogermanischer Text- und Sprachmaterialien*) duomenų bazę (<<http://titus.uni-frankfurt.de/texte/texte2.htm#lit>>)¹⁰. TITUS'o tekstai parengti pagal bendrą ženklų sekvencijų kodavimo sistemą („WordCruncher“ programinės sistemos formatu, konvertuoti į HTML duomenų struktūrų kalbą), jie indeksuoti pagal struktūrinius teksto elementus, žodžių formos susietos tarpusavyje ir iš dalies su vertimo šaltiniais¹¹. Paieška („WordCruncher“ ir HTML aplinkoje) šiuo metu galima tik pagal kalbinių ženklų sekas (žodžių formas, šaknis, automatiškai generuotas lemas) kontekstuose, ją taip pat galima modifikuoti vadinamaisiais pakaitos simboliais (angl. *wildcards*, vok. *Reguläre Ausdrücke*). Šiuo metu TITUS'o tekstai konvertuojami į bazinę XML struktūrą, kuri įgalina tiesioginį žodžių formų žymėjimą ir kurią naudojant toliau bus įvedamos reikalingos pažymos. Keturių tekstų (DzP, MžK 1547, MžFK 1559, EnK 1722) bazinė XML struktūra rengiama be tarpinio formato.

Du tekstai (WP 1573, DK 1595) taip pat įtraukti į Lietuvių kalbos institute rengiamą ir nuolat pildomą „Senųjų raštų duomenų bazę“ (<<http://www.lki.lt/seniejirastai/db.php>>)¹². Ji apima tekstus nuo 1573 iki 1816 m. Tai didžiausias ir dokumentinio tekstų perteikimo atžvilgiu patikimiausias senosios lietuvių kalbos tekstynas. Šiuo metu išoriniam vartotojui prieinama 50 tekstų (apie 2 mln. teksto žodžių, HTML) ir jų konkordansai (KWIC [angl. *Keyword in Context*] formatu: bazinis žodis kontekste). Tekstus ir konkordansus galima parsisiųsti („Word-doc“ formatu). Linijinė morfologinė anotacija yra pradėta, bet SLIEKKO tekstų kol kas neapima¹³.

Pasirinkti tekstai atspindi senosios lietuvių kalbos laikotarpio teksto žanrų, rūšių ir tipų įvairovę: a) religiniai (Biblijos ištraukos, maldos, katekizmai, giesmynai, pamokslai) ir pasaulietiniai tekstai, b) proza ir poezija, c) verstiniai, originalūs ir mišrūs tekstai, d) rankraščiai ir spaudiniai. Tekstai taip pat atspindi tris rašomuosius lietuvių k.

¹⁰ Visi tekstai SLIEKKO projekte aktualizuojami pagal kritinius leidimus resp. originalus. Donelaičio „Metų“ tekstas, TITUS'e parengtas pagal akademinį 1977 m. leidimą (Donelaitis 1977), bus sinoptiškai išplėstas pagal kritinius principus parengtu rankraštiniu tekstu ir pagal kritinius principus parengtais DMRh 1818, DMSch 1865 bei DMN 1869 leidimų tekstais.

¹¹ Plačiau žr.: Gippert 2005.

¹² URL: <<http://www.lki.lt/seniejirastai/db.php?source=37>>, <<http://www.lki.lt/seniejirastai/db.php?source=1>>.

¹³ *Senųjų raštų* anotacijos principus parengė Ona Aleknavičienė, plg.:

<<http://www.lki.lt/seniejirastai/Downloads/Indeksai/PB.html>>,

<<http://www.lki.lt/seniejirastai/Downloads/Indeksai/RgM.html>>,

<<http://www.lki.lt/seniejirastai/Downloads/Indeksai/RsG.html>> ir kt.

variantus: vakarinį (VE 1579, LyK 1719, EnK 1722, DM 1765 / 1775), vidurinį (DK 1595) ir rytinį (DzP) bei kelių variantų samplaiką (MžK 1547, MžGA 1549, MžFK 1559, WP 1573). Tekstai skiriasi (orto)grafija ir akcentografija (kirčiuoti tekstai: DK 1595, DM 1765/75; iš dalies LyK 1719, EnK 1722). Tekstų heterogeniškumas ne tik iškalbingai atspindi kalbamojo laikotarpio įvairialypumą, bet ir verčia ieškoti naujų teksto technologinių problemų sprendimų.

Dabar prieinami kritiniai šaltinių tekstų leidimai (įvairūs lotyniški maldų ir giesmių tekstai; Luther 1529; Hubatsch 1968; Rospond 1948; Neumann 1941; Kalisz 2007; Jakštienė, Palionis 1995) taip pat bus skaitmeninami ir ruošiami tolesnei anotacijai. Vokiškų ir lotyniškų tekstų pildymas pažymomis vyks bendradarbiaujant su „Referenzkorpus Altdeutsch“. Lenkiškų tekstų anotacijų principai dar turi būti parengti.

3. ANOTACINIAI SLUOKSNIAI. PUSIAU AUTOMATINIS ANOTAVIMAS

Kaip minėta, SLIEKKO pirminis uždavinys yra parengti pusiau automatinę procedūrą, kuri įgalintų skaitmeninius teksto duomenis išplėsti numatytais pažymų sluoksniais. Nesant lietuvių kalbos istorinio žodyno ir atskirų laikotarpių ar autorių žodynų, kuriuose būtų suregistruota visa numatytoms pažymoms reikalinga kalbinė informacija, neįmanomas tokios informacijos automatinis priskyrimas tekstyno tekstų žodžių formoms. Todėl kuriama dalinės žodžių formų anotavimo automatizacijos procedūra, naudojami tokie resursai:

- duomenų bazės TITUS lietuviškų tekstų žodžių formų sąrašas (41 213 žodžių formų tipų [angl. *types*]¹⁴;
- lemų sąrašai iš atskirų kalbamojo laikotarpio autorių ir tekstų žodynų;
- automatinės morfologinės lietuviškų žodžių analizės programinė įranga „Lemuoklis“¹⁵;
- struktūrizuotų duomenų rinkmenos, sukurtos skaitmeninant „Lietuvių kalbos žodyną“¹⁶.

Anotavimo procedūros metu kiekvienai žodžio formai automatiškai ieškoma atitiktoms žodžių formų žodyne, iš kurio išrenkama ir įrašoma į anotaciją formos gramatinė charakteristika bei nustatoma lema. Paskui iš lemų žodyno automatiškai išrenkama ir įrašoma į anotaciją lemos charakteristika. Tais atvejais, kada automatinė paieška žodynuose (tiek žodžių formų žodyne, tiek lemų žodyne) nėra viena-reikšmė (paieškos kriterijus atitinka daugiau negu vienas įrašas žodyne), automatiškai suformuojama pasirinkimo iš kelių galimų variantų užduotis: ekrane pateikiama

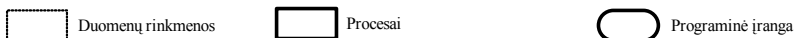
¹⁴ URL: <http://titus.fk1.uni-frankfurt.de/database/titusinx/wordwhl.asp?lxlang=&lxsublang=32807&lxxbd=*3F00*&lxop=1&C=H>.

¹⁵ Plačiau žr.: Zinkevičius 2000.

¹⁶ Plačiau žr.: Zinkevičius 2008.

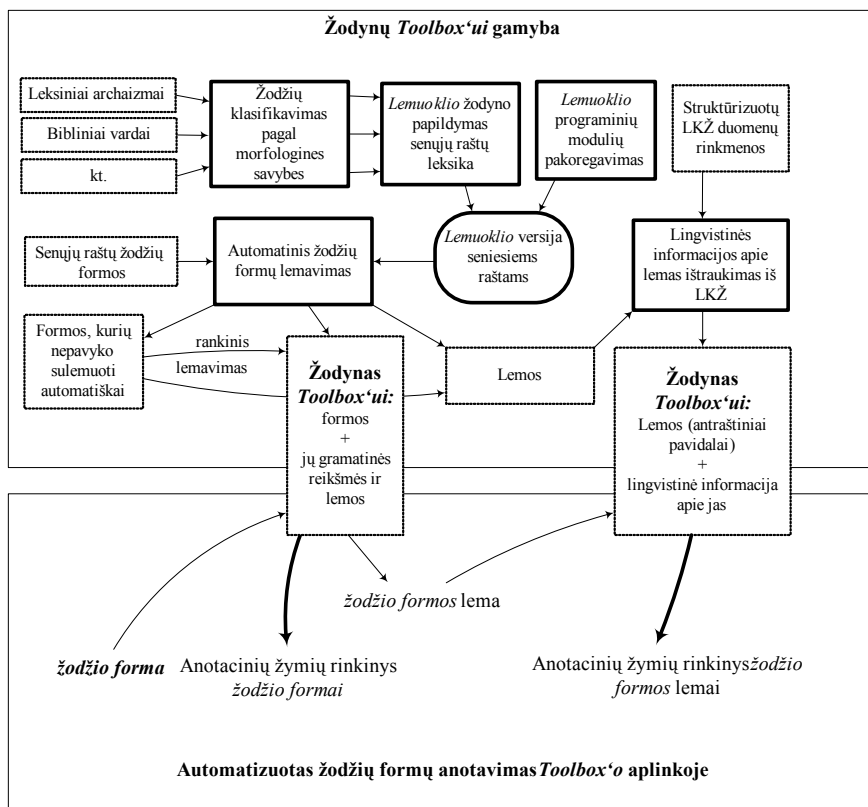
Schema: Žodžių formų anotavimo automatizacija SLIEKKE

Schemoje pažymėta:



Lemuoklis - morfologinės analizės bei lemavimo programų biblioteka

LKŽ - Lietuvių kalbos žodynas, 20 tomų



2 schema. Žodžių formų anotavimo automatizavimas SLIEKKE

informacija su skirtingais paieškos rezultatų variantais, o anotavimo operatorius, atsižvelgdamas į konkretų anotuojamą tekstą, gali pasirinkti vieną teisingą variantą. Ši pusiau automatinė žodžių anotavimo procedūra, įgyvendinama programos „Toolbox“ aplinkoje, pavaizduota 2 schemos apatinėje dalyje. Viršutinė schemos dalis vaizduoja tokiai procedūrai reikalingų žodžių formų ir lemos žodynų generavimo iš įvairių resursų procesą.

Aprašyti procesai programoje „Toolbox“ atrodo taip, kaip pavaizduota 1–3 paveiksluose.

kiami informacija apie atskirų ženklų grafinę specifiką bei apie klaidas ir korektūros sluoksnius (žr. 1 lentelę).

1 lentelė. Struktūrinių ir tekstologinių SLIEKKO pažymų sluoksniai¹⁸

Pažymų sluoksniai	Paaiškinimas
Faksimilė B	autografas, spaudinys; paženklinis lygiagretinimas su referenciniu (kritiniu) tekstu (Referencinis tekstas B) ¹⁹
Faksimilė W	autografas, spaudinys; pažodinis lygiagretinimas su referenciniu (kritiniu) tekstu (Referencinis tekstas W)
Tekstas B	tekstas, paženkliui – SLIEKKAS, generuojama automatiškai
Tekstas R	rankraščio arba spaudinio grafinės specifikos komentavimas: šrifto pasikeitimas (pvz., jeigu visas rankraštis parašytas humanistiniu kursyvu, nurodoma, kurios vietos parašytos gotikiniu kursyvu ir pan.), šrifto stiliaus ar dydžio pasikeitimas (pvz., nurodoma, jeigu pavadinimas ar kuri teksto vieta parašyta didesniu šrifto nei visas tekstas; įterpta kito stiliaus raidė spausdintame tekste), korektūros vietos, nubraukimai, pabraukimai, trynimai, rašalo spalvos kitimai, santrumpos, diakritikai, ligatūros ir t. t. – SLIEKKAS
Tekstas W	tekstas, pažodžiui – SLIEKKAS, generuojama automatiškai
Tekstas EB	aiškiai nustatytos teksto klaidos autografe, spaudinyje; paženkliui: ištaisyta forma – SLIEKKAS
Tekstas EW	klaidos autografe, spaudinyje; pažodžiui: ištaisyta forma – SLIEKKAS
Tekstas KB1a	korektūrų sluoksnis <i>taisyta iš</i> ; paženkliui: paties autoriaus taisymai rašant – SLIEKKAS
Tekstas KW1a	korektūrų sluoksnis <i>taisyta iš</i> ; pažodžiui: paties autoriaus taisymai rašant – SLIEKKAS
Tekstas KB1b	korektūrų sluoksnis <i>taisyta arba taisytina iš</i> ; paženkliui: vėlesni paties autoriaus taisymai, resp. <i>corrigenda</i> sąrašas spaudinyje – SLIEKKAS
Tekstas KW1b	korektūrų sluoksnis <i>taisyta arba taisytina iš</i> ; pažodžiui: vėlesni paties autoriaus taisymai, resp. <i>corrigenda</i> sąrašas spaudinyje – SLIEKKAS
Tekstas KB2a, KB2b, KB2c ir t. t.	korektūrų sluoksnis <i>taisyta iš</i> ; paženkliui: kitų amžininkų rankomis daryti taisymai – SLIEKKAS
Tekstas KW2a, KW2b, KW2c ir t. t.	korektūrų sluoksnis <i>taisyta iš</i> ; pažodžiui: kitų amžininkų rankomis daryti taisymai – SLIEKKAS
Tekstas KB3a, KB3b, KB3c ir t. t.	korektūrų sluoksnis <i>taisyta iš</i> ; paženkliui: kitomis rankomis daryti vėlesnių laikų taisymai – SLIEKKAS
Tekstas KW3a, KW3b, KW3c ir t. t.	korektūrų sluoksnis <i>taisyta iš</i> ; pažodžiui: kitomis rankomis daryti vėlesnių laikų taisymai – SLIEKKAS
Referencinis tekstas B	kritinio leidimo tekstas; paženkliui – SLIEKKAS, generuojama automatiškai
Referencinis tekstas R	kritinio leidimo teksto specifikos komentavimas – SLIEKKAS
Referencinis tekstas W	kritinio leidimo tekstas; pažodžiui – SLIEKKAS, generuojama automatiškai
Tg1a skyrius	SLIEKKAS, pagal preindeksaciją generuojama automatiškai
Tg1b poskyris	SLIEKKAS, pagal preindeksaciją generuojama automatiškai
Tg2a versus	poetinio teksto strofa; biblinio teksto eilutė – SLIEKKAS, pagal preindeksaciją generuojama automatiškai

¹⁸ Dabar vykdomo SLIEKKO projekto (žr. išn. 1) metu pildomi sluoksniai, prie kurių pažymėta SLIEKKAS. Kiti sluoksniai bus pildomi planuojamo tolesnio kooperacinio projekto metu.

¹⁹ Siekiant įgalinti teksto ir autografo ar spaudinio faksimilės lygiagretinimo procedūrą šiuo metu Nijmegeno MPI rengiama nauja ELAN'o programos versija.

1 lentelės tęsinys

Pažymų sluoksniai	Paaiškinimas
Ms1 lapas, puslapis	rankraštiniam originale, spaudinyje – SLIEKKAS, pagal preindeksaciją generuojama automatiškai
Ts1 puslapis Ed.1	pagrindiniame kritiniame leidime
Ts1 puslapis Ed.2	kritiniame leidime, su kuriuo lyginamas pagrindinis
Ms2 eilutė	rankraštiniam originale, spaudinyje – SLIEKKAS, pagal preindeksaciją generuojama automatiškai
Ts2 eilutė Ed.1	pagrindiniame kritiniame leidime
Ts2 eilutė Ed.2	kritiniame leidime, su kuriuo lyginamas pagrindinis
Žodžio numeris	žodžio eilės numeris visame tekste – SLIEKKAS, pagal preindeksaciją generuojama automatiškai
Referencijos numeris	referencinio žodžio, kuris susiejamas su tam tikru teksto žodžiu, eilės numeris tekste, pvz., tekste pateikta lotyniška citata ir toliau jos vertimas, taigi referencijos numeris žymi, kuris lotyniškas žodis atitinka lietuvišką vertimo žodį ir pan. – pagal preindeksaciją generuojama automatiškai

3.2. LEMAVIMAS IR GLOSAVIMAS

Visos SLIEKKO teksto žodžių formos yra lemuojamos. Pirmiausia pateikiama standartizuota teksto žodžio forma, tokia, kokia ji būtų atstatyta istoriniame žodyne (dabartiniais rašmenimis) ir antraštinis žodis, t. y. lema – pagrindinė forma istoriniame žodyne (dabartiniais rašmenimis). Toliau lema glosuojama, t. y. pateikiama atitinkama dabartinės lietuvių kalbos forma ir jos reikšmės. Taip pat glosuojama konkreti tekste pavartota žodžio forma, pateikiama jos reikšmė konkrečiame kontekste, kuri nuo lemos reikšmės gali skirtis (žr. 2 lentelę). Kitomis kalbomis (lotynų, graikų, lenkų, vokiečių) lietuviškame tekste parašytų žodžių formos taip pat lemuojamos ir glosuojamos pagal atitinkamos kalbos taisykles. Atskirame pažymų sluoksnyje žymima kiekvieno teksto žodžio kalba (pagal rankiniu būdu atliktą preindeksaciją generuojama automatiškai, pvz.: *olit, lat, gnh*). Tais atvejais, kai nėra aiškių kriterijų tiksliai nustatyti kalbą – pvz., neaišku, ar santrumpa „2 capito:“ yra lietuviška, ar lotyniška, – žymimos kelios galimos (pvz., *olit, lat*)²⁰.

²⁰ Kelios galimos pažymos viena nuo kitos atskiriamos kableliu be tarpo. Kalbos žymėjimas gali būti probleminis vardų ir onimų atveju, pvz.: „kaip S Augustinus kalba“ (WP 1573, fol. 47v,35) formą „Augustinus“ galima traktuoti kaip lietuvišką, jeigu tame pačiame tekste randama kitų *u* kamieno linksnių formų, pvz.: gen. „Augustinaus“; jeigu tokių formų nepasitaiko, gen. forma yra „Augustino“, o vardas visur tekste reguliariai kaitomas pagal *a* kamieno paradigmą (vartojamas ir nom. „Augustinas“), vardininko formą „Augustinus“ tenka anotuoti *lat* arba *lat, olit*.

2 lentelė. SLIEKKO lemavimas ir glosavimas

Pažymų sluoksniai	Paaiškinimas
Standartizuota B	standartizuota teksto žodžio forma (atstatyta kaitoma forma istoriniame žodyne): paženkliui pvz.: ž v a i z d ė s, r a š t a, m a t e r e s, G r e g o r i j u s
Standartizuota W	standartizuota teksto žodžio forma (atstatyta kaitoma forma istoriniame žodyne): visas žodis pvz.: žvaizdės, rašta, materes, Gregorijus
Lema	antraštinis žodis (lema, pagrindinė forma istoriniame žodyne) pvz.: žvaizdė, raštas, materis, Gregorijus
Glosa (lema) L	lemos glosavimas: pateikimas bendrine kalba, lemos reikšmės, vertimas (skolinių, tarmybių ir nebevertojamų, dabartiniam vartotojui nebeatpažįstamų žodžių, kitomis kalbomis parašytų žodžių atitikmuo bendrinėje kalboje) pvz.: žvaigždė, raštas, moteris, Grigalius
Glosa (lema) E	lemos glosavimas anglų kalba (ir kitomis kalbomis)
Glosa (konkreči žodžio forma) L	konkrečios žodžio formos glosavimas: pateikimas bendrine kalba, reikšmė konkrečiame kontekste pvz.: konkrečiame tekste žodis <i>žvaigždė</i> gali reikšti „Aušrinė žvaigždė, Venera“; o <i>Grigalius</i> – „Grigalius Didysis“
Normalizuota forma	konkrečios tekste pavartotos žodžio formos pateikimas dabartine lietuvių kalba pvz.: žvaigždės, rašto, moters, Grigalius
Glosa (konkreči žodžio forma) E	konkrečios tekste pavartotos žodžio formos glosavimas anglų kalba (ir kitomis kalbomis)
Kalba	standartizuota kalbos santrumpa

3.3. LINGVISTINĖ ANOTACIJA

Tolesni pažymų sluoksniai skirti lingvistinei anotacijai. Kaip ir rašant gramatiką ar žodyną, lingvistinė anotacija yra pagrįsta hierarchijos principu. SLIEKKO lingvistinėje anotacijoje informacija pateikiama sluoksniais, pradedant nuo stambiausių kategorijų ir detalizuojant iki smulkesnių fleksinės morfologijos požymių. Lingvistinė anotacija iš esmės yra morfologinė ir tam tikrais atvejais sintaksinė morfologinė (apie tokius atvejus žr. žemiau 3.3.1.). Morfologinė anotacija apima kalbos dalių (gramatinių klasių) žymėjimą (angl. *POS-tagging*) ir morfologinius žodžių formų aprašus²¹. Sudarytas visų numatytų gramatinių pažymų kompleksas SLIEKKASTS, čia pristatoma jo dalis.

²¹ Orientuojamasi į „Referenzkorpus Altdeutsch“ parengtą gramatinių pažymų sistemą (DDDTS = „Deutsch Diachron Digital-Tagset“): <<http://www.deutschdiachrondigital.de/data/home/manual/dateien/TagSetPoS.pdf>>. DDDTS yra išplėstas ir senosios vokiečių kalbos morfosintaksinei analizei pritaikytas pažymų kompleksas, kurio pagrindas – vokiečių kalbos tekstynų išplėstiniai anotaciniai principai (vad. didysis pažymų kompleksas, vok. *großes Tagset*), 1995–1999 m. sudaryti Stuttgarto ir Tübingeno universitetuose (STTS = „Stuttgart–Tübingen–TagSet“): <<http://kitt.cl.uzh.ch/clab/hilfe/stts/stts1999.pdf>>; <<http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html>>. Į DDDTS įtrauktos ir lotyniškų tekstų morfosintaksinės anotacijos pažymos, pagal kurias anotuojami lotyniški vokiškų vertimų šaltinių tekstai. Pagal DDDTS bus anotuojami lotyniški ir vokiški SLIEKKO tekstų vertimo šaltinių tekstai.

3.3.1. KALBOS DALIŲ (GRAMATINIŲ KLASIŲ) ŽYMĖJIMAS (*POS-TAGGING*)

Lemos žymimos pagal 10 gramatinių klasių.

3 lentelė. Gramatinės klasės (kalbos dalys) ir jų pažymos

1. ADJ – būdvardis (<i>adjectivum</i>)	7. N – daiktavardis (<i>nomen</i>): NA bendrinis (apeliatyvinis), NT tikrinis
2. ADV –rieveiksmis (<i>adverbium</i>)	8. P – įvardis (<i>pronomen</i>): PD demonstratyvinis, PI indefinitinis, PK klausiamasis (interogatyvinis), PPER asmeninis
3. AP – prielinksnis, polinksnis (<i>adpositio</i>)	9. PTK – dalelytė (<i>particula</i>)
4. CARD – kiekinis (<i>cardinale</i>) skaitvardis, ADJO kelintinis (<i>ordinale</i>) skaitvardis	10. V – veiksmažodis (<i>verbum</i>): VA pagalbinis, VV pagrindinis
5. ITJ – jaustukas (<i>interiectio</i>), išiktukas (<i>onomatopoeicum</i>)	
6. KO – jungtukas (<i>coniunctio</i>)	

Kiekviena iš lemos dešimties gramatinių klasių turi įvairaus detalumo subklasifikaciją, kuri priklauso nuo konkrečios pavartotos žodžio formos morfologijos. Kiekvienos žodžio formos gramatinė klasė žymima dvejopai: pagal lema (antraštinio žodžio gramatinė klasė) ir pagal konkrečiame tekste pavartotą žodžio formą.

4 lentelė. Morfologinių SLIEKKO pažymų sluoksniai

Pažymų sluoksniai	Paaškinimas
M1a lema	antraštinio žodžio (lemos) kalbos dalis (lemos POS)
M1b forma	konkrečios tekste pavartotos žodžio formos morfologinė (tam tikrais atvejais sintaksinė morfologinė) charakteristika (konkrečios žodžio formos POS)

Lemos pagrindinė gramatinė klasė atskiriama nuo konkrečios formos morfologinės charakteristikos. Lietuvių kalbos istorijai svarbu nustatyti (patvirtinti ar paneigti esančius teiginius), ar tokios dabar kaip izoliuotos suprantamos formos senuoju laikotarpiu traktuotinos kaip paradigminės, ar ne, pvz.: dabartinėje lietuvių kalbojerieveiksmiais laikomi *bematai*, *nedelsiant*, *vidunakčia*, *vidunakčiu*, *namie*, *namopri*, *trumpam*, *šalin*, *dviese* ir t. t.²² Lemos POS ir konkrečios žodžio formos POS skyrimas padeda:

- žymėti ir sekti gramatinių klasių kitimus senuoju lietuvių kalbos laikotarpiu, pvz., būdvardžių daiktavardėjimą, dalyvių būdvardėjimą bei daiktavardėjimą; daiktavardžių irrieveiksmių virtimą adpozicija (prielinksnium ar polinksnium), daiktavardžių, būdvardžių, dalyviųrieveiksmėjimą ir pan.;
- apibūdinti ambivalentiškus atvejus, kai neišku, kuriai kalbos daliai priskirti konkrečią žodžio formą, pvz., sustabarėjusių linksniųrieveiksmėjimo atveju konkreti

²² Ne visada aiškus reikšminės izoliacijos principas, pvz., *vidunakčiu* LKŽ yra lemuojamas kaiprieveiksmis, nors šios formos išsiskyrimas iš paradigmos nėra akivaizdus, plg.: *atejo vidunakčiu* (rieveiksmis) ir *džiaugiasi vidunakčiu* (daiktavardis), orieveikšminės lemos semantika nesiskiria nuo daiktavardžio lemos semantikos.

pavartota forma yra aplinkybinis daiktavardis *atėjo vidunakčiu* (NAA), o pati lema *vidunaktis* – daiktavardis (NA) (plg. 5 lentelę);

- adpozicijos (prielinksnio) klasėje skirti prielinksnius (prepozicijas) ir polinksnius (postpozicijas), pvz.: *link* (APPR) *namų* ir *namų link* (APPO); taip pat išskirti konkrečiame kontekste prielinksnėjančiusrieveiksmius, pvz.: *aukščiau* (APPR, lema *aukštai* ADV) *bambos*;

- atitinkamai žymėti sudėtinės veiksmažodines konstrukcijas (sudėtinės laikų formas). Lemų morfologinėse pažymose veiksmažodis skiriamas į pagrindinį [VV] ir pagalbinį [VA], pvz.: *Jeruzalėje buvo žmogus* (konkreti forma *buvo* VVFIN [finitinė pagrindinio veiksmažodžio forma], lema *būti* VV [pagrindinis veiksmažodis]), *tas žmogus buvo pranašas* (konkreti forma *buvo* VAFIN [finitinė pagalbinio veiksmažodžio forma], lema *būti* VA [pagalbinis veiksmažodis])²³;

- pateikti aiškų neasmenuojamų veiksmažodžio formų apibūdinimą, pvz., *kalbėdamas tarė* konkreti pavartota forma yra pusdalyvis (predikatyvinis VVSP), o lema *kalbėti* – veiksmažodis (VV). Toks lemos ir konkrečios formos skyrimas svarbus aprašant dalyvius, gerundyvus (reikiamybės formas), pusdalyvius (semiparticipijus), padalyvius (gerundijus) ir supiną; taip pat išskirti tokius veiksmažodinius būdvardžius kaip *nešinas*, *vedinas* (tarpinius variantus tarp pusdalyvio ir būdvardžio), kurių lema yra veiksmažodis (*nešti*, *vesti*).

5 lentelė. Lemos ir konkrečios žodžio formos POS pavyzdys

Pažymų sluoksniai	Pavyzdžiai		
	[atėjo] <i>trumpam</i>	[atėjo] <i>vidunakčiu</i>	<i>nepailktinai</i> [klausykima]
Lema	trumpas	vidunaktis	nepailktinai
Glosa (lema)	trumpas, neilgai trunkantis, greit pasibaigiantis	vidunaktis, vidurnaktis, nakties vidury	nenuobodžiaujant, be nuobodulio
Glosa (forma)	trumpam, neilgam laiko tarpui, trumpai	vidunaktį, vidurnaktį, nakties viduryje	nenuobodžiaujant, be nuobodulio
M1a lema	ADJ (būdvardis)	NA (bendrinis daiktavardis)	ADV (prieveiksmis)
M1b forma	ADJA (aplinkybinis būdvardis)	NAA (aplinkybinis bendrinis daiktavardis)	ADV (prieveiksmis)

Tam tikrais atvejais konkrečios formos anotacija apima ir sintaksinę morfologijos lygmenį. Sintaksinės funkcijos žymėjimas leidžia išryškinti skirtynes, kurios neatsispindi grynai morfologinėse pažymose. Tai svarbu tokiais atvejais, kai ta pati morfologinė forma turi skirtingą sintaksinę funkciją, pvz.: *pasirodė pranašas* (bendrinis

²³ Gretiminė vertimo šaltinio (lotyniško, vokiško, lenkiško teksto) morfologinė ir sintaksinė morfologinė analizė padės išskirti ir sugrupuoti tuos atvejus, kada kopula lietuviškame vertime vartojama, o kada praleidžiama. Taip pat bus matyti, ar analitinių formų ir laikų pagalbinis veiksmažodis *būti* bei *tapti* (*darytis*, *stoti* ir pan.) vartojamas priklausomai nuo vertimo šaltinio (pvz., *yra / buvo / tapo siūstas* – *missus est / erat / fuit* – *ist / war / wurde gesandt*) ir koks tokios koreliacijos laipsnis.

daiktavardis NA); *žmonės džiaugėsi pranašu* (bendriniis daiktavardis NA); *Jonas buvo pranašas / pranašu* (predikatyvinis bendrinis daiktavardis NAP). Sintaksinės funkcijos skyrimas anotuojant konkrečias formas leidžia toliau tikslingai tirti tokius reiškinius kaip įvairuojanti būtiną (vardažodinis tarinys) ir laisvojo (tarininis pažymins) predikatyvo morfologinė forma: *Jonas buvo pranašas / pranašu*; *Jonas atėjo pirmas / pirmu*. Konkrečių formų POS pažymose skiriama atributinė, predikatyvinė (tarininė) ir adverbialinė (aplinkybinė) funkcija²⁴.

3.3.2. MORFOLOGINĖ ANOTACIJA

Tolesnė morfologinė anotacija taip pat yra hierarchinė. Ji susideda iš trijų sluoksnių. Pirmas ir antras sluoksnis žymi tokias nekintamas, inherentines kalbos dalių morfologines kategorijas kaip kaitybos klasė (kaitybos kamienas) ir giminė (giminėmis nekaitomų klasių: daiktavardžių ir daiktavardiškų skaitvardžių). Kaip ir žymint kalbos dalis (POS), čia žymimos pirmiausia lemos, o tada konkrečios žodžio formos kategorijos. Toks lemos ir žodžio formos skyrimas leidžia fiksuoti kaitybos klasės ir giminės įvairavimą, pvz.:

- lema – *amžius* (ju5_Masc) ir tekste vartojamos įvairių kamienų formos *amžias* (ja_Masc), *amžas* (a_Masc), *amžis* (ja_Masc);
- lema – *gražus* pagrindinė būdvardžio forma nelyginamuoju laipsniu (*u5*, *jo8* kamienas, vyriškoji ir moteriškoji giminė), o konkreti žodžio forma *gražiausias* aukščiausiuoju laipsniu (*ja* kamienas).

Trečias sluoksnis žymi kintamus fleksinės morfologijos požymius: giminę (giminėmis kaitomų klasių), skaičių, linksnį, laipsnį, asmenį, nuosaką, laiką, rūšį (diatezę), apibrėžtumą (įvardžiuotines, arba definitines, formas).

6 lentelė. Morfologinių kategorijų anotacijos sluoksniai

Pažymų sluoksniai	Paaiškinimas	Pavyzdys	
		<i>gražiausiame</i>	<i>amže</i>
M2a fleksija lema	lemos (antraštinio žodžio) nekintamos morfologinės kategorijos (kaitybos klasė ir giminėmis nekaitomų klasių giminė)	gražus u5,jo8	amžius ju5_Masc
M2b fleksija forma	konkrečios žodžio formos nekintamos morfologinės kategorijos (kaitybos klasė ir giminėmis nekaitomų klasių giminė)	ja	a_Masc
M2c fleksija forma	konkrečios žodžio formos fleksinės morfologijos požymiai	Sup_Masc_Sg_Loc	Sg_Loc

Pavyzdžiui, pagal sudarytą SLIEKKO pažymų kompleksą daiktavardžio aprašymo tvarka ir galimų morfologinių ypatybių pažymos atrodo taip:

²⁴ Nekaitomų formų (pvz., prieveiksmių) predikatyvinė funkcija neskiriama. Pusedalyvių (semiparticipijų) ir padalyvių (gerundijų) funkcija apibrėžiama kaip „predikatyvinė arba adverbialinė“, atskirai jų nedetalizuojant.

7 lentelė. Daiktavardis (SLIEKKO pažymų kompleksas)

Daiktavardis	
M1a lema: NA (bendrasis), NT (tikrinis)	
M1b forma: NA, NAA (adverbialinis), NAP (predikatyvinis), NAT (bendrasis iš tikrinio), NATA (adv.), NATP (pred.), NT (vardas), NTA (adv.), NTO (onimas), NTOA (adv.), NTOP (pred.), NTP (vardas pred.)	
M2a fleksija lema	kaitybos klasė_giminė ²⁵
M2b fleksija forma	kaitybos klasė_giminė
M2c fleksija forma	skaičius_linksnis
kaitybos klasė	lema: a1, ja2, ja3, u4, ju5, o6, jo7, e8, i9, i10, Cn, Cr ²⁶ konkreiti forma: a, ja, u, ju, o, jo, e, i, Cn, Cr
giminė	Masc, Fem, Com
skaičius	Sg, Pl, Du
linksnis	Nom, Gen, Dat, Acc, Ins, Loc, Ad, Ill, All, Voc

3.3.3. BAZINĖ SINTAKSINĖ ANOTACIJA

Kol kas SLIEKKO projekte detali sintaksinė anotacija nėra numatyta. Planuojama pateikti tik bazinę informaciją, t. y. remiantis jau atlikta tekstų preindeksacija (TITUS) automatiškai generuoti informaciją apie sakinių ribas, taip pat tolesniame etape linijinės schemos forma žymėti finitiškumą ir sintaksinį sakinių statusą.

4. CITATŲ ANOTACIJA

Papildomi pažymų sluoksniai skirti nustatytiems citatų šaltiniams nurodyti ar pažymėti, kad atitinkamos lietuviško teksto vietos gali būti ne originalus, bet cituojamas tekstas. Citatomis šiuo atveju vadinamas eksplisicinis arba implicitinis Biblijos, perikopių, autorių citavimas, atpasakojimai, aliuzijos ir pan. Cituojama gali būti tiek originaliame, tiek verstiniame tekste. Cituotų teksto atkarpų išskyrimas leidžia žymėti ne tik originalaus ir verstinio lietuviško teksto sluoksnius, bet ir trečią sluoksnį – citatas verstiniame tekste, kuriomis verstinis tekstas papildytas vertimo ar perrašinėjimo proceso metu ir kurių vertimo šaltinyje gali ir nebūti. Citatos, neturinčios pagrindo vertimo šaltinyje, ar citatos originaliame lietuviškame tekste nurodomos pagal kritinius arba standartinius mokslinius laikomus tekstų leidimus.

²⁵ Pažymos rašomos didžiosiomis ir mažosiomis raidėmis be tarpų, viena nuo kitos atskiriamos brūkšniu apačioje „_“.

²⁶ Lemų paradigmų numeriai žymimi remiantis DLKG 1994: a1 (*vyras, miškas*), ja2 (*elnias, vėjas, galvijai*), ja3 (*brolis, arklys*), u4 (*sūnus, dangus*) ju5 (*sodžius, pavojus*), o6 (*ranka*), jo7 (*žinia, sauja, marti*), e8 (*bitė*), i9 (*širdis*), i10 (*žvėris, dantis*), Cn (=i11 *akmuo, šuo*), Cr (=i12 *sesuo, duktė*).

8 lentelė. Citatų pažymų sluoksniai

Pažymų sluoksniai	Paaškinimas
CE ekspllicitinė citata	Biblija, perikopės, autorių citatos ir kt.
CEN ekspllicitinės citatos šaltinis	jeigu įmanoma, nurodomas pagal kritinį šaltinio leidimą
CI implicitinė citata	Biblija, perikopės, autorių citatos, atpasakojimai, aliuizijos ir kt.
CIN implicitinės citatos šaltinis	jeigu įmanoma, nurodomas pagal kritinį šaltinio leidimą

5. VERTIMO ŠALTINIŲ TEKSTŲ ANOTACIJA. TEKSTŲ LYGIAGRETINIMAS

Dauguma senojo lietuvių kalbos laikotarpio ir bandomojo tekstyno tekstų (išskyrus DM 1765/75 ir atskiras WP 1573 dalis) yra vertimai iš lotyniškų, lenkiškų ar vokiškų šaltinių (Biblijos vertimų atveju bus būtina atsižvelgti ir į senuosius graikiškus bei hebrajiškus šaltinius). Siekiant nustatyti vertimo strategijas, sekti jų realizavimą kiekviename tekste ir patvirtinti ar paneigti galimą interferenciją tarp tekstų būtina lietuviškus tekstus lygiagretinti su nustatytais vertimo šaltinių tekstais. Tokio lygiagretinimo pagrindas yra vertimo šaltinio ir verstinio teksto susiejimas pažodiniu lygmeniu. Lygiagretinimas gali būti atliekamas tik tada, kai vertimo šaltinių tekstai bus suskaitmeninti, lemuoti, glosuoti ir anotuoti gramatiškai (angl. *part of speech-tagging*, morfologinė, bazinė sintaksinė informacija)²⁷. To paties žanro ir turinio lietuviški tekstai (pvz., katekizmai, Biblijos ištraukos, poteriai ir pan.) bus taip pat lygiagretinami tarpusavyje. Remiantis bandomuoju tekstynu bus lygiagretinami penki katekizmai (MžK 1547 + VE 1579 + DK 1595 + LyK 1719 + EnK 1722), jų pavyzdžiu bus renjamas ir tobulinamas lygiagretinimo procesas.

6. VIZUALIZAVIMAS. PAIEŠKOS SISTEMA

Visi SLIEKKO darbų metu parengti ir sukurti skaitmeniniai duomenys turi būti prieinami vartotojui per atvirąją prieigą (angl. *open access*). Tekstyno vizualizavimo, paieškos sistemos formulavimo ir paieškos rezultatų vizualizavimo sprendimas turi atitikti giliosios anotacijos referencinio istorinio tekstyno poreikius: užtikrinti tūkslų visų tekstuose vartojamų ženklų perteikimą ir įdiegti visus pažymų sluoksnius bei sulygiagretintus paralelius tekstus (lietuvišką–lietuvišką ir lietuvišką–nelietuvišką) apimančius paieškos įrankius.

²⁷ Vertimo šaltinių tekstams nenumatyta paleografinė ir tekstologinė anotacija.

7. PABAIGA

Modernūs kalbos istorijos tyrimai reikalauja ne tik priėjimo prie patikimų empirinių duomenų, bet ir kokybiškos jų paieškos galimybės pagal įvairius lingvistinius kriterijus. Suskaitmenintus, filologiškai užtikrintus, anotuotus ir su faksimilėmis bei vertimo šaltiniais sulygiagretintus tekstus „Senosios lietuvių kalbos tekstynas“ (SLIEKKAS) siekia paruošti tolesniems moksliniams tyrimams, visų pirma lingvistiniams, bet taip pat ir istoriniams plačiąja prasme. Jis turi tapti patikimu bei inovatyviu lietuvių kalbos istorijos tyrimų resursu ir padėti rasti atsakymus į įvairius lingvistinius klausimus. SLIEKKAS taip pat turi padėti ir pradėti įgyvendinti du didžiausius diachroninės lituanistikos darbus: senosios lietuvių kalbos gramatikos ir istorinio lietuvių kalbos žodyno rengimą.

Straipsnio priede „*Senosios lietuvių kalbos referencinis tekstynas (1500–1800)*. Anotacijos pavyzdys“ pateikiamas SLIEKKO teksto ir pažymų sluoksnių pavyzdys.

Jolanta Gelumbeckaitė, Mindaugas Šinkūnas, Vytautas Zinkevičius

OLD LITHUANIAN REFERENCE CORPUS (SLIEKKAS). A NEW CONCEPT OF A HISTORICAL CORPUS

Summary

The *Old Lithuanian Reference Corpus* (Lith. *Senosios lietuvių kalbos tekstynas*; acronym SLIEKKAS, Germ. *Referenzcorpus Altlitauisch*), a comprehensive, deeply annotated reference corpus of Old Lithuanian, is being developed in cooperation between the Goethe-University of Frankfurt am Main (Germany), the Institute of Lithuanian Language (Vilnius, Lithuania), and the University of Pisa (Italy). Its ultimate goal is to develop the linguistic and text-technological basis for the creation of a reference corpus of Old Lithuanian (1500–1800, ca. 10 m. text words) and to test it on the basis of an exemplary corpus comprising ca. 350 000 Old Lithuanian tokens. The attempt to start with a test corpus is driven by the necessity to establish complex multilayered structures that are needed for a diachronic corpus, and to apply them gradually. The envisaged annotation scheme of the Corpus embraces the following structural features:

- a thorough linguistic and textological annotation, including header information, lemmatisation, grammatical information (*Part of Speech-Tagging*, morphological and basic syntactical information), glossation (in Modern Lithuanian, English, and possibly other languages), information about the text structure (text subdivision into words, sentences, lines, verses, paragraphs etc.), palaeographic and textological information;
- a multi-level architecture of the annotations;
- multi-modality of the corpus through the alignment of the texts with facsimile reproductions of the originals.

Since most of the Old Lithuanian texts are translations from Latin, German, or Polish sources,

the source texts (in the case of the test corpus ca. 190 000 text words) will be accumulated and annotated in the same way as the Lithuanian ones. This will allow for the alignment of the Old Lithuanian texts with their sources. Furthermore, Old Lithuanian texts of the same genre will be aligned with each other in order to allow for an assessment of possible mutual influences within one and the same genre, as well as across genres.

The *Old Lithuanian Reference Corpus* will be designed to provide an innovative scientific resource for historical and comparative linguistics as well as literary, religious and cultural studies concerning the Baltic countries, including the controversy between pre-Christian and Christian cultures and the confessional spinoff processes of the area as well as their backgrounds. In this way, essential knowledge of the cultural development of Lithuania and the Baltic countries in the given period will be gained. With regard to historical linguistics, the *Old Lithuanian Reference Corpus* is expected to provide a basis for an efficient development and implementation of further research programmes concerning the diachronic grammar and the lexicon of Lithuanian.

ŠALTINIAI

Bechtel 1882 – *Bartholomäus Willent's litauische Übersetzung des Luther'schen Enchiridions und der Episteln und Evangelien, nebst den Varianten der von Lazarus Sengstock besorgten Ausgabe dieser Schriften*, hrsg. von Fritz Bechtel, *Litauische und Lettische Drucke des 16. Jahrhunderts* 3, hrsg. von Adalbert Bezzenberger, Göttingen: Robert Peppmüller, 1882.

Bezzenberger 1874 – *Litauische und Lettische Drucke des 16. Jahrhunderts* 1. *Der litauische Katechismus vom Jahre 1547*, hrsg. von Adalbert Bezzenberger, Göttingen: Robert Peppmüller, 1874.

Bezzenberger 1875 – *Litauische und Lettische Drucke des 16. Jahrhunderts* 2. II. *Der lettische Katechismus vom Jahre 1586*, III. *Das litauische Taufformular vom Jahre 1559*, IV. *Anhang: Das (angeblich altpreußische) lettische Vaterunser des Simon Grunau*, hrsg. von Adalbert Bezzenberger, Göttingen: Robert Peppmüller, 1875.

CChSL 143B – *S. Gregorii Magni Moralia in Iob* 3. *Libri XXIII–XXXV*, cura et studio Marci Adriaen, *Corpus Christianorum. Series Latina* 143B. *S. Gregorii Magni opera*, Turnholti: Typographi Brepols editores pontificii, 1985.

CPL – *Clavis Patrum Latinorum*, recludit Eligius Dekkers, Aemilius Gaar, *Corpus Christianorum. Series Latina (CChSL)*, Turnhout: Brepols, 31995.

Dini 1993 – *Heinrich Johann Lysius. Mažasis Katechizmas*, par. Pietro U. Dini, Vilnius: Mokslo ir enciklopedijų leidykla, 1993.

Donelaitis 1977 – *Kristijonas Donelaitis. Raštai*, Vilnius: Vaga, 1977.

Ford 1969 – FORD, GORDON B. JR., 1969: *The Old Lithuanian Catechism of Baltramiejus Vilentas (1579). A Phonological, Morphological and Syntactical Investigation*, Slavistic Printings and Reprintings 71, ed. by

C. H. van Schooneveld, The Hague, Paris: Mouton. Ford 1971 – *The Old Lithuanian Catechism of Martynas Mažvydas (1547)*, ed. by Gordon B. Ford Jr., Assen: Van Gorcum & Comp. N.V., 1971.

Gelumbeckaitė 2008a, b – *Die litauische Wolfenbütteler Postille von 1573, Faksimile, kritische Edition und textkritischer Apparat* 1, *Einleitung, Kommentar und Register* 2, hrsg. von Jolanta Gelumbeckaitė, *Wolfenbütteler Forschungen* 118.1–2, Wiesbaden: Harrassowitz Verlag in Kommission.

Hubatsch 1968 – HUBATSCH, WALTER, 1968: *Geschichte der evangelischen Kirche Ostpreussens* 3. *Dokumente*, Göttingen: Vandenhoeck & Ruprecht.

Jakštienė, Palionis 1995 – *Mikalojaus Daukšos 1595 metų katekizmas. Katechismus von Mikalojus Daukša vom Jahre 1595*, par. Vida Jakštienė, Jonas Palionis, Vilnius: Mokslo ir enciklopedijų leidykla, 1995.

Kalisz 2007 – *Jana Seklucjana Pieśni chrześcijańskie dawniejsze i nowe...*, wstęp i opracowanie Anna Kalisz, *Biblioteka tradycji* 62, Kraków: Collegium Columbinum, 2007.

Lebedys, Palionis 1964 – LEBEDYS, J[URGIS] / PALIONIS, J[ONAS], 1964: „Seniausias lietuviškas rankraštinius tekstas“, *Bibliotekininkystės ir bibliografijos klausimai* 3, Vilnius: Valstybinė politinės ir mokslinės literatūros leidykla, 109–135.

Luther 1529 – *D. Martin Luthers Werke. Kritische Gesamtausgabe. Schriften* 30.1. *Katechismuspredigten 1528. Großer und Kleiner Katechismus 1529*, Weimar: Hermann Böhlhaus Nachfolger, 1910.

Nesselmann 1869 – *Christian Donalitis. Littauische Dichtungen nach den Königsberger Handschriften*, hrsg. von G[eorg] H[einrich] F[erdinand] Nesselmann, Königsberg: Verlag von Hübner & Matz, 1869.

Neumann 1941 – *Studien zum polnischen frühreformatorischen Schrifttum 1. Die Katechismen von 1545 und 1546 und die Polemik zwischen Seklucyan und Maletius*, hrsg. von Friedrich Wilhelm Neumann, *Slavisch-baltische Quellen und Forschungen* 11, hrsg. von Reinhold Trautmann, Leipzig: Markert & Peters Verlag, 1941.

Panzer 1993 – *Der Kleine Catechismus D.M. Lutheri. Māžas Katgismas D. Mertino Lutteraus. Besorgt von D. Henrich Lysius*, hrsg. von Baldur Panzer, Frankfurt am Main: Peter Lang, 1993.

Ročka 1974 – *Martyras Mažvydas. Pirmoji lietuviška knyga*, [hrsg. von] Marcelinas Ročka, Vilnius: Vaga, 1974.

Rospond 1948 – *Druki mazurskie XVI w. Mazurian Printings of the 16-th Century*, [hrsg. von] Stanisław Rospond, *Prace Instytutu mazurskiego* 2, Olsztyn: Wydawnictwo Instytutu mazurskiego, 1948.

LITERATŪRA

DLKG 1994 – *Dabartinės lietuvių kalbos gramatika*, red. Vytautas Ambrazas, Vilnius: Mokslo ir enciklopedijų leidykla, 1994.

GELUMBECKAITĖ, JOLANTA, 2011: „Senosios lietuvių kalbos referencinis korpusas (SLIEKKAS). Lingvistinė anotacija“, pranešimas tarptautiniame moksliniame seminare *A. Bezenbergeris – etnografas ir akademinės baltistikos pradininkas* (Juodkrantė, Neringa 2011 09 9–10), URL: <<http://titus.uni-frankfurt.de/personal/gelumbeckaite/bezenberger.pdf>>.

GELUMBECKAITĖ, JOLANTA; JOST GIPPERT, MINDAUGAS ŠINKŪNAS, VYTAUTAS ZINKEVIČIUS, 2012: „Referenzkorpus Altlitauisch“, *Sprache als komplexes System. 34. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*, Johann Wolfgang Goethe-

Universität Frankfurt am Main (6.–9. März 2012), 290. Plakatas „Präliminarien eines Referenzkorpus Altlitauisch (1500–1800)“, URL: <<http://titus.uni-frankfurt.de/dgfs2012/poster/gelumbeckaite.pdf>>; aprašas, URL: <http://titus.uni-frankfurt.de/dgfs2012/poster/gelumbeckaite_handout.pdf>.

GIPPERT, JOST, 2005: „Baltisches Textmaterial in der TITUS-Datenbank“, *Das Baltikum im sprachgeschichtlichen Kontext der europäischen Reformation*, Tagungsband des internationalen Arbeitsgesprächs an der Herzog August Bibliothek (21.–23.05.2003 Wolfenbüttel, Deutschland), hrsg. von Jolanta Gelumbeckaitė, Jost Gippert, *Bibliotheca Archivi Lithuanici* 4, Vilnius: Lietuvių kalbos instituto leidykla, 88–110.

MARCINKEVIČIENĖ, RŪTA, 2010: *Lietuvių kalbos kolokacijos*, Kaunas: Vytauto Didžiojo universitetas. URL: <http://vddb.laba.lt/fedora/get/LT-eLABa-0001:B.03-2010-ISBN_978-9955-12-656-0/DS.001.0.01.BOOK>.

PALIONIS, JONAS, ²1995: *Lietuvių rašomosios kalbos istorija*, Vilnius: Mokslo ir enciklopedijų leidykla.

RIMKUTĖ, ERIKA, 2006: *Morfologinio daugiareikšmiškumo ribojimas kompiuteriniame tekстыne*, daktaro disertacija, Kaunas: Vytauto Didžiojo universitetas. URL: <http://fcim.vdu.lt/~erika_rimkute/straipsniai/disertacija.pdf>.

ZINKEVIČIUS, VYTAUTAS, 2000: „Lemuoklis – morfologinei analizei“, *Darbai ir dienos* 24, Kaunas: Vytauto Didžiojo universitetas, 245–273. URL: <<http://done-laitis.vdu.lt/publikacijos/zinkevicius.pdf>>.

ZINKEVIČIUS, VYTAUTAS, 2008: „The Digitization of the Dictionary of the Lithuanian Language“, *The Third Baltic Conference on Human Language Technologies (October 4-5, 2007)*, Kaunas, Vilnius: Vytauto Didžiojo universitetas, Lietuvių kalbos institutas, 349–355.

1 PRIEDAS

SENOSIOS LIETUVIŲ KALBOS REFERENCINIS TEKSTYNAS (SLIEKKAS, 1500–1800).
ANOTACIJOS PAVYZDYS

„Wolfenbüttelio postilė“, WP 1573, fol. 263r_{21–26}

21. kaip ir anfaī dak=

22. tarus rafchta f: wardu Gregorius kalba ant

23. tū ņadziū Dawida teip tardams: Mala

24. quae nos hic præmunt, ad deum ire compellunt.

25. Wargai kure mus ņcha wargina, tiei mus <←nus>

26. Diewap eit ragina.

Faksimilė B, W	kaip	ai	anfas	dak =	tarus	rafchta, si	wardu	Gregorius	kalba	ant	tu	Spađim ¹	Da w i d a
Tekstas B	kaip	ir	an fai	dak =	tarus	rafchta	wardu	Gregorius	kalba	ant	tu	f 3 a d 3 i u	D a w i d a
Tekstas R	kaip	ir	an fai	dak =	tarus	rafchta	wardu	Gregorius	kalba	ant	tu	f 3 a d 3 i u	D a w i d a
Tekstas W	kaip	ir	an fai	dak =	tarus	rafchta	wardu	Gregorius	kalba	ant	tu	f 3 a d 3 i u	D a w i d a
Tekstas KB1a													
Tekstas KW1a													
Ref. tekstas B	kaip	ir	an fai	dak =	tarus	rafchta	wardu	Gregorius	kalba	ant	tu	f 3 a d 3 i u	D a w i d a
Ref. tekstas W	kaip	ir	an fai	dak =	tarus	rafchta	wardu	Gregorius	kalba	ant	tu	f 3 a d 3 i u	D a w i d a
Standart. B	kaip	ir	an sai	daktarus	raštus	rašta	wardu	Gregorius	kalba	ant	tu	ž a d ž i u	D a w i d a
Standart. W	kaip	ir	an sai	daktarus	raštus	rašta	wardu	Gregorius	kalba	ant	tu	ž a d ž i u	D a w i d a
Lema	kaip	ir	an sai	daktarus	raštus	rašta	wardu	Gregorius	kalbėti	ant	tas	žodis	Dovydas
Glosa (lema) L	kaip	irgi taip pat	anas	daktaras, mokslininkas, išminčius	raštas	šventasis raštas, Biblija	wardas	Grigalius	kalbėti, sakyti	ant, apie	tas	žodis	Dovydas
Glosa (forma) L			ansai					Grigalius Didysis		apie			
Norm. forma	olit	olit	olit	olit	olit	rašto	švento					žodžių	Dovydo
Kalba	olit	olit	olit	olit	olit	olit	olit	olit	olit	olit	olit	olit	olit
M1a lema	KO	PTK	PD	PTK	NA	NA	ADJ	NT	VV	AP	PD	NA	NT
M1b forma	KOKOM ¹	PTK	PD	PTKE ²	NA	NA	ADJ	NT	VVFIN	APPR	PD	NA	NT
M2a fleksija			a	u4_Masc	a1	a1_Masc	a1	ju5_Masc	ėti		a	ja3_Masc	a1_Masc
lema													
M2b fleksija				u_Masc	a	a_Masc	a	ju_Masc	a-Pres		a	ja_Masc	a_Masc
forma													
M2c fleksija			Masc_Sg_Nom_3_D	Sg_Nom	Sg_Gen	Sg_Gen	Masc_Sg_Gen_D	Sg_Nom	Ind_Pres_3		Masc_Pl_Gen_3	Pl_Gen	Sg_Gen
forma													
S1a sakiny													
Msl lapas	263r				22								
M2d eilutė	21										23		
Žodžio nr.	1	2	3	4	5	6	7	8	9	10	11	12	13
Referenc. nr.													
CE ekspl. citata													
CEN ekspl. cit. šaltinis													
CI impl. citata šaltinis													
CIN impl. cit. šaltinis													

¹ KOKOM = lyginamasis jungtukas, jungia sakinio dalis arba sudėtinio prijungiamojo sakinio dėmenis.

² PTKE = dalytė, enklitinė įvardžio forma.

Faksimilė B, W	šūj	tar dam s	Mala	q3	nos	hic	praemunt	ad	deum	ire	pellunt
Tekstas B	teip	tar dam s	: Mala	q3	nos	hic	praemunt	, ad	deum	ire	pellunt
Tekstas R				A			L				A
Tekstas KB1a	teip	tar dam s	: Mala	q3	nos	hic	praemunt	, ad	deum	ire	pellunt
Tekstas KW1a											
Ref. tekstas B	teip	tar dam s	: Mala	quae	nos	hic	praemunt	, ad	deum	ire	compellunt
Ref. tekstas W	teip	tar dam s	: mala	quae	nos	hic	praemunt	, ad	deum	ire	compellunt
Standart. B	teip	tar dam s	: mala	quae	nos	hic	praemunt	, ad	deum	ire	compellunt
Standart. W	teip	tar dam s	: mala	quae	nos	hic	praemunt	, ad	deum	ire	compellunt
Lema	taip	tarti	malum	qui	ego	hic	praemere	ad	deus	ire	compellere
Glosa (lema) L	taip	tarti, sakyti	blogis, blogybė, vargas	kuris	as	čia	spausiti, slėgti, varginti	prae, iki, pas	dievas	eti	versti, vartoti prievarta
Glosa (forma) L	taip, tokiu būdu										priversti
Norm. forma		tar damas									
Kalba	olir	olir	lat	lat	lat	lat	lat	lat	lat	lat	lat
M1a lema	ADV	VV	NA	DW	PPER	ADV	VV	\$, AP	NA	VV	VV
M1b forma	ADV	VVSP	NA	DWREL ³	PPER	ADV	VVFIN	\$, APPR	NA	VVIN	VVIN
M2a fleksija lema		ti	a_Neut				C		a_Masc	irr	C
M2b fleksija forma		a	a_Neut				C		a_Masc	irr	C
M2c fleksija forma		Masc_Sg	PL_Nom	Neut_PL_Nom	PL_Acc_1		Ind_Pres_PL_3		Sg_Acc	Act_Pres	Ind_Pres_PL_3
S1a sakiny											
Ms1 lapas	263r			24							
Ms2 eilutė	23										
Žodžio nr.	14	15	16	17	18	19	20	21	22	23	24
Referenc. nr.			25	26	27	28	29	32	32	33	34
CE ekspl. citata											
CEN ekspl. cit. šaltinis											
CI impl. citata											
CIN impl. cit. šaltinis											
Mali enim bonos magis ab huius mundi desideritis expediti, dum affligit, quia dum multa eis hic uiolenta ingerunt, festinare illos ad superna compellunt.											
Gregorius Magnus, <i>Monalia in Iob</i> , l. 26, p. 13, lin. 18–21 (CChSL 143B, p. 1280 [CPL 1708])											

³ DWREL = determinatyvas, klausiamasis, santykinis (DDDDTS pažyma).

	Wargai	kurie	mus	sha	wargina	fisi	mus	Diewap	Rinwaj	cit	wargina
Faksimilė B, W											
Tekstas B	Wargai	kure	mus	fcha	wargina	tici	mus	Diewap	cit	wargina	
Tekstas R											
Tekstas W	Wargai	kure	mus	fcha	wargina	tici	mus	Diewap	cit	wargina	
Tekstas KB1a							nus				
Tekstas KW1a							mus				
Ref. tekstas B	Wargai	kure	mus	fcha	wargina	tici	mus	Diewap	cit	wargina	
Ref. tekstas W	Wargai	kure	mus	fcha	wargina	tici	mus	Diewap	cit	wargina	
Standart. B	wargai	kurie	mus	čia	wargina	tieji	mus	diewap	cit	wargina	
Standart. W	wargas	kuris	mus	čia	warginti	tas	mus	diewap	cit	wargina	
Lema	wargas	kuris	aš	čia	warginti	tas	aš	diewas	etiti	warginti	
Glosa (lema) L	wargas, nelaimė	kuris	aš	čia, šioje vietoje šioje žemėje, šiame gyvenime	warginti, slėgti	tas	aš	diewas	etiti	warginti	
Glosa (forma) L											
Norm. forma											
Kalba	olit	olit	olit	olit	olit	tie		diewop			
M1a lema	NA	PK	PPER	ADV	VV	\$, PD	olit	olit	olit	olit	
M1b forma	NA	PKREL ⁴	PPER	ADV	VVFIN	\$, PD	PPER	NA	VV	VV	
M2a fleksija	a1_Masc	ja			inti	a	PPER	NA	VVINF	VVFIN	
M2b fleksija	a_Masc	ja			a-Pres	aE	a1_Masc	a_Masc	ti	inti	
M2c fleksija	PI_Nom	Masc_PI_Nom_3	PI_Acc		Ind_Pres_3	Masc_Sg_Nom_3_D	PI_Acc	Sg_All	Inf	Ind_Pres_3	
S1a sakiny											
Ms1 lapas	263r										
Ms2 eilutė	25							26			
Žodžio nr.	25	26	27	28	29	30	31	32	33	34	
Referenc. nr.	16	17	18	19	20			21,22	23	24	
CE ekspl. citata											
CEN ekspl. cit. šaltinis											
CI impl. citata											
CIN impl. cit. šaltinis											

⁴ PKREL = įvardis, santykinis, būdvardiškas.