

## INFORMACIJA

doi: 10.15388/Baltistica.54.2.2401

### RENGINYS, SKIRTAS SENŲJŲ TEKSTŲ SKAITMENINIMUI

2019 m. lapkričio 28–29 dienomis Vilniaus universiteto Filologijos fakultete įvyko senųjų lietuviškų tekstų skaitmeninimo problemų ir perspektyvų aptarimui skirtas renginys „Senųjų lietuviškų tekstų rengimas tyrimams“. Vienas svarbiausių renginio tikslų – suburti šioje srityje dirbančius mokslininkus, kad jie pristatytų savo darbų rezultatus ar planus, pasidalytų skaitmeninimo patirtimi. Seminaro metu kilusios idėjos ir įgyti įgūdžiai padės ateityje vykdyti taikomuosius lietuvių kalbos tyrimus panaudojant skaitmenines technologijas. Modernūs filologiniai tyrimai šiais laikais yra nebeįsivaizduojami be suskaitmenintų tekstų, automatizuotų paieškų ir galimybės statistiškai apdoroti duomenis, todėl antroji renginio dalis buvo skirta praktinėms dirbtuvėms, kurių dalyviai buvo supažindinti su tarptautiniu tekstų anotavimo formatu TEI (*Text Encoding Initiative*) ir patys atliko užduotis dirbdami su šiam formatui privaloma XML (*Extensible Markup Language*) programavimo kalba – išmoko parengti struktūrines teksto anotacijas.

Pirmoji skaitmeninės humanitarikos problemoms skirta konferencija buvo surengta 2018 m. rugpjūčio 2-ąją tarptautinėje Salų kalbotyros vasaros moky-

kloje *Academia Grammaticorum Salensis Quinta Decima* – jai buvo skirta viena savaitę trunkančios mokyklos diena. Dabartinio renginio, kurį inspiravo minėtoji konferencija, tikslai buvo platesni, didesnis buvo ir dalyvių būrys – ne tik iš Lietuvos, bet ir iš Latvijos, Lenkijos, Rusijos bei Vokietijos. Pranešimus apie skaitmeninių duomenų bazių kūrimą ir tekstų anotavimą skaitė filologai, istorikai, bibliotekininkai ir programuotojai – reikšminga, kad skirtingų sričių mokslininkai rado bendrą kalbą ir dalyvavo diskusijoje.

Su skaitmeninių senųjų lietuviškų tekstų duomenų bazių kūrimu buvo susiję trys pranešėjai filologai – Jolanta Gelumbeckaitė (Frankfurto prie Maino universitetas), Gina Kavaliūnaitė (Vilniaus universitetas) ir Mindaugas Šinkūnas (Lietuvių kalbos institutas), ir du programuotojai iš Vokietijos – Arminas Hoenenas (Frankfurtas prie Maino) bei Wolfas Dieteris Syringas (Bukstehudė).

J. Gelumbeckaitė pristatė skaitmeninimo projektą, kuriam finansinę paramą 2019 m. skyrė Fritzo Thysseno fondas (Kelnas). Projekto tikslas – parengti visapusiškai anotuotą (t. y. apimantį ne tik struktūrines, bet

ir lingvistines anotacijas) Kristijono Donelaičio tekstų korpusą, pavadintą *CorDon* (<https://www.fritz-thyssen-stiftung.de/en/fundings/altilitauisch-digital-corpora-des-kristijonas-donelai-1714-1780-akronym-cordon/>). Parengtoje duomenų bazėje tyrėjas galės pateikti tekstologines, filologines ir lingvistines užklaudas, pagal kurias programa atliks analizę ir sugeneruos rezultatus. Analogiška gramatinio senųjų lietuviškų tekstų anotavimo sistema buvo sukurta 2012–2014 metais vykdant projektą SLIEKKAS (<http://titus.fkidg1.uni-frankfurt.de/sliekkas/>). Naujajame tekstų korpuse bus duomenų apie tekstų struktūrą, paleografines, tipografines ir kitas tekstologines šaltinių ypatybes, apie leksiką (leksemos pateiktos transliteruotos, standartizuotos, lemuotos bei glosuotos), gramatines tekstų ypatybes (sakinio dalis, morfologinę ir morfosintaksinę informaciją). Iš techninės (programavimo) pusės apie projektą *CorDon* kalbėjo Frankfurto prie Maino Gėtės universiteto Empirinės kalbotyros instituto programuotojas Arminas Hoenenas. Jis kėlė klausimus, kaip interneto svetainėje vartotojui pateikti įvairius lingvistinius duomenis, kurie standartai turėtų būti taikomi. Jo kuriama programa sudarys galimybę duomenų bazėje vykdyti paieškas pagal pasirinktus kriterijus, pavyzdžiui, pagal linksnį, giminę, kalbos dalį, arba šių kriterijų derinius. Tekstų korpuso anotacijos *CorDon* projekte diegiamos naudojantis ELAN formatu.

G. Kavaliūnaitė papasakojo apie šiais metais startavusią Chylinskio Biblijai skirtą interneto svetainę [\[linskibible@flf.vu.lt\]\(mailto:linskibible@flf.vu.lt\), pademonstravo jos paieškos sistemas ir kitas funkcijas. Ši interaktyvi senųjų lietuviškų tekstų interneto svetainė skirta pirmiausia lituanistams, be to, ji taip pat aktuali ir Biblijos vertimų tyrėjams. Svetainėje tyrimui galima rinktis pirmąją ar paskutinę rankraščio teksto redakciją, atskirai analizuoti vertėjo atliktus redakcinius morfologijos, leksikos, morfosintaksės ir sintaksės pakeitimus; nagrinėti žodžių formų tiesioginį, atgalinį ir dažnumo sąrašą. Joje galima vykdyti paieškas pagal Biblijos eilutę ir pagal rankraščio puslapį; pasirinktas žodis ar teksto fragmentas vienu mygtuko paspaudimu susiejamas su jo kontekstu, o kiekviena Biblijos eilutė susieta su atitinkama vieta Naujojo Testamento rankraštyje. Svetainė buvo sukurta pagal 2016–2018 m. įvykdytą LMT finansuotą projektą „Samuelio Boguslavo Chylinskio Naujasis Testamentas. Rankraščio tyrimas, faksimilinis ir interaktyvus skaitmeninis leidimas“. Jos kūrėjai – Felixas Thiesas, Bartłomiejus Kowalis, Valentinas Kuliničius, Gina Kavaliūnaitė ir Wolfas Dieteris Syringas \(programuotojas\). Projekto rezultatas taip pat buvo ir išleistas antrasis Chylinskio Biblijai skirtos serijos tomas su Naujojo Testamento vertimo rankraščio faksimile ir dviem įvadiniais straipsniais.](http://www.chy-</a></p></div><div data-bbox=)

M. Šinkūnas pristatė nuo 1996 metų kuriamą Lietuvių kalbos instituto duomenų bazę (<http://seniejirastai.lki.lt>) bei ją paremtą naują eksperimentinę paieškų sistemą (<http://sr.lki.lt>). Virtualioje senųjų tekstų saugykloje kaupiami lituanistikos bei lituanikos šaltinių diplomatiniai perrašai ir jų pagrindu parengtos konkordancijos bei indeksai.

LKI skaitmeninių tekstų rinkinys apima 425 šaltinius, iš kurių viešai prieinami 75, o trylika iš jų įdiegta į minėtą paieškos sistemą. Paieškai parengtą skaitmeninį vienetą sudaro šaltinio tekstas ir metainformacija – šaltinio ar jo fragmentų kalbos autorystė, rašymo ar spaudos vieta, laikas, bibliografinė informacija, puslapiai, spaudos lankai, turinio dalys (skyriai, giesmės, maldos, perikopės ir pan.), versetai, klaidų taisymai ir kt.

Su XX a. rašytų lietuviškų laiškų duomenynu „Mūsų laiškai“ ([www.musulaiskai.lt](http://www.musulaiskai.lt)) supažindino iš Mainco atvykusi kalbininkė Aurelija Tamošiūnaitė. Šis duomenynas yra egodokumentų kaip empirinių duomenų kaupimo iniciatyvos dalis. Dauguma sukauptų egodokumentų (laiškų, atvirukų, dienoraščių, užrašų knygelėlių, atminimų) nėra saugomi jokiose atminties institucijose, jie surinkti ir suskaitmeninti iš privačių kolekcijų. 2013 m. pradėjusiame veikti tinklapyje paskelbta daugiau nei 1 300 laiškų, atspindinčių lietuvių rašomosios kalbos vartoseną ir kaitą per visą XX amžių. Pranešėja aptarė duomenų svetainės rengimo principus, sandarą, aptarė kalbinių ir sociolingvistinių tyrimų galimybes, pasidalijo ateities planais. Tamošiūnaitė pristatė ir platesnį kontekstą – pademonstravo panašius projektus, sukurtus kitose šalyse: portugalų ir ispanų kalbų privačių laiškų tekstyną (<http://ps.clul.ul.pt/index.php>), olandų XVII–XIX a. laiškų tekstyną (<http://brievenalsbuit.inl.nl/>), Minesotos universiteto XVIII–XIX a. imigrantų laiškų internetinį archyvą (<https://www.lib.umn.edu/ihrca/dil>).

F. Thiesas kalbėjo apie Vokietijoje vykdytą projektą „Liuterio raštų poveikis Vidurio ir Rytų Europos kalboms“ („Die Ausstrahlung der Schriften Luthers auf die Sprachen Mittel- und Osteuropas“) ir sukurtą šiuo metu dar viešai neprieinamą duomenų bazę, įtraukusią kai kuriuos senuosius lietuviškus tekstus (Mažvydo, Vilento, Zengštoko ir Kleino giesmynus bei kt.). Projektas 2017–2019 m. buvo vykdytas Vokietijos Halės prie Zalės universiteto Germanistikos katedroje. Juo buvo siekiama apžvelgti protestantišką tradiciją turinčių Rytų Europos kalbų medžiagą – apibūdinti jose vyraujančias vertimo strategijas, atlikti kiekybinę analizę ir aprašyti Liuterio kalbos įtakos mastą. Didelis dėmesys buvo skirtas žodžių darybai ir sintakse, taip pat tirta ir tekstų metrinė struktūra.

Seminaro svečiai Everita Andronova (Ryga ir Sankt Peterburgas), Anta Trumpa (Ryga) ir Pēteris Vanagas (Ryga ir Stokholmas) pristatė nuo 2013 m. kaupiamą latvių kalbos senųjų tekstų korpusą bei jo pagrindą 2016 m. pradėtą rengti istorinį XVI–XVII a. latvių kalbos žodyną. Skaitmeniniame korpuse (*Latviešu valodas tekstu korpus*, [www.senie.korpuss.lv](http://www.senie.korpuss.lv)) šiuo metu prieinami 74 nevienodos apimties XVI–XVIII a. latviški tekstai (apie 1 088 486 žodžiai), jų konkordancijos, žodžių dažnių lentelės ir trumpa bibliografinė informacija. Prie dalies publikuotų perrašų teikiama ir faksimilė. Šis korpusas sudaro istorinio latvių kalbos žodyno pagrindą (*Latviešu valodas vēsturiskā vārdnīca (16.–17. gs.)*, [www.tezaurs.lv/lvvv](http://www.tezaurs.lv/lvvv)). Šiuo metu žodyno svetainėje

publikuoti 1 182 pavyzdiniai straipsniai bei sukurtos įvairios paieškos galimybės. A. Trumpa ir P. Vanagas apibūdino žodyno kūrimo principus ir jų pasirinkimo priežastis, nurodė straipsnių rašymo problemas ir jų sprendimo būdus.

Seminaro metu pranešimus skaitė ne tik duomenų bazių bei svetainių kūrėjai, bet ir jų vartotojai. Varšuvos universiteto doktorantas Pawełas Brudzyński pasakojė, kokiomis programomis ir tinklapiais jis naudojosi tirdamas senuosius lietuviškus tekstus. Tiriamosios nuosakos formų tyrimas, atliktas remiantis minėta Chylinckio Biblijos svetaine, padėjo jam nustatyti, kad išlikęs Chylinckio Naujojo Testamento vertimo rankraštis yra ankstesnio teksto redakcija. Vilma Zubaitienė kalbėjo apie tai, kaip suskaitmeninti Jacobo Quando Biblijos leidimai (1727 m., 1728 m. ir 1735 m.) ir Lietuvių kalbos institute sukurtas automatinis eilučių gretinimo įrankis padėjo nustatyti Biblijos ir Mažosios Lietuvos žodynų tarpusavio ryšius. Pranešėjai priėjo išvadą, kad kol kas dar nėra sukurta labai lengvai įvaldomų elektroninių tekstų tyrimo programų. Apibendrinamas P. Brudzyński pabrėžė būtinybę sukurti vieną didelę, visus senuosius lietuviškus tekstus jungiančią duomenų bazę su patogiai naudotis interneto sąsaja, įvairiapuse paieškų sistema ir galimybe bazėje esančius tekstus naudoti kuriant savo duomenų bazes ir analizės įrankius.

Renginyje, be filologų, dalyvavo ir istorikai bei bibliotekininkai. Lietuvos vyriausiojo archyvaro tarnybos projekto „Elektroninės paslaugos EAIS „Skaitmeninė skaityklė“ sukūrimas“ vadovas

Andrius Zilnys apibūdino archyvo informacinėje struktūroje veiksiančias paslaugas „Skaitmeninė skaityklė“, „Genealogija“ ir „Asmenvardžių archyvas“ – jos turėtų užtikrinti vieningą prieigą prie valstybės archyvuose sukauptų skaitmeninių išteklių ir didinti dokumentų prieinamumą. Egidijus Miltakis supažindino su Klaipėdos universiteto Baltijos regiono istorijos ir archeologijos instituto kuruojamais skaitmeninių istorinių dokumentų kopijų archyvais ([www.e-fontes.ku.lt](http://www.e-fontes.ku.lt)), reikšmingais lituanistiniais tyrimams. Šioje bazėje sutelkti įvairūs XIV–XXI a. Vakarų Lietuvos tyrimams svarbūs istoriniai šaltiniai iš įvairių bažnyčių, teismo aktų, dvarų ir krašto savivaldos archyvų. Nuolat pildomoje ir atnaujinamoje bazėje kaupiamos skaitmeninės šių šaltinių kopijos. Lituanistiniais tyrimams šiuo metu dažniausiai naudojama sukaupta informacija apie asmenvardžius ir vietovardžius. Lietuvos nacionalinės Martyno Mažvydo bibliotekos Dokumentinio paveldo tyrimų departamento direktorė Jolanta Budriūnienė pasakojė apie nacionalinių bibliotekų vaidmenį skaitmeninant dokumentinį paveldą bei 2018–2019 m. vykdytą Baltijos jūros regiono šalių projektą, siekusį sukurti tvarų tarptautinį ir tarpinstitucinį tinklą, užtikrinantį kultūrinio paveldo naudojimą, tyrimus, sklaidą.

Abiejų renginio dienų popietinėje dalyje buvo organizuojamos praktinės dirbtuvės, kurių metu dalyviai supažindinti su svarbiausiais dokumentais ir standartais, tarptautinėje praktikoje taikomais skaitmeninant istorinius tekstus, tai yra TEI tekstų anotavimo formatu ir

XML programavimo kalba. Dirbtuvių dalyviai atliko ir praktines užduotis – pagal pateiktą standartą anotavo tekstus struktūrinėmis žymėmis. Dirbtuves vedė programuotojas dr. Wolfas Dieteris Syringas (Vokietija) ir kompiuterine lingvistika besidomintis Vilniaus universiteto Matematikos ir informatikos fakulteto bakalaurantas Ignas Rudaitis.

Nemenkas susidomėjimas renginiu ir teigiami dalyvių atsiliepimai rodo tokio pobūdžio renginio aktualumą ir svarbą. Seminaro diskusijose buvo pasakyta mintis, kad tekstų tyrėjams būtų labai paranku, jei iki šiol skirtingais principais rengiamas lietuvių senųjų tekstų skaitmeninimas būtų suvienodintas. Ar šis renginys buvo žingsnis tokio susivienijimo link, atsakyti kol kas dar sunku. Vis dėlto galima viltis, kad humanitarai ims intensyviau domėtis ir

įsitrauks į skaitmeninimo veiklas, kurios yra būtinos moderniems filologiniams ir istoriniams tyrimams<sup>1</sup>.

Renginio organizavimas ir praktinės dirbtuvės buvo paremtos Valstybinės lietuvių kalbos komisijos (nr. K-20/2019), o dalyvių iš užsienio baltistikos centrų išlaidos padengtos iš ESFA remiamo projekto „Užsienio baltistikos centrų ir Lietuvos mokslo ir studijų institucijų bendradarbiavimo skatinimas“ (nr. 09.3.1-ESFA-V-709-01-0002).

*Gina KAVALIŪNAITĖ*

[gina.holvoet@flf.vu.lt]

*Birutė KABAŠINSKAITĖ*

[birute.kabasinskaite@flf.vu.lt]

*Ernesta KAZAKĖNAITĖ*

[ernesta.kazakenaite@flf.vu.lt]

---

<sup>1</sup> Daugiau informacijos apie renginį, tezes bei pranešimų skaidres galima rasti čia: <https://baltnexus.lt/news/841/19/Renginio-Sen%C5%B3j%C5%B3-lietuvi%C5%A1k%C5%B3-tekst%C5%B3-rengimas-tyrimams-med%C5%BEiaga>.