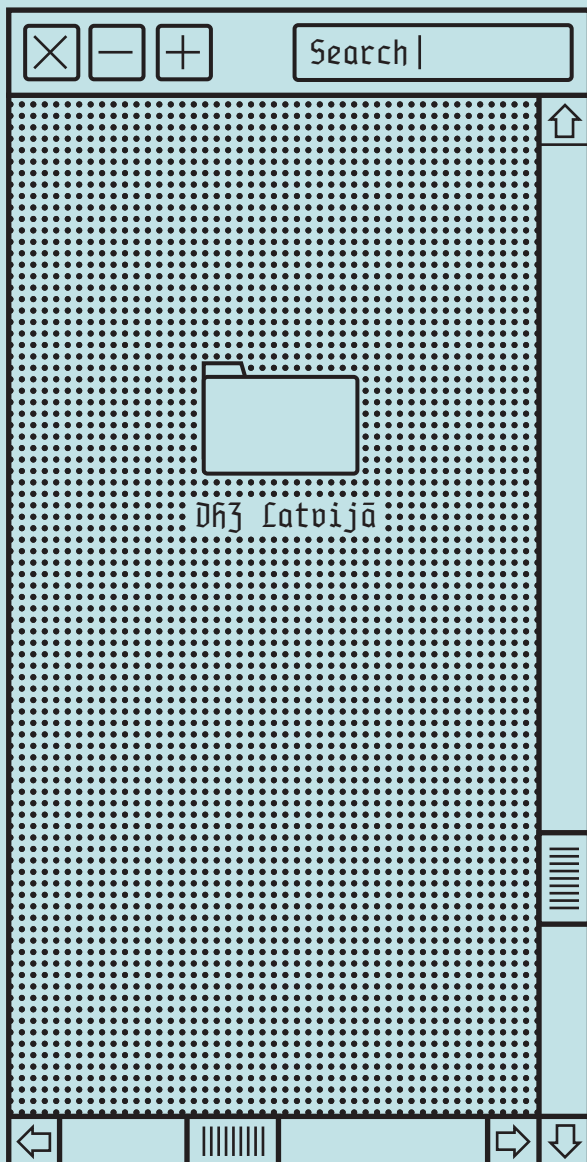


L E T

O

N



I C A



Izdevējs / Publisher

LATVIJAS UNIVERSITĀTES LITERATŪRAS,
FOLKLORAS UN MĀKSLAS INSTITŪTS / THE
INSTITUTE OF LITERATURE, FOLKLORE AND
ART OF THE UNIVERSITY OF LATVIA

Izdevniecības vadītāja / Publishing director
UNA BALODE

Redakcijas kolēģija / Editorial board
EPA ANNUSA (EPP ANNUS, Igaunija / Estonia),
UNA BERGMANE, DACE BULA, DACE
DZENOVSKA (Lielbritānija / The United Kingdom),
MĀRA GRUDULE, ZANDA GŪTMANE, JĒRGIS
HAKMANIS (JÖRG HACKMANN, Vācija / Germany),
BENEDIKTS KALNAČS, METJŪ KOTS (MATTHEW
KOTT, Zviedrija / Sweden), TOMS ĶENCIS, STELLA
PELŠE, SANĪTA REINSONE, GUNTIS ŠMIDCHENS
(ASV / The USA), PĀVELS ŠTOLLS (PAVEL ŠTOLL,
Čehija / The Czech Republic), EDĪTE TIŠHEIZERE,
GITANA VANAGAITE (GITANA VANAGAITĒ,
Lietuva / Lithuania), KĀRLIS VĒRDIŅŠ

Žurnāls *Letonica* ir iekļauts *Scopus*, *ERIH* un *EBSCO*
datubāzēs / Journal *Letonica* is included in the *Scopus*,
ERIH and *EBSCO* databases

Žurnālā ievietotie zinātniskie raksti ir anonīmi recenzēti /
Articles appearing in this journal are peer-reviewed

L E T

O N

47
2022

I C A

IEVADS

6

*Anda Baklāne, Sandis
Laime, Sanita Reinsone.*
Priekšvārds

RAKSTI

12

*Jānis Daugavietis, Agnese
Karlsonē, Ilona Kunda,
Ance Kristāla.* Latvijas
digitālo humanitāro
zinātņu rīku un resursu
izstrādāšanas prakses

52

*Sanita Reinsone, Sandis
Laime.* Latviešu folkloras
krātuves digitālais
arhīvs *garamantas.lv*:
priekšvēsture un attīstība

70

*Sanita Reinsone, Haralds
Matulis, Ilze Ļaksa-
Timinska.* Metadatos
balstīta dienasgrāmatu
teksta korpusa analīze

88

Ieva Ančevska. Liepājas
Universitātes folkloras
krājuma veidošanās
un apstrāde, digitālā

pieejamība un integrēšana
studiju un pētniecības
procesā

104

*Zita Kārkla, Eva Eglāja-
Kristsonē.* Liriskās
ģeogrāfijas: literārās
telpas kartēšana latviešu
rakstnieču romānos un
īsprozā

128

*Madara Eversone, Jānis
Oga.* Ceļojumu kartēšana
un sociālpolitiskās
tendences: Latvijas
Padomju rakstnieku
savienības biedru braucieni
20. gadsimta 70.–80. gados

150

*Anda Baklāne, Valdis
Saulespurēns.* Latento
Dirihlē sadalījumu modeļa
izmantojums laikraksta
Latvijas Kareivis tematu
analīzē: Oskara Kalpaka
gadījuma izpēte

168

*Mārīte Apenīte, Uldis
Bojārs, Ansis Garda, Anita
Goldberga, Maira Kreislere,
Anita Rašmane, Elza
Ungure.* LNB atvērtie dati
pētniecībai

188

Everita Andronova, Anna Frīdenberga, Renāte Siliņa-Piņķe, Elga Skrūzmane, Anta Trumpa, Pēteris Vanags. Variantums kā konvertācijas izaicinājums: Georga Mancela tekstu atveide mūsdienu rakstībā

208

Sanita Martena, Anna Briška, Nikole Naua. Latgaliešu valodas korpuss citu Eiropas mazāk lietoto valodu kontekstā: korpusa raksturojums, lietojums un potenciālā iespējošana

226

Angelika Juško-Štekele, Antra Kļavinska. Mūsdienu latgaliešu valodas runas korpusa izveide mazāk lietoto valodu dokumentēšanas kontekstā

244

Ilze Auziņa, Roberts Darģis, Baiba Saulīte, Normunds Grūzītis, Mikus Grasmanis, Andrejs Spektors, Kaspars Stepanovs. Specializēta latviešu valodas runas korpusa un izrunas vārdnīcas izveide vizuālās diagnostikas

izmeklējumu lingvistiskai analīzei un sistemātiskai transkribēšanai

264

Ilze Auziņa, Roberts Darģis, Inga Kaija, Kristīne Levāne-Petrova, Kristīne Pokratniece. Valodas korpusu izmantošana latviešu valodas uzdevumu automātiskā ģenerēšanā

284

Baiba Saulīte, Gunta Nešpore-Bērzkalne, Laura Rītuma, Viesturs Jūlijs Lasmanis, Normunds Grūzītis. Latviešu valodas FrameNet korpuss

298

Sanda Rapa. Vārdnīcas digitālā struktūra

312

Inguna Skadiņa, Ilze Auziņa, Roberts Darģis, Arnis Voitkāns. CLARIN valodas resursu un rīku pētniecības infrastruktūra humanitārajām un sociālajām zinātnēm

6

Anda Baklāne, Sandis Laime,
Sanita Reinsone

Žurnāla *Letonica* tematiskais numurs “Digitālās humanitārās zinātnes Latvijā” veltīts Latvijā veidotajiem humanitāro zinātņu digitālajiem resursiem, rīkiem un to pētniecībai, mēģinot sniegt priekšstatu par jomas situāciju Latvijā kopumā un šādi sperot soli ārpus datorlingvistikas jomas, kas Latvijā līdz šim bijusi vienīgā humanitāro zinātņu disciplīna, kuras paspārnē ilgoši un mērķtiecīgi izstrādāti digitāli latviešu valodas resursi, rīki un izmantotas digitālās pētniecības metodes. Liela daļa krājumā apkopoto materiālu tapuši Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi” projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība”.

Apzīmējums “digitālās humanitārās zinātnes” (angļu val. – *digital humanities*) ir salīdzinoši jauns – tas sāka iegūt popularitāti ap 2004. gadu, kad tika publicēta rokasgrāmata “*A Companion to Digital Humanities*” (“Ceļvedis digitālajās humanitārajās zinātnēs”). Tomēr digitālo humanitāro zinātņu jomai ir ap 70 gadu sena vēsture – par jomas priekštečiem var uzskatīt gan digitālo informācijas avotu un rīku izstrādi datorzinātnē un datorlingvistikā kopš 20. gadsimta 50. gadiem, gan kvantitatīvo, matemātisko un statistisko metožu un kontentanalīzes izmantojumu sociālajās un humanitārajās zinātnēs, kas sākotnēji ne vienmēr bija nepieciešami saistīts ar datortehnoloģiju izmantojumu. Neliels ieskats nozares vēsturē Latvijā atrodams šajā krājumā iekļautajā Jāņa Daugavieša un autoru kolektīva izstrādātajā rakstā “Latvijas digitālo humanitāro zinātņu rīku un resursu izstrādāšanas prakses”. Datubāzu veidošanas un avotu digitalizācijas vēstures aspekts folkloras materiālu digitalizācijas kontekstā aplūkots Sanitas Reinsones un Sanda Laimes rakstā “Latviešu folkloras krātuves digitālais arhīvs *garamantas.lv*: priekšvēsture un attīstība”, savukārt ieskats Latvijas digitālo vārdnīcu veidošanas vēsturē atrodams Sandas Rapas rakstā “Vārdnīcas digitālā struktūra”.

Digitālās humanitārās zinātnes mūsdienās ietver vairākas savstarpēji saistītas jomas, kas aptver visus datu aprites posmus – datu veidošanu, glabāšanu, analīzi, pētniecību, vizualizāciju un pētniecības rezultātu tālāknodešanu un praktisku izmantošanu. Matemātisko un statistisko metožu izmantojums joprojām ir viens no digitālās pētniecības stūrakmeņiem, tomēr mūsdienīga digitālā pētniecība ne vienmēr ir formālistiska – tā ietver gan kvantitatīvas, gan kvalitatīvas un interpretējošas pētniecības metodes, kas tiek lietotas, lai ar datorprogrammām virtuālā darba vidē pētītu digitalizētus vai digitāli radītus informācijas avotus. Par digitālā pētījuma priekšmetu var kļūt dažādi datu veidi – datubāzēs strukturēta informācija, tekstuāli avoti, attēli, skaņu ieraksti un video materiāli, telpiski objekti un vide – ikviens informācijas nesējs, kuru iespējams reprezentēt digitālā formā. Līdztekus zinātniskajai pētniecībai digitālo humanitāro zinātņu laukam piederīga ir arī plašākas sabiedrības iesaiste digitālā kultūras mantojuma iepazīšanā un veidošanā. Balstoties pētnieku darbā, tiek veidotas plašāki sabiedrībai paredzētas digitālas mantojuma kolekcijas un izstādes, savukārt sabiedrības iesaistes kampaņās ikviens, kam ir interese un nepieciešamās prasmes, var iesaistīties digitālo resursu veidošanā, piemēram, transkribējot rokrakstus, papildinot informāciju datubāzēs, atzīmējot kartē ģeogrāfiskas vietas u. c. Visbeidzot – digitāli resursi un rīki var kalpot par pamatu arī dažādās praktiskās dzīves jomās izmantojamiem tehnoloģiskiem risinājumiem – tā, piemēram, valodas digitālie resursi tiek izmantoti mašīntulkošanai, runas automatizētai transkripcijai, teksta ģenerēšanai, automātiskai kopsavilkumu veidošanai utt.

Valsts pētījumu programmas projekts “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” norisinājās no 2020. līdz 2022. gadam, un tajā bija iesaistītas astoņas institūcijas, kas veic pētniecisku un izglītojošu darbu digitālo humanitāro zinātņu jomā – Latvijas Universitātes Literatūras, folkloras un mākslas institūts (LU LFMI), LU Matemātikas un informātikas institūts (LU MII), LU Lībiešu institūts (LU LI) un LU Latviešu valodas institūts (LU LaVI), Rēzeknes Tehnoloģiju akadēmija (RTA), Liepājas Universitāte (LiepU), Latvijas Kultūras akadēmija (LKA), Rīgas Tehniskā universitāte (RTU) un Latvijas Nacionālā bibliotēka (LNB). Tas bija pirmais humanitāro zinātņu digitālajiem resursiem un to pētniecībai veltītais projekts, kas apvienoja tik lielu skaitu zinātnisko institūciju un tika īstenots ar mērķi iedibināt ciešāku starpinstitūciju sadarbību digitālo resursu un rīku izstrādes jomā – lai novērstu digitālo resursu fragmentāciju un dublēšanos, sekmētu datu risinājumu atkārtotu izmantojumu un plašāku resursu redzamību un pieejamību. Projektā tika pētītas institūciju resursu veidošanas un uzturēšanas prakses un mērķauditoriju viedoklis, lai izstrādātu ieteikumus turpmākai jomas attīstībai un resursu izstrādei. Tika papildināti un pilnveidoti tādi resursi kā Nacionālā korpusu kolekcija *korpus.lv*, Nacionālās korpusu kolekcijas instance LNB infrastruktūrā *nosketch.lnb.lv*, elektronisko vārdnīcu platforma *Tēzaurus.lv*, teksta analīzes iespēju demonstrācijas vietne “Latviešu prozas skaitītājs”, Latviešu folkloras krātuves digitālais arhīvs *garamantas.lv*, LU LFMI digitālais arhīvs *dati.lulfmi.lv*, Latvijas un latviešu literatūras vortāls *literatura.lv*, Tautas vietvārdu datubāze *vietvardi.lv* un Lībiešu digitālo resursu platforma *Livonian.tech*. Ievērojama daļa projekta finansējuma tika novirzīta arī izglītojošām darbībām – lekciju kursu īstenošanai, vieslektoru piesaistei un izglītojošu materiālu izstrādei.

Lai arī šis krājums veidots ar mērķi sniegt priekšstatu par digitālo humanitāro zinātņu situāciju Latvijā, tajā netiek sistemātiski aprakstītas katras apakšnozares vēsture un aktuālās problēmas – drīzāk esam izveidojuši momentuzņēmumu. Raksti sniedz ieskatu pētījumos un ar resursu izstrādi saistītās darbībās, kas projektā “Humanitāro zinātņu digitālie resursi” iesaistītajām institūcijām bijušas aktuālas 2021. un 2022. gadā (ņemot vērā projekta izteikto orientāciju uz resursu izstrādi, vairāk rakstu veltīts tieši iepazīstināšanai ar resursiem, mazāk – resursu pētniecības rezultātiem). Šajā momentuzņēmumā pagaidām nav iekļuvuši pārspriedumi, kas saistīti ar attēlu, skaņas, multimediju, objektu, ēku vai telpas pētniecību – visi raksti veltīti datubāzu, pētniecības infrastruktūras, tekstuālu resursu vai transkribētu runas resursu izveidei vai pētniecībai. Tātad Latvijas digitālo humanitāro zinātņu jomas aptveršanas darbs ir tikai aizsācies, un to nepieciešams turpināt.

Vispārīgāko pārskatu par kopējo ainavu sniedz Jāņa Daugavieša, Agneses Karlsones, Ilonas Kundas un Ances Kristālas raksts “Latvijas digitālo humanitāro zinātņu rīku un resursu izstrādāšanas prakses”, kurā resursu izstrādes pieredzes izvērtēšana veikta, analizējot desmit zinātnisko iestāžu paspārnē veidotu resursu veidotāju paustos viedokļus. Ja minētais pētījums piedāvā vispārīnātu pieredzes apkopojumu, nepakavējoties pie individuālu resursu izstrādes niansēm, tad pārējie piecpadsmit krājuma raksti veltīti konkrētu resursu vai resursu kopu detalizētiem aprakstiem un pētījumiem, kas veikti, analizējot šos resursus.

Trīs publikācijas veltītas folkloras digitālajiem resursiem. Sanita Reinsone un Sandis Laime rakstā “Latviešu folkloras krātuves digitālais arhīvs *garamantas.lv*: priekšvēsture un attīstība”

apkopojusi informāciju par latviešu folkloras materiālu digitalizācijas vēsturi un LU LFMI Latviešu folkloras krātuves digitālā arhīva tapšanas posmiem, rādot evolūciju no vēsturisku fragmentāru datubāzu integrācijas līdz multifunkcionālai platformai, kas apvieno vairāku institūciju resursus un kuras nozīmīga sastāvdaļa ir lietotāju līdzdalības moduļi un kartēšanas rīki; rakstā detalizēti skatīti digitālā arhīva uzbūves principi, metadatu tipi un kolekcijas. Rakstā “Metadatos balstīta dienasgrāmatu teksta korpusa analīze” Sanita Reinsone, Haralds Matulis un Ilze Ļaksa-Timinska sniedz pārskatu par Latviešu folkloras krātuves autobiogrāfiju digitālās kolekcijas veidošanas vēsturi un korpusa izveides principiem; tiek ieskicēti iespējamie šī korpusa pētniecības virzieni un iezīmētas dažādas grūtības un risināmi jautājumi, ar kuriem var nākties saskarties dienasgrāmatu datu apstrādes procesā. Savukārt Ieva Ančevska rakstā “Liepājas Universitātes folkloras krājuma veidošanās un apstrāde, digitālā pieejamība un integrēšana studiju un pētniecības procesā” dokumentē LiepU folkloras materiālu krājuma tapšanu un digitalizāciju; LiepU folkloras krājuma materiāli arī tikuši iekļauti *garamantas.lv* digitālajā arhīvā.

Literatūras digitālajiem resursiem un metodēm veltītas divas publikācijas – Zitas Kārklas un Evas Eglājas-Kristones raksts “Liriskās ģeogrāfijas: literārās telpas kartēšana latviešu rakstnieču romānos un īsprozā” un Madaras Eversones un Jāņa Ogas raksts “Ceļojumu kartēšana un sociālpolitiskās tendences: Latvijas Padomju rakstnieku savienības biedru braucieni 20. gadsimta 70.–80. gados”. Abi pētījumi saistīti ar vienu no LU LFMI specializācijas jomām – digitālās kartēšanas metožu izmantojumu. Kārklas un Eglājas-Kristones raksts sniedz izsmelšu ieskatu literārās ģeogrāfijas metodoloģijas lietojumā pasaulē un demonstrē Sieviešu prozas korpusa (1872–1940) kartēšanas rezultātus. Savukārt Eversone un Oga pievēršas Rakstnieku savienības biedru braucieni (1970–1989) datu kopas analīzei un vizualizācijai kartēs; izklāsta gaitā uzskatāmi aprakstīts *literatura.lv* un *Google Maps* piedāvāto kartēšanas rīku izmantošanas process.

Digitālās vēstures tematiku piesaka Andas Baklānes un Valda Saulespurēna raksts “Latento Dirihlē sadalījumu modeļa izmantojums laikraksta *Latvijas Kareivis* tematu analīzē: Oskara Kalpaka gadījuma izpēte”. Pētījums veikts, lai izmēģinātu automātiskas tematu noteikšanas (tematu modelēšanas) metodoloģiju vēsturisko laikrakstu datu pētniecībai; izmēģinājumam izmantots laikraksta “Latvijas Kareivis” teksta korpusa, kas veidots no LNB digitalizētās periodikas kolekcijas materiāliem.

Ieskatu LNB digitālo resursu klāstā sniedz arī Mārītes Apenītes, Ulda Bojāra, Anša Gardas, Anitas Goldbergas, Mairas Kreisleres, Anitas Rašmanes un Elzas Ungures raksts “LNB atvērtie dati pētniecībai”, kurā sniegta informācija par nacionālās bibliogrāfijas, Nacionālā tēzaura un digitālās kolekcijas “Rainis un Aspazija” datu kopām, vienlaikus skaidrojot jēdzienus un tehnoloģijas, kas saistošas šo datu kopu kontekstā: atvērtie dati, saistītie dati, datu anotēšanas un entīšu veidošanas principi un mērķi, kontrolēto un strukturēto metadatu vārdnīcu veidi un lietojums.

Everita Andronova, Anna Frīdenberga, Renāte Siliņa-Piņķe, Elga Skrūzmane, Anta Trumpa un Pēteris Vanags rakstā “Variantums kā konvertācijas izaicinājums: Georga Manceļa tekstu atveide mūsdienu rakstībā” skar aktuālu tematu visiem humanitāro zinātņu pētniekiem, kuri strādā

ar avotiem, kas iespiesti gotiskajā rakstā. Gotisko rakstzīmju atveide digitālā formā pati par sevi saistīta ar virkni risināmu jautājumu, savukārt, lai nodrošinātu plašāku pieejamību un daudzveidīgākas teksta pētniecības iespējas, lietderīga ir teksta modernizācija. Balstoties Georga Manceļa korpusa piemērā, raksts vēsta par gotiskā teksta konvertāciju mūsdienu rakstībā.

Krājumā iekļauti pieci raksti, kas veltīti dažādu valodas korpusu izveidei un izmantojumam. Divas publikācijas veltītas latgaliešu valodas korpusiem – mūsdienu latgaliešu tekstu korpusam MuLa un pirmā mūsdienu latgaliešu runas korpusa MuLaR izveidei. Sanita Martena, Anna Briška un Nikole Naua rakstā “Latgaliešu valodas korpusu citu Eiropas mazāk lietoto valodu kontekstā: korpusa raksturojums, lietojums un potenciālā iespējošana” vēsta par MuLa 2012. un 2022. gada versiju veidošanas vēsturi, principiem un saturu, kā arī sniedz ieskatu korpusa lietotāju aptaujas rezultātos, iezīmējot iespējamās korpusa izmantojuma veidus nākotnē; rakstā vērsta uzmanība arī uz Eiropas mazāk lietoto valodu korpusu problemātiku kopumā. Angelika Juško-Štekele un Antra Kļavinska rakstā “Mūsdienu latgaliešu valodas runas korpusa izveide mazāk lietoto valodu dokumentēšanas kontekstā” aplūko tādas problēmjautājumus kā dažādu izlokšņu atspoguļošana dialekta korpusā, spontānas runas dokumentēšana, runas ierakstu transkribēšana un runas korpusa lietošanas platformas izvēle.

Ilze Auziņa, Roberts Dargis, Baiba Saulīte, Normunds Grūzītis, Mikus Grasmanis, Andrejs Spektors un Kaspars Stepanovs rakstā “Specializēta latviešu valodas runas korpusa un izrunas vārdnīcas izveide vizuālās diagnostikas izmeklējumu lingvistiskai analīzei un sistematiskai transkribēšanai” stāsta par specializēta medicīnas runas korpusa un vārdnīcas izveidi. Šajā gadījumā korpusa veidošanas galvenais mērķis nav diagnostikas valodas pētniecība – rakstā aplūkoti valodas resursi tiek veidoti ar mērķi izstrādāt tehnoloģiju, kas ļaus automātiski atpazīt un sintezēt runu medicīnas jomas vajadzībām.

Praktiskas valodas korpusu datu izmantošanas iespējas tiek aplūkotas arī Ilzes Auziņas, Roberta Dargā, Ingas Kaijas, Kristīnes Levānes-Petrovas un Kristīnes Pokratnieces rakstā “Valodas korpusu izmantošana latviešu valodas uzdevumu automātiskā ģenerēšanā”. Valodas apguvēju korpusi ir izmantojami, lai analizētu apguvēju tipiskās pieļautās kļūdas, savukārt līdzsvarotie, paralēlie u. c. valodas korpusi izmantojami, lai automātiski ģenerētu piemērus valodas apguves uzdevumiem. Rakstā sīkāk analizēts valodas apguvēju korpusa LaVA, kā arī sniegti izvērsti uzdevumu ģenerēšanas piemēri.

Baiba Saulīte, Gunta Nešpore-Bērzkalne, Laura Rituma, Viesturs Jūlijs Lasmanis un Normunds Grūzītis rakstā “Latviešu valodas *FrameNet* korpusu” vēsta par darbu, kas veikts, veidojot *FrameNet*-LV datu kopu, kurā, balstoties freimu semantikas teorijā, detalizēti tiek marķētas leksisko vienību semantiskās lomas. Rakstā aprakstīts un ar piemēriem ilustrēts *FrameNet* konceptuālais modelis, un tiek uzskatāmi parādīts latviešu *FrameNet* korpusa semantiskās marķēšanas darba process; lai šis valodas resurss būtu pieejams pētniekiem, ir izstrādāta arī tīmekļvietne.

Sanda Rapa rakstā “Vārdnīcas digitālā struktūra” pievēršas elektroniskās leksikogrāfijas – elektronisko vārdnīcu un datubāzu sastādīšanas – problemātikai. Autore reflektē par

atšķirībām starp iespīstām un elektroniskām vārdnīcām, raksturo dažādus elektronisko vārdnīcu strukturēšanas veīdus, to priekšrocības un ierobežojumus, kā arī snīdz ieskatu latviešu valodas elektronisko vārdnīcu veidošanas vēsturē un tagadnē.

Daudzveīdīgo valodas resursu pārskatu noslēdz Ingūnas Skadiņas, Ilzes Auzīņas, Roberta Darģa un Arņa Voītkāna publikācija “CLARIN valodas resursu un rīku pētnīecības infrastruktūra humanitārajām un sociālajām zinātnēm”. Autori uzsver starptautiskās valodas digitālo resursu infrastruktūras CLARIN nozīmi un iepazīstīna lasītāju ar CLARIN ERIC un CLARIN Latvījas mezglpunkta CLARIN–LV piedāvātajām iespējām. Demonstrējot infrastruktūras funkcionalitātes, kā piemēri aplūkoti CLARIN jau iekļautie Latvījā veīdotie digitālie valodas resursi.

Latvijas digitālo humanitāro zinātņu rīku un resursu izstrādāšanas prakses

Jānis Daugavietis, Agnese Karlsons,
Ilona Kunda, Ance Kristāla

Raksts izstrādāts ar projekta “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (VPP-IZM-DH-2020/1-0001) atbalstu. Izsakām pateicību visiem intervētajiem ekspertiem, kā arī kolēģiem (īpaši Ingūnai Skadiņai un Sanitai Martenai) un recenzentam par vērtīgiem komentāriem un norādēm darbā pie manuskripta.

Ievads

Raksta mērķis ir sniegt Latvijas digitālo humanitāro zinātņu resursu un rīku (turpmāk – DH RR) veidošanas un uzturēšanas prakšu izvērtējumu no to veidotāju perspektīvas un pieredzes, vienlaikus to kritiski salīdzinot ar teorētiskās literatūras un tajā aprakstītās ārvalstu pieredzes secinājumiem par DH RR veidošanu, uzturēšanu un ilgtspējības nodrošināšanu. Mērķa sasniegšanai izvirzīti trīs pētnieciskie jautājumi: (1) kāda ir Latvijas DH RR izstrādātāju un uzturētāju pieredzes resursu veidošanā; (2) kāds ir viņu vērtējums par dažādiem šīs prakses aspektiem; (3) ko viņi uzskata par labu DH RR?

Viena no digitālo humanitāro zinātņu (turpmāk – DH) darbības jomām ir tādu digitālo resursu un rīku izstrāde, kas var būt paredzēti gan zinātniekiem, gan plašākai sabiedrībai vai noteiktām tās grupām. Tie ir runātās un rakstītās valodas tekstu korpusi; mašintulkošanas, runas tehnoloģijas, stilometrijas, sentimenta analīzes, mašīnmācīšanās, 3D modelēšanas, sociālo tīklu analīzes rīki; vārdnīcas, enciklopēdijas; audiovizuālo materiālu arhīvi un rīki; spēles u. c.¹ Tie kalpo kā zināšanu pārnesei (no vienas akadēmiskās kopienas uz citu, no pētniekiem vai sabiedriskajiem zinātniekiem (*citizen scientists*) – sabiedrībai), tā arī jaunu zināšanu radīšanai, tāpēc tie ir svarīgi ne tikai zinātnei, bet tiem var būt arī nozīmīga loma dažādu sabiedrisko (sociālo, ekonomisko, politisko, kultūrpolitisko) mērķu īstenošanā. Tātad DH RR izstrāde ir gan viena no DH lauka zinātniskajām pamataktivitātēm (Thoden et al. 2017), gan prakse, kas iziet ārpus tās (Schreibman, Hanlon 2010; Poole 2017b). Līdz ar to tā daudz nepastarpinātāk nekā daudzas citas zinātnes mijiedarbojas ar sabiedrību, un daudzos gadījumos DH RR tiek izstrādāti tieši viņiem – nezinātniekiem, sabiedrībai.

Latvijā izpēte digitālo humanitāro zinātņu laukā aizsākās 20. gs. 50. gadu beigās datorlingvistikas laukā (Skadiņa 2021). Folkloristika šajā laukā sāk iekļauties 70. gadu vidū latviešu Rietumu diasporā

1 Ir dažādas DH RR taksonomijas. Piemēram DARIAH rīkus klasificē šādi: komunikācijai un sadarbībai, datu analīzei, datu ieguvei, datu publicēšanai un izplatīšanai, datu strukturēšanai un uzlabošanai, empīriskai pētniecībai, projektu vadībai (Dallas et al. 2017: 156). DH RR var dalīt reprezentācijas, programmēšanas, analīzes, arhivēšanas un daudzos citos veidos. Par tipiskiem DH rīkiem un resursiem var uzskatīt, piemēram, vārdnīcas, tekstu korpusus, tekstu analīzes un vizualizācijas rīkus, datubāzes, digitālos arhīvus, multimediju kolekcijas, kodēšanas un programmēšanas resursus, spēles, metadatu veidošanas rīkus, sociālos tīklus, mācību platformas, raidierakstus, skaņas apstrādes programmatūras utt.

(Vikis-Freibergs, Freibergs 1978) un 80. gados Padomju Latvijā (Klotiņš, Muktupāvels 1985; Pūtelis 2021; Skadiņa 2021; Spektors 2022), kad tiek veidoti pirmie elektroniskie daņu korpusi. 80. gadu sākumā Latvijā tiek veidotas pirmās elektroniskās vēstures, valodniecības un literatūrzinātnes bibliogrāfiskās datubāzes (Prikulis 1982; Bērziņš un Grauda 1987; Bērziņš 1988). Lai turpinātu attīstīt mašintulkošanu, elektroniskās vārdnīcas, programmrīkus valodas un runas automatizētai apstrādei, 90. gadu sākumā paplašinās latviešu valodas datorfondu (dažādu korpusu) izveide (Spektors 1999; Skadiņa un Spektors 2009), kā arī rodas jaunas nozaru datubāzes, piemēram, Latvijas rakstnieku un literātu biogrāfiskā un bibliogrāfiskā datubāze (Briedis 1992; Rožkalne 1992b; 1992a; 2021).

Mūsdienās Latvijā tādās institūcijās kā LU Matemātikas un informātikas institūts (LU MII) un tā Mākslīgā intelekta laboratorija (AiLab), LU Literatūras, folkloras un mākslas institūts (LU LFMI), Latvijas Nacionālā bibliotēka (LNB), SIA "Tilde" u. c. ir izveidoti un darbojas daudzi labi attīstīti DH rīki un resursi. Tomēr kopumā Latvijas DH joprojām ir salīdzinoši jauns zinātniskās, sociālās un ekonomiskās darbības lauks, lai gan darbs pie pirmo rīku un resursu izstrādes, kurus šodien dēvējam par pieredzīgiem DH laukam, ir sāksis pirms dažām desmitgadēm. Pirmie resursi ir beiguši pastāvēt vai veiksmīgi transformējušies un iekļāvušies modernākās digitālajās infrastruktūrās, bet pieredze un institucionālā atmiņa turpina būt klātesoša mūsdienu praksēs.

Šajā rakstā stingri nenošķir DH "rīkus" un "resursus", šos terminus lietojot kopā vai kā sinonīmus, lai gan tie nav par tādiem uzskatāmi. Ar "DH rīku" jeb instrumentu (*tool*) visbiežāk saprot kādu programmatūru vai kodu, kas paredzēts diezgan specifisku uzdevumu veikšanai (piemēram, attēla vai teksta apstrādei), savukārt "DH resurs" ir plašāks jēdziens. Tā var saukt datu avotus, krājumus, kolekcijas, korpusus, kas ievietoti kādā IT sistēmā kopā ar līdzekļiem, kas ļauj darboties ar šīm vienībām vai nu sistēmā vai izgūstot tās. Par resursiem var saukt arī tādās IT sistēmas, kas tikpat labi varētu tikt dēvētas par digitālajām infrastruktūrām.

Raksta fokuss ir nevis konkrētu rīku vai resursu tehnoloģiskās funkcionēšanas, bet gan visas ekosistēmas analīze. Tāpēc raksts nepievēršas DH rīku un resursu jēdzienu striktai definēšanai. Lai DH rīki un resursi varētu tikt efektīvi izmantoti akadēmisko vai sabiedrisko uzdevumu veikšanai, tiem jābūt labiem, t. i., tādiem, ar kuriem var veikt tiem paredzētos uzdevumus, un tādiem, kurus izmanto tam paredzētās mērķauditorijas. Analizējot mūsdienu Latvijas DH RR veidošanas un uzturēšanas prakses un pieredzes, iespējams saprast, kā notiek šī lauka attīstība, kas palīdz un traucē arvien veiksmīgāku rīku un resursu izveidei. Tam veltīts šis raksts.

Liela mērā šis ir skats uz Latvijas DH it kā no malas. Lai arī bieži tiek apgalvots, ka DH ir starpdisciplinārs lauks, parasti tiek nosaukti tikai divi tā kopēji: humanitāro un datorzinātņu pārstāvji: "tehnoloģiju un humanitāro zinātņu laulība" (Klein 2018: 5). Tomēr nereti šis "(..) populārais DH raksturojums kā divu kultūru – humanitāro zinātņu un tehnoloģiju – savienojums aizēno to, cik liela nozīme ir humanitāro zinātņu un trešās kultūras – sociālo zinātņu – robežu šķērsošanai" (turpat: 25).

Lai atbildētu uz sākumā izvirzītajiem pētnieciskajiem jautājumiem, veikta teorētiska un empīriskā analīze. Pirmkārt, iepazīta attiecīgā akadēmiskā literatūra, veikta teorētisku modeļu izvēle un adaptācija; otrkārt, īstenotas Latvijas DH RR izstrādātāju un satura veidotāju interneta aptauja un daļēji strukturētas intervijas. Iegūto datu apstrāde strukturēta atbilstoši šādiem teorētiskās un empīriskās analīzes uzdevumiem:

1. izveidot teorētisku DH RR ekosistēmas modeli;
2. aprakstīt tipiskās Latvijas DH RR veidošanas prakses, aptverot visus galvenos procesa posmus, sākot ar ideju un beidzot ar uzturēšanu un ilgtspējas nodrošināšanu;
3. analizēt Latvijas DH RR veidotāju un uzturētāju priekšstatus par dažādiem šīs prakses aspektiem, tostarp jautājot, kāds būtu uzskatāms labs DH RR gan pētniekiem, gan plašākai publikai.

Teorētiskā pieeja un analīzes modeļu izvēle

Šajā nodaļā veikta akadēmiskās literatūras analīze ar mērķi adaptēt un veidot atbilstošākos teorētiskās analīzes modeļus turpmākajai empīrisko datu analīzei. Nodaļa dalīta divās apakšnodaļās: (1) DH RR konceptualizēšana, izmantojot kultūras ražošanas pieeju un no bioloģijas pārņemto jēdzienu “ekosistēma”; (2) svarīgāko DH RR veidošanas un uzturēšanas procesa aspektu izvēle.

DH RR kā ekosistēmas daļa

DH RR konceptualizācijā visbiežāk izmantots šādu četru pamata komponentu modelis: saturs, lietotāji, mēdžments, izplatīšana². Tas norāda, ka DH pētniecībā un praksē DH rīki un resursi tiek skatīti paplašinātā kontekstā. Tai pašā laikā šādā modeli it kā aizmirsts svarīgs posms – paša DH RR izstrāde un tā izstrādātāji.

Līdz šim DH RR pētniecībā visbiežāk tiek pētīti lietotāji neatkarīgi no tā, vai tie ir profesionāļi, pieredzējuši amatieri vai iesācēji (Walsh et al. 2016). Diezgan daudz zināms par dažādām DH RR lietotāju grupām, tostarp tik specifiskām kā digitālie talkotāji (*crowdsourcers*). Salīdzinoši daudz pētīti arī humanitāro zinātņu profesionāļi kā DH RR izmantotāji (sk. Gibbs, Owens 2012; Warwick 2012; Green, Courtney 2015 u. c.). Tomēr dažādu iemeslu dēļ paši rīku un resursu izstrādātāji un arī pats izstrādes process ir visai maz pētīti, lai gan tiem ir nozīmīga loma DH RR veidošanā un attīstībā (Given, Willson 2018).

2 The LAIRAH Digital Humanities Checklist, 2007 University College London. Pieejams: <https://www.ucl.ac.uk/infostudies/research/circah/lairah/features> [skatīts 15.06.2021.].

Var pieņemt, ka Latvijā situācija ir līdzīga gan DH RR (jo kas gan cits veidos DH resursus, ja ne paši DH vai, plašāk raugoties, humanitāro zinātņu pētnieku kopienas pārstāvji), gan šī procesa izpētē. Latvijā DH RR veidotāji un veidošana nav pētīta gandrīz nemaz³, un tas ir iemesls, kāpēc šī raksta mērķi un uzdevumi ir diezgan plaši un vispārīgi⁴.

Kristina Šteinere (*Christina Steiner*) ar kolēģiem pieejas DH RR analīzē grupē trīs galvenajos tipos: lietotāju orientētas, sistēmorientētas un sistēmiskas (Steiner et al. 2014: 58). Rakstā izvēlēta trešā pieeja, balstoties sociologu Ričarda Petersona (*Richard A. Peterson*) kultūras ražošanas (*production of culture*) pieejā un Pjēra Burdjē (*Pierre Bourdieu*) lauku teorijā. Tās ir tuvas pieejas, kas sociālo realitāti dala segmentos (sistēmās, laukos), kuru centrā ir kāda specifiska ekonomiskā, kultūras vai politiskā ražošana. Šo ražošanu veic konkrēti indivīdi, kurus iegrožo un vada ne tikai dažādas sociālās normas, vērtības un sociālie institūti (ģimene, baznīca, valsts u. c.), bet arī dažādas institūcijas (organizācijas, uzņēmumi, asociācijas utt.). Tātad šos segmentus veido indivīdi, indivīdu grupas, sociālie institūti un konkrētas institūcijas un attiecības starp tiem. Komunikācija katrā segmentā notiek vairākos līmeņos un atšķirīgos mērogos, tāpēc to analīzē nepieciešama gan mikroanalīze, gan mezoanalīze. Tāpat jāņem vērā laika dimensija, jo nemitīgi notiek visu iesaistīto elementu transformācijas, un konkrētā brīža situācijas izprašana reti kad iespējama, nezinot vēsturisko trajektoriju.

“Ieliekot” DH rīkus un resursus šādā plašākā kontekstā, tā turpmāk dēvēta par ekosistēmu. Tā ietver gan pašus rīkus un resursus (“nedzīvā apkārtējā vide [biotops]”, runājot bioloģijas valodā, no kurienes pārņemts šis termins), gan to izstrādātājus, uzturētājus, lietotājus un citus tieši vai pastarpināti iesaistītos indivīdus un to grupas (“dzīvie organismi [biocenoze]”). Atsevišķs šīs DH RR ekosistēmas parametrs ir šo elementu savstarpējās saites un komunikācija, kā arī institūti un institūcijas, bez kuru klātbūtnes nebūtu iespējama indivīdu sadarbība.

No sociālās teorijas viedokļa raugoties, DH RR ir instruments kādam, kurš ar to vēlas īstenot savas intereses un apmierināt savas vajadzības. Tāpēc DH RR ekosistēmas modeļa centrā ir ne tikai pats rīks vai resurss, bet arī aģenti, kas to veido, uztur un izmanto. Ap DH RR grupējas tā pasūtītāji, finansētāji, izstrādātāji, uzturētāji un lietotāji. Tie var būt gan atsevišķi indivīdi, gan to grupas, organizācijas vai institūcijas. Aģentu prakses un to rezultātus ietekmē resursi to rīcībā (kapitāli, zināšanas, spējas), tāpēc arī tie ir būtiski parametri, kas nosaka gan konkrēta DH RR kvalitātes un likteni, gan arī plašāku DH ekosistēmu funkcionēšanu. Cita svarīga ekosistēmas pazīme, līdz ar to arī analīzes objekts, ir tā nemitīgā mainība laikā, tāpēc autori savā modeli izdalījuši divus būtiski atšķirīgus DH RR ekosistēmas cikla vai procesa posmus: izstrādes un uzturēšanas posmu (sk. 1. modeli).

3 Viena no retajām publikācijām ir “Latviešu valoda digitālajā vidē: datorlingvistika: informatīvi izglītojoša semināru cikla materiāli: rakstu krājums” (Baldunčiks, Auziņa 2012).

4 Kaut kādā mērā šis ir arī vēsturisks pētījums par Latvijas DH lauku. Rakstā šim aspektam pieskaramies tikai garārojot, bet pētījuma laikā ievāktie dati (īpaši ekspertu intervijas) var kalpot padziļinātai šī temata izpētei.

<i>intereses, vajadzības, resursi</i> <i>(pieprasījums un pasūtījums)</i> intereses var atšķirties un pat konfliktēt kapitāli, zināšanas, spējas	<i>aģenti (iespējamie)</i> <i>lomas var pārklāties</i>	<i>cikla posmi</i>
	pasūtītājs finansētājs izstrādātājs (HUM + IT) lietotājs (potenciālais)	ideja, plāns: izstrādāšana
	pasūtītājs finansētājs izstrādātājs (HUM + IT) lietotāji (arī nelietotāji, bijušie lietotāji) uzturētāji	gatavs resurss: resursa uzturēšana

1. modelis. DH RR ekosistēmas uzbūve.

Skatoties plašākā tvērumā uz viena DH RR ekosistēmu, tā ir citu, lielāku, specifiskāku vai vispārīgāku, ekosistēmu daļa. Piemēram, konkrēts DH rīks vai resurss ir iekļauts plašākā institūcijas digitālās infrastruktūras sistēmā, kura savukārt ir institūta zinātniskās infrastruktūras daļa, bet tā – lielākas institūcijas vai visas valsts zinātnes infrastruktūras un politikas daļa. Zinātnes un politikas lauki vismaz daļēji pārklājas un ir savstarpēji ietekmēti no citiem laukiem, piemēram, ekonomikas. Tas norāda uz to, ka dažāda mēroga DH ekosistēmas ir saistītas un atkarīgas arī no sociālajām normām, vērtībām, konvencijām, tradīcijām, no institūcijas vai nācijas ekonomiskās labklājības, no institūcijas iespējām un pētnieku rīcības spējas u. c. faktoriem (Collins et al. 2015: 45). Pat skatoties tikai uz atsevišķu DH RR, tas pats par sevi jau ir salīdzinoši sazarota un daudzslāņaina sistēma (Zorich 2008: 39).

Būtiskākie aspekti DH RR veidošanā

Viens no raksta pētnieciskajiem jautājumiem veltīts svarīgāko DH RR izstrādāšanas un uzturēšanas procesa aspektu noskaidrošanai. Tāpēc šajā apakšnodaļā apkopoti teorētiskajā un analītiskajā DH literatūrā sastopamie priekšstatī un pieņēmumi par to, kas tiek uzskatīts par svarīgāko šajā procesā. Tālāk, empīriskās analīzes daļā, tie salīdzināti ar ievāktajiem datiem par Latvijas DH RR izstrādātāju un uzturētāju vērtējumiem.

Autoru iepriekš pieņemtā pieeja paredz to, ka DH rīki un resursi nav gatavas un autonomas IT sistēmas, bet gan ekosistēma, kas ietver arī to pasūtītājus, izveidotājus, uzturētājus un lietotājus, viņu vajadzības, zināšanas un spējas, neaizmirstot par plašākiem sociāli politiskajiem kontekstiem, kultūras konvencijām, ekonomisko un politisko pieprasījumu.

Kā norāda dažādie DH RR tipi un nosaukumi (piemēram, vārdnīca, tekstu korpusi, tekstu analīzes un vizualizācijas rīki, datubāze, digitāls arhīvs, spēle u. c.), tie var būtiski atšķirties vairākos parametros, un varbūt vienīgais kopīgais ir tikai tas, ka tie ir digitāli un ka to saturs vai funkcijas tieši vai netieši ietilpst humanitāro zinātņu un zināšanu sfērā. Tas liek uzdot pirmo jautājumu: vai maz iespējams kāds teorētisks ideālā DH RR veidošanas un uzturēšanas procesa modelis? Visdrīzāk – nav. Tāpēc arī DH akadēmiskajā literatūrā ir sastopams

daudz dažādas specializācijas un mērogu teorētisko modeļu, bet maz mēģinājumu veidot ko universālu. Tādēļ turpmāk akadēmiskajā literatūrā meklētas atbildes uz mazāk globāliem jautājumiem: kas ir tipiskie un biežāk minētie DH RR izstrādes un uzturēšanas labās prakses piemēri un kas raksturo labu DH RR?

Izstrādātāji: intereses, sadarbība

Atsaucoties uz autoru pieņemto sistēmisko jeb ekoloģisko pieeju DH RR izpētei, svarīgākie aģenti to izstrādē ir to izstrādātāji un uzturētāji – indivīdi, grupas un institūcijas. Tātad viņu intereses, vajadzības, resursi un no tiem izrietošā prakse ir noteicošie DH RR funkcionēšanā, tāpēc izpētes vērti.

Pirmkārt, būtu jāmēģina saprast, kādas ir tās intereses un motivācija, kas indivīdiem un institūcijām liek pasūtīt vai pašiem veidot DH RR. Viena no dominējošajām psiholoģiskajām motivācijas teorijām ir Edvarda Dīsi (*Edward L. Deci*) un Ričarda Raiena (*Richard M. Ryan*) 70. gados izstrādātā pašnoteikšanās teorija (*Self-Determination Theory*). Autori uzskata, ka jebkurai brīvprātīgai cilvēka rīcībai ir divas pamata motivācijas: iekšējā un ārējā. Iekšējā motivācija nozīmē, ka indivīdiem liek rīkoties ieinteresētība gūt iekšēju gandarījumu par pašu procesu, bet ārējā motivācija liek cilvēkiem rīkoties, lai saņemtu kaut ko citu, nevis tikai gandarījumu par darbības veikšanu, piemēram, atlīdzību vai izvairīšanos no soda (Ryan, Deci 2017). Šīs divas motivāciju kategorijas vēlāk tika precizētas, iekļaujot trešo starpkategoriju – internalizēta ārējā motivācija, t. i., indivīdi rīkojas, jo ir apzinājušies savas rīcības ieguvumus (turpat). Šo psiholoģiskās motivācijas teoriju papildinām ar sociologa Pjēra Burdjē kapitālu teoriju, tādējādi stiprinot motivācijas modeļa sociālā aspekta izpratni (Daugavietis 2021). Blakus ekonomiskajam kapitālam Burdjē liek sociālo, kultūras un simbolisko kapitālu, kas rīcības motivācijas kontekstā interpretējami kā resursi (katrs šī kapitāla veids un tā īpatsvars indivīda “kapitālu krājumā”), kas vienlaikus ir gan indivīda rīcības spējinautāji, gan arī iekārojami labumi (Bourdieu 1986). Tātad mūsu izmantotā motivācijas modeļa vienā “galā” ir psiholoģiskā iekšējā motivācija (darbošanās “sava prieka pēc”), bet otrā – utilitārā (piemēram, ekonomiskā peļņa), savukārt “vidū” ir no Burdjē teorijas pārņemtās sociālo motivāciju kategorijas (Daugavietis 2021).

Kā vairumā teorētisko modeļu, arī šī klasifikācija ir tikai ideālie tipi jeb teorētiskas kategorijas, kas tīrā veidā varbūt nemaz nepastāv, turklāt to empīriskās robežas nav strikti nodalāmas. Tāpat būtiski piebilst, ka indivīda rīcību vienlaikus var motivēt dažādi un atšķirīgi apsvērumi.

Otrkārt, DH RR praktiski vienmēr ir sadarbības rezultāts. Pat tajos gadījumos, kuros rīku vai resursu izstrādājis viens indivīds. Vismaz netiešā veidā viņš ir komunicējis ar citiem indivīdiem, ar vairākām kopienām, sākot ar kodēšanas valodas autoriem un adaptētājiem, beidzot ar visdažādākajām konvencijām, gaumēm un kultūras simboliem, datu avotiem utt. Tomēr parasti darbs notiek grupā vai sadarbojoties vairākām grupām, institūcijām, tāpēc komunikācija un sadarbība ir neizbēgams DH RR veidošanas un uzturēšanas procesa elements. DH laukā ir tipiska sadarbība starp vismaz divu zinātnisko nozaru pārstāvjiem, t. i., humanitāro zinātņu un datorzinātņu, kas savukārt norāda uz īpašu akadēmiskās

komunikācijas veidu – starpdisciplināritāti⁵. Atšķirībā no daudzām citām zinātnes nozarēm te tā nav uzspiesta “no augšas” (kā labas zinātniskās darbības indikators), bet gan “dabiska”. Tāpēc DH literatūrā šis ir biežs temats, un kā galvenā te tiek izcelta savstarpējās izpratnes jeb tulkšanas problēma (Poole 2017a: 1777). Dažādās nozarēs un disciplīnās izglītoti un nodarbināti zinātnieki un speciālisti nāk ar savu valodu, terminoloģiju, tradīcijām un izpratni, un daudzos gadījumos tās nesakrīt, pat ja tiek lietoti vieni un tie paši vai līdzīgi termini. Papildus tam jāuzsver, ka DH RR veidošanā un uzturēšanā nereti ir iesaistītas arī privātās, publiskās un nevalstiskās institūcijas, kas nāk ar citu komunikācijas un darba kultūras pieredzi, standartiem un tradīcijām. Tas viss prasa sistemātisku procesu un zināšanu organizāciju (Golub et al. 2021).

Lietojamība un tās testēšana

Viens no bieži izmantotiem jēdzieniem, runājot par rīka novērtēšanu, ir tā lietojamība (*usability*). Nav grūti iedomāties, ka tas ir tikai novērtēšanas procesa operacionalizācijas sākums, jo ir jāsāk definēt, kas ar šo jēdzienu tiek saprasts. Tas var nozīmēt daudzus un dažādus DH RR aspektus. Tāpēc jāpieņem, ka jēdziens “lietojams rīks” faktiski ir sinonīms jēdzienam “labs rīks”.

Šis jēdziens nereti parādās DH RR izmēģināšanas/pārbaudes kontekstā, sakot, ka tiek “testēta lietojamība”. Lai arī daži autori apgalvo, ka lietojamības problēmas parasti nav rīkspecifiskas (*tool-specific*), bet gan vispārējas (*generic*) un kopīgas lielai daļai DH RR (Thoden et al. 2017: 4), konkrētu potenciālo lietojamības problēmu noteikšanas un mērīšanas gadījumā vēl jo vairāk nepieciešama šī jēdziena saskaidrošana daudzos un konkrētos parametros. Tāpēc tik plašs jēdziens kā “lietojamība” nav uzskatāms par lietderīgu analītisko rīku. Tomēr tas dažreiz implicīti, bet citreiz eksplīcīti norāda uz nepieciešamību rēķināties ar gala lietotāja (*the end user*) jeb mērķauditorijas vajadzībām un spējām (Steiner et al. 2014).

Mērķauditorija: komunikācija, izpēte

Mērķauditorija jeb lietotāji allaž ir viens no lielajiem DH tematiem, ieskaitot DH RR izvērtēšanu. Pats jēdziens “mērķauditorija” norāda uz to, ka tā ir DH RR izstrādes mērķis un pamatojums, viens no centrālajiem DH ekosistēmas elementiem. Tāpēc jau pirms konkrēto DH RR izstrādes būtu vēlams izprast, kas ir tā iedomātā mērķauditorija, ko viņi vēlas un ko spēj. Tam kalpo komunikācija ar lietotājiem un to izpēte. Tās ir prakses, kuru uzdevumi un rezultāti daļēji pārklājas, un jo mazāk resursu tiem tiek veltīti, jo mazāk izpēte (lietotāju problēmu vai nelietotāju motivācijas noskaidrošana) atšķiras no komunikācijas (auditorijas piesaistīšanas, pieradināšanas, apmācības).

5 Jāpiezīmē, ka tas ir ideālais modelis, bet praksē nereti DH RR izveides un uzturēšanas tehnoloģiskā funkcija ir ārpakalpojums, ko veic nevis datorzinātnieki, bet gan IT privātuzņēmumi vai atsevišķi noliegti IT speciālisti.

Tā kā atšķiras ne tikai dažādi DH rīki un resursi, bet vēl jo vairāk – dažādas lietotāju (un nelietotāju) grupas, tad arī tie ir neiespējami radīt universālu teoriju vai koherentus lietotāju vajadzību, kompetenču, rīcību modeļus. Tā vietā autori parasti norāda uz katrā teorētiskās vai empīriskās analīzes gadījumā (piemēram, dažāda tipa DH RR izpētē), viņuprāt, svarīgākajiem aspektiem, kas būtu jāņem vērā, veidojot un uzturot DH RR. Visbiežāk tiek minēta saskarnes (*interface*) nozīme (Steiner et al. 2014; Thoden et al. 2017).

Vienkāršība vai komplicētība?

Viena no netriviālām lietotāju un lietojamības problēmām ir DH RR “vienkāršība”. Bieži tiek lietots arī jēdziens “intuitīvs”, lai raksturotu saskarnes vai paša rīka vienkāršību un saprotamību. No vienas puses vienkāršība, kas jāattiecinā uz vairākiem parametriem, sākot ar dizainu, beidzot ar dokumentāciju, ir viena no pamatprasībām, veidojot rīkus un resursus (Gibbs, Owens 2012), jo pat daļa pētnieku nereti “jūtas apmulsuši par daudzām vietnēm, kas ir paredzētas jomas ekspertiem” (Warwick et al. 2007: 299). Tai pašā laikā ir zinošākas vai pieredzējušākas lietotāju grupas, kas spēj šos rīkus izmantot salīdzinoši pilnvērtīgi un kuriem tie arī bijuši radīti. Tas norāda arī uz DH RR saskari ar profesionālajām, politiskajām un finansiālajām vajadzībām un iespējām. Vai, izstrādājot rīku, ir jādomā par maksimāli iekļaujošu produktu, kas nozīmē rēķināšanos ar vairākām un ļoti atšķirīgām mērķauditorijām? Vai ir pieļaujama šauri specializēta rīka uzbūve, un kuram no ceļiem ir pietiekams finansiālais segums?

Tāpat DH RR izstrādē un uzturēšanā jāņem vērā arī sociālie un kultūras aspekti, jo rīks ir arī kultūras artefakts, vēl jo vairāk – digitālo humanitāro zinātņu rīks un resurss. Tā dizains pats par sevi ir mākslas darbs, kas ietver dažādus kultūras simbolus un atsaucis, kamēr tā saturā var glabāties vēl daudz vairāk neviennozīmīgāk vērtētu, strīdīgu vai pat vienai vai otrai ideoloģijai, režīmam vai ticībai naidīgu kultūras objektu. Tāpat DH RR izveides mērķis un tā izmantošanas veidi norāda uz noteiktām kultūras un politiskajām interesēm, par kurām sabiedrībā reti kad ir sasniegta pilnīga vienprātība. Piemēram, ja DH RR ir instruments kādas minoritātes interešu un vajadzību īstenošanai, tas visdrīzāk raisīs dažādus jautājumus no majoritātes pārstāvjiem. Jāatceras, ka paši DH pētnieki ir izteikta minoritāte, kas ir atkarīga no valsts zinātnes politikas prioritātēm un pieejamā finansējuma.

DH RR ilgtspēja, finansējums, atvērtā zinātne

DH RR ilgtspēja, finansējums un atvērtā zinātne ir cieši saistīti aspekti. Jēdziens “ilgtspēja” nācis no 80. gadu globālās vides politikas lauka, kur tas sākotnēji nozīmēja tādu dabas resursu izmantošanu un saimniecisko darbību vispār, kas planētu saglabā arī nākamajām paaudzēm, bet vēlāk tika paplašināts, iekļaujot arī sociālo un kultūras dimensiju (ar īpašu uzsvāru uz marginālo un mazo grupu interešu respektēšanu) (Daugavietis 2015: 10). DH RR kontekstā ilgtspēja tiek saprasta tehnoloģiskāk. Konkrētajam rīkam vai resursam vajadzētu spēt būt ne tikai funkcionālam un lietotam ilgi: “(..) [I]lgtspējas gala mērķis ir nevis stagnācija, bet gan

transformācija un atkārtota izmantošana.” (Edmond, Morselli 2020: 1024) Varam prognozēt, ka arvien biežāk no DH RR tiks prasīta arī atbilstība citām ilgtspējīgās attīstības pamatprasībām, ieskaitot energoefektivitāti un minoritāšu kultūras vajadzību ievērošanu.

Aprakstot ilgtspējīgu DH RR, bieži tiek minēti tie paši objekti un principi, uz kuriem balstās atvērtā zinātne, it īpaši – atvērtie pētniecības dati, datu repozitoriji un atvērtās infrastruktūras. Tie ir atvērtā koda programmatūra, pieejamība, sadarbīpēja, sabiedriskā zinātne un digitālā talkošana (*crowdsourcing*), brīvpieeja (*open access*) publikācijām, datiem un DH RR, sekundārā pētniecība, datu un metadatu standartizācija, informācijas un zināšanu sistematiska organizēšana un daudzi citi (sk. Liu 2004; Callison-Burch, Dredze 2010; Berry 2012; Blumberg 2013; Boyd, Larson 2014; Collins et al. 2015; Sabharwal 2015; Hai-Jew 2017; Hedges, Dunn 2017; Poole 2017b; Woodcock et al. 2017; Levenberg et al. 2018; Golub et al. 2021 u. c.). Tāpēc viens no vienkāršākajiem un drošākajiem veidiem, kā vērtēt DH RR ilgtspēju, ir izmantojot atvērto pētniecības datu novērtēšanas FAIR modeli⁶.

Viegli saprotams, ka DH RR ilgtspējas nodrošināšanai ir vajadzīga stabila un nepārtraukta finanšu plūsma. Ar retiem izņēmumiem, kas parasti ir bagātas valstis, kur arī salīdzinoši liels IKP īpatsvars tiek veltīts pētniecībai un attīstībai (angliski – R&D), realitāte ir cita: “(..) [DH RR] bieži vien ir atkarīgi no īstermiņa, uz projektiem balstītiem grantiem [...], cīnās ar līdzīgām problēmām: ilgtspēja, projektu vadība, institucionālā pozīcija un sava darba zinātniskais novērtējums.” (Poole 2017a: 1772) Tā ir sistematiska problēma, ar kuru saskaras to valstu pētnieki, kuru veikums un piedāvājums tiek atbalstīts konkursu kārtībā ar īstermiņa zinātniskajiem grantiem, un no viņiem drīzāk tiek prasītas inovācijas, nevis jau gatavu produktu uzturēšana (Maxwell et al. 2019).

Rekomendācijas DH RR ilgtspējas nodrošināšanai ir tipiskas un labi zināmas: “Mēs iesakām pāriet uz finansēšanas modeļiem, kas nav balstīti uz projektiem, bet gan uz digitālo infrastruktūru sertificēšanu, atbilstošu finansējuma novērtēšanu, Eiropas mēroga politiku un stratēģijas, kā arī digitalizācijas un metadatu standartu un vārdnīcu paraugpraksi pieņemšanu. Pētniekiem paraugu piemēri jāmeklē tādās lielās organizācijās kā *Research Data Alliance* un *DARIAH*.” (Collins et al. 2015: 45)

Kā norāda Alekss Pūls (*Alex Poole*), DH RR ilgtspēja ir diezgan jauna un augstākās pakāpes problēma (Poole 2017b). Domājams, ka tas īpaši labi raksturo ne tikai šī brīža Latvijas DH lauka situāciju, bet Latvijas zinātni kopumā, kurā finansējums ir salīdzinoši mazs un ne-prognozējams, strikti atvērtās zinātnes principi vēl nav ieviesti pat politikas dokumentos, un humanitāro un mākslas zinātņu pētnieku zināšanas un prakses Latvijas zinātņu nozaru vidū raksturojamas kā visneapmierinošākās (Skadiņa 2019; Bite u. c. 2020).

6 Akronīms: *Findable* – atrodamais; *Accessible* – pieejamais; *Interoperable* – savietojamais; *Reusable* – atkārtoti lietojams (Wilkinson et al. 2016; Bite u. c. 2020).

DH RR ekosistēmas modelis

Mūsu izveidotā DH RR ekosistēmas modeļa četri pamata elementi ir: (1) pats RR kā tehnoloģiska, materiāla IT sistēma (digitāla infrastruktūra un tās saturs); (2) aģenti (indivīdi, grupas, organizācijas, institūcijas), kas pasūta, finansē, izstrādā, uztur un/vai izmanto; (3) aģentu intereses; (4) aģentu kapitāli. Tie ir mainīgi lielumi, kuru kvalitāte mainās laikā un atšķiras telpā, tāpēc atsevišķa DH RR dimensija ir procesi. DH RR ekosistēmā izdalām divus veidošanas cikla posmus – izstrādi un uzturēšanu (sk. 1. modeli “DH RR ekosistēmas uzbūve”), turklāt jāņem vērā vēl vairāki citi aspekti un dimensijas.

Pirmkārt, jāiedomājas, ka ekosistēmā ir vairāki mērogi (mikro, mezo, makro), kur katrā var darboties atšķirīgas likumsakarības un aģentu rīcības loģika. Otrkārt, aģenti (gan indivīdi, gan institūcijas) vienlaikus var ieņemt dažādas lomas un īstenot atšķirīgas (pat konfliktējošas) intereses. Treškārt, ekosistēma ir nemitīgi mainīga laikā, daudzi procesi (piemēram, DH RR izstāde un uzturēšana) notiek vienlaikus. Ceturtkārt, ekosistēmas elementi un to savstarpējās attiecības nepastāv pašas par sevi un atsevišķi. Piektkārt, kā jau iepriekš norādīts, katrs konkrētais DH RR ir kādas lielākas digitālās infrastruktūras un ekosistēmas daļa.

Rezumējot teorētiskās un lietišķās izpētes literatūras apskatu, kas veltīta DH RR veidošanai, uzturēšanai un novērtēšanai, var secināt, ka nav vienotu un globālu teorētisko modeļu. DH rīki un resursi mēdz būt dažādi, to mērķauditorijas jeb lietotāju grupas veido vēl lielāku dažādību, tāpēc te nav iespējams modelis, kas derīgs visam. Tā vietā pētnieki un praktiķi mēģina formulēt dažādus vairāk vai mazāk vispārīgus principus un aspektus, kuriem būtu jāpievērš īpaša uzmanība divos galvenajos DH RR cikla posmos: izstrādē un uzturēšanā.

DH RR ekosistēmas ciklu dalām divos iedomātos procesos (izstrādē un uzturēšanā), bet procesu analīzē koncentrējamies uz trim plašākajiem no sešiem Pūla piedāvātajiem DH RR cikla elementiem: sadarbība un komunikācija (t. sk. starpdisciplināritāte); plānošana un projektu vadība; informēšana un zināšanu pārnese (Poole 2017a). Savukārt pārējie trīs elementi, t. i., datu izmantošanas problēmas, resursi, tehnoloģijas, ir specifiskāki, un tos šajā pētījumā padziļināti nepētām, kaut gan tie, tāpat kā citi aspekti, ir nodalāmi vien analītiski un tāpēc nevar tikt pavisam izolēti no teorētiskās un empīriskās analīzes. Tā kā šis ir pirmais šāda veida pētījums par Latvijas DH lauku, empīriskajai analīzei izvēlēti tikai daži no svarīgākajiem “DH RR ražošanas procesa” aspektiem:

- izstrādātāji – motivācija un intereses, ideju un plānu saskaņošana, institucionālā un individuālā sadarbība;
- lietotāji, mērķauditorija – testēšana, komunikācija, izpēte;
- DH RR uzturēšana un pilnveide, tostarp ilgtspējas, finansējuma, atvērtās zinātnes problemātika;
- izstrādātāju priekšstati par labu DH rīku un resursu.

Datu ieguves metodes

Raksta mērķis ir Latvijas DH RR izstrādāšanas un uzturēšanas prakšu izpēte, bet uzdevums – iegūt un analizēt datus par dažādiem izstrādātāju un uzturētāju iesaistes un līdzdalības aspektiem, sākot ar motivāciju iesaistīties resursu veidošanā, beidzot ar pašu izstrādātāju vērtējumiem par šo procesu un galarezultātu. Pētījumā padziļinātai analīzei izvēlēti 10 pētnieciskajā projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” iesaistīto partnerinstitūciju DH rīki un resursi, ko veido sešas no septiņām projektā iesaistītajām institūcijām (sk. 1. tabulu). Tātad pētījums ir par lielāko valsts finansēto pētniecisko DH rīku un resursu veidotāju institūciju ekosistēmu. Lai arī apskatītie rīki un resursi neveido visu DH RR ekosistēmu attiecīgajās institūcijās, var pieņemt, ka prakses to izstrādē un uzturēšanā vienas institūcijas robežās neatšķiras, un tas ļauj izpētes secinājumus ekstrapolēt uz visu ekosistēmu.

<i>Resurs</i>	<i>Institūcija</i>
Periodika (http://periodika.lv/)	Latvijas Nacionālā bibliotēka
Zudusi Latvija (https://zudusilatvija.lv/)	Latvijas Nacionālā bibliotēka
Vietvārdu talka (https://vietvardi.lv/talka/)	Latvijas Universitātes Latviešu valodas institūts
Latviešu folkloras krātuves digitālais arhīvs (http://garamantas.lv/)	Latvijas Universitātes Literatūras, folkloras un mākslas institūts
Literatura.lv (https://literatura.lv/)	Latvijas Universitātes Literatūras, folkloras un mākslas institūts
LIV-EE-LV [Lībiešu] vārdnīca (http://www.livones.net/lili/lv/vardnica)	Latvijas Universitātes Lībiešu institūts
Latviešu valodas tekstu un runas korpusi (http://www.korpuss.lv/)	Latvijas Universitāte, Matemātikas un informātikas institūts
Tēzauris.lv (https://tezauris.lv/)	Latvijas Universitāte, Matemātikas un informātikas institūts
Mūsdienu latgāliešu tekstu korpus (http://www.korpuss.lv/id/MuLa2022)	Latvijas Universitāte, Matemātikas un informātikas institūts; Rēzeknes Tehnoloģiju akadēmija
Īsapazeisim (https://iepapziesimies.rta.lv/)	Rēzeknes Tehnoloģiju akadēmija

1.tabula. Padziļinātai analīzei izvēlētie VPP “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” partnerinstitūciju DH rīki un resursi.

Datu avots ir sešu institūciju darbinieku interneta aptauja un daļēji strukturētās ekspertu intervijas. Respondentiem un intervētajiem jautāts gan par viņu personīgo pieredzi, veidojot konkrētus savu pārstāvēto institūciju DH rīkus un resursus, gan arī lūgti vērtējumi par kopīgo situāciju šajā laukā Latvijā. Galvenās aptaujas un interviju tēmas ir vienas un tās pašas, lai par tām iegūtu gan kvantitatīvus, gan kvalitatīvus datus.

Aptauja

Aptaujas instrumentāriju pamatā veido slēgtie jautājumi, tomēr tajā iekļauti arī vairāki atvērtie jautājumi, kas ļauj veikt refleksiīvāku analīzi. Kā datu ieguves tehnika izmantota slēgta⁷ elektroniskā tiešsaistes aptauja, izmantojot atvērtā koda aptauju rīku *LimeSurvey* (Schmitz 2018), kas iepriekš instalēts LU LFMI serverī (aptauja.garamantas.lv).

Respondentu izlase veidota, konsultējoties ar sešu institūciju pārstāvjiem (LNB, LU LaVI, LU LFMI, LU LI, LU MII, RTA) un lūdzot iesniegt to darbinieku sarakstu, kas strādā vai ir strādājuši pie DH rīku un resursu izstrādes vai uzturēšanas. Tādā veidā izveidota 97 indivīdu izlase, kura ievietota aptaujas rīka respondentu datubāzē (vārds, uzvārds, e-pasta adrese). Izmantojot *LimeSurvey* rīku, izlasē iekļautajiem respondentiem izsūtīti individualizēti uzaicinājumi piedalīties aptaujā, kā arī līdz četriem atgādinājumiem, ja konkrētais respondents nav iesācis vai pabeidzis aptauju (respondentu anonimizācija sistēmā noris automātiski). Lauka darbs: 2021. gada jūnijs–septembris.

Aptaujas gaitā daži respondenti atteicās piedalīties, jo vai nu neuzskatīja sevi par pietiekami kompetentiem, vai arī uzskatīja, ka nav pietiekami strādājuši ar DH rīkiem vai resursiem. Kopējais sasniegtais respondentu skaits ir 92, no kuriem anketu pabeiguši 69 respondenti (sk. 2. tabulu). Analīzē izmantoti arī nepabeigtu anketu dati. Kopumā var uzskatīt, ka sasniegtā izlase labi reprezentē ģenerālo kopu – iepriekš minēto sešu institūciju darbiniekus, kas strādā ar DH rīkiem un resursiem.

<i>institūcija</i>	<i>izlase</i>	<i>atvērta anketa</i>	<i>pabeigta anketa</i>
LNB	9	8	7
LU LFMI	41	41	29
LU LI	9	7	4
LU LaVI	9	9	5
LU MII	21	20	19
RTA	8	7	5
<i>kopā:</i>	97	92	69

2.tabula. Aptaujas izlase.

Aptaujas anketas jautājumus var dalīt četrās lielās grupās: sociāli demogrāfiskie (dzimums, vecums, izglītība, darbošanās ilgums DH laukā), vispārīgs DH resursu vērtējums (piemēram, labākie Latvijas DH resursi, kas ir paredzēti pētniekiem un plašākai sabiedrībai; laba resursa pazīmes), jautājumi par respondenta institūcijas praksēm, veidojot DH resursus (piemēram, dokumentēšana; sadarbība ar citām institūcijām), jautājumi par vienu konkrētu DH resursu,

⁷ Katram izlasē iekļautajam respondentam tika izveidota unikāla piekļuves adrese (URL).

pie kura respondents ir darbojies tā izstrādes vai uzturēšanas gaitā (piemēram, par funkcijām, kuras apsvērts ieviest resursa izstrādes laikā, par mērķauditoriju).

Dati pārsvarā apstrādāti, izmantojot viendimensionālu deskriptīvo analīzi. Ne tikai tāpēc, ka izlase ir par mazu, lai salīdzinātu dažādu institūciju respondentu atbildes uz vieniem un tiem pašiem jautājumiem, bet arī tāpēc, ka pētījuma uzdevumi ir samērā plaši, fokusēti uz attiecīgo institūciju veidotās DH rīku un resursu kopainas izpēti. Aptaujas anketa, aptaujas datu fails un metodoloģiskais apraksts deponēts brīvpieejas datu arhīvā⁸.

Ekspertu intervijas

Otra empīrisko datu ievākšanas metode ir DH rīku un resursu izstrādātāju daļēji strukturētās ekspertu intervijas. To uzdevums ir iegūt kvalitatīvus datus, kas papildina un dod skaidrojumu aptaujā iegūtajiem kvantitatīvajiem datiem. Šīm intervijām pamatā atlasīti pētnieki un IT speciālisti, kas DH laukā strādājuši ilgāku laiku un atradušies vadošos amatos vai pildījuši līdzīgas, bet neinstitucionalizētas lomas. Gandrīz visi intervētie ir DH VPP partnerorganizāciju pārstāvji, kas darbojas vai darbojušies ar konkrētu DH rīku un resursu izveidi. Izņemot vienu interviju klātienē, visas pārējās veiktas tiešsaistē (izmantojot *Zoom* programmatūru, ierakstot un saglabājot audiofailu). Pēc dažām intervijām e-pasta vēstulēs iegūti saturiski papildinājumi un papildu dokumenti vai norādes uz tiem. Kopumā veiktas 13 daļēji strukturētas intervijas (intervēto ekspertu saraksts pievienots pielikumā).

Interviju jautājumu tēmas ir līdzīgas aptaujas saturiskajai struktūrai, tie ir jautājumi par intervējamo pieredzi DH laukā, vispārīgi par DH, par labu DH rīku un resursu, par resursu izstrādes un uzturēšanas procesu, par sadarbību ar partneriem, par mērķauditoriju un lietotāju nozīmi. Atbilstoši ekspertu interviju praksei un iepriekš vienojoties par to ar intervētajiem, pie analīzē izmantotajiem citātiem norādīti autoru vārdi. Tāpat kā aptaujas dati, interviju metodoloģija un to audioierakstu faili deponēti brīvpieejas datu arhīvā iespējamās sekundāras analīzes vajadzībām⁹.

DH rīku un resursu veidošana Latvijā

Šajā nodaļā tiek analizēti pētījuma gaitā ievāktie empīriskie dati. To tvērums ir DH ekosistēma, ko veido nozīmīgākās Latvijas DH publiski finansētās akadēmiski pētnieciskās institūcijas. Tā ir nozīmīga daļa, bet ne viss Latvijas DH lauks. Ārpus šī pētījuma robežām paliek privātā un t. s. nevalstiskā sektora DH aktivitātes rīku un resursu veidošanā, kā arī daudzas valsts un pašvaldību institūcijas, kuru primāro darbības mērķu vidū nav akadēmiska pētniecība (piemēram, muzeji vai citas nemateriālā kultūras mantojuma un atmiņas institūcijas).

8 "Latvijas digitālo humanitāro resursu izstrādātāju aptauja (metodoloģija un datu fails)" (2022). Pieejams: <https://doi.org/10.5281/zenodo.6542120>

9 Sk. "Latvijas humanitāro zinātņu digitālie resursi". Pieejams: https://zenodo.org/communities/dh_lv

Turpmākajās sadaļās ar iegūtajiem datiem ilustrēta Latvijas DH nozīmīgāko publiski finansēto akadēmiski pētniecisko institūciju un to DH RR ekosistēma, dalībnieku lomas un saiknes, komunikācija un sadarbība, izsekojot DH resursa izveides ceļu no iniciatīvas rašanās brīža līdz izveidotā resursa uzturēšanai un papildināšanai ilgtermiņā. Atsevišķas apakšsadaļas veidotas ap šādiem tematiem: izstrāde; sadarbība; mērķauditorija; uzturēšana; izstrādātāju priekšstati par labu resursu. Būtiski atzīmēt, ka analīze veidota, salīdzinot visu iesaistīto un ieinteresēto pušu (humanitāro zinātņu pārstāvju, datorzinātnes un matemātikas pārstāvju) skatījumu.

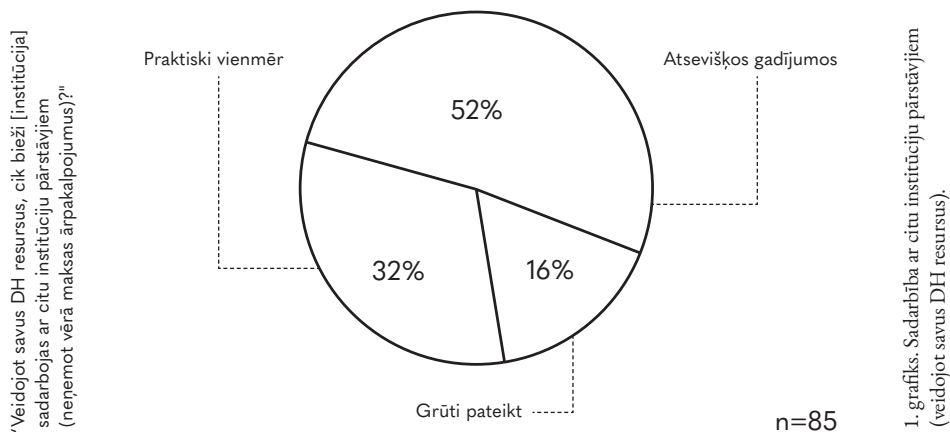
Ekosistēmas aģenti un to attiecības

Ja ekosistēma ir kompleksa sistēma ar noteiktām robežām, kuru veido elementi mijiedarbē, tad no sākuma svarīgi iezīmēt, kas ir pētāmās ekosistēmas aģenti, to attiecības un lomas. Piedāvājam analīzi, kuras centrā novietoti izvēlētie Latvijas DH rīki un resursi un ar to izveidi, uzturēšanu un lietošanu saistīto lomu īstenotāji (sk. 1. tabulu). Tādējādi iegūstam ierobežotu skaitu resursu, kas atrodas atšķirīgās “gatavības” un autonomijas pakāpēs attiecībā pret citiem: daži resursi jau šobrīd integrē citu veidotus resursus (plašākā vai specifiskākā mērogā), citi resursi raksta gatavošanas laikā (2021. gada rudenī) gatavoja kāda specifiska resursa integrācijai papildus savai ierastajai funkcionēšanai un DH laukā jau pazīstamam “zīmolam”, vēl citi ir tapšanas un testēšanas posmā kā pagaidām autonomi resursi. Var norādīt, ka visai izteiktais uzsvars uz integrāciju šķietami izriet tikai no konkrēta projekta rakstura, taču veiktās ekspertu intervijas vairākkārt apstiprina, ka tieši fragmentācija, sadrumstalotība ir raksturīga pašreizējās Latvijas DH ekosistēmas iezīme – tāpēc arī aktuāli centieni šo stāvokli pārvarēt.

Mums katrs arhīvs, katra pašvaldība vai bibliotēka mēģina veidot savu mājaslapu (...) šī resursu sadrumstalotība joprojām ir Latvijas problēma. Mēs esam ļoti maza valsts, kurā būtībā būtu jābūt vienam šādam resursam, kur katram ir sava lapa (...). Šai videi ir jābūt vienotai. (Brūvelis 2021)

Nav iespējams raksturot konkrēto DH ekosistēmu, neskatot tās būtiskāko organizatorisko izkārtojumu – projektu. Projekts, kā gandrīz vienmēr neprognozējams finansējuma avots (jo notiek konkursa veidā), var sniegt vai nesniegt iespējas īstenot darbības konkrētā laika nogrieznī, ar konkrētu resursu klātbūtni un apsoliņumu par noteiktu mērķu sasniegšanu. No vienas puses, eksperti atzīmē, ka institūciju iekšienē šāda projektu sistēma ļauj kritiski izvērtēt vajadzības un izvēlēties prioritāras.

No otras puses, projektu finansētājiem kā ekosistēmas dalībniekiem ir savi noteikumi (piemēram, ar projekta nosacījumiem sekmēt starpdisciplināritāti vai kādus plašākus sociālās politikas mērķus), kas ietekmē citu DH ekosistēmas dalībnieku izvēles. Projektus var izvēlēties īstenot kopā ar tiem, ar kuriem jau ir sadarbības pieredze un tādējādi mazināt iespēju nesasniegt finansētājam solītos rezultātus.



Atgriežoties pie konkrētās DH ekosistēmas dalībniekiem, tālāk raksturotas būtiskākās aģentu lomas: pasūtītājs, izstrādātājs, pētnieks, lietotājs. Šis nošķirums gan ir visai nosacīts tādā ziņā, ka tās ne vienmēr ir pilnīgi atsevišķas grupas, jo, kā atklājis īstenotais pētījums, minēto zinātņu nozaru pārstāvji var darboties dažādās, bieži saplūstošās lomās: piemēram, pētnieks var būt gan pasūtītājs, gan izstrādes posma dalībnieks, gan izstrādātā resursa lietotājs. Vairākums ekspertu piemēru par DH RR izstrādi bija tieši par to, kā veidojas sadarbība ar pētniekiem visās pieminētajās lomās.

Būtiska ekspertu piezīme ir, ka sadarbība parasti ir starp konkrētiem indivīdiem, izvēloties cilvēkus ar nepieciešamajām prasmēm un kvalifikāciju, nevis institūcijām – ja konkrētais pētnieks strādātu citā institūcijā, tieši ar šo citu institūciju tiktu noslēgts sadarbības līgums.

Ja mums nepieciešama sadarbība, mēs nevēršamies pie institūcijas, mēs vēršamies pie konkrētiem cilvēkiem, kas strādā konkrētās institūcijās. Ja šis cilvēks, piemēram, nomainītu institūciju, mēs līdz ar to sadarbotos ar citu institūciju, tāpēc ka mums vajag konkrēto cilvēku. (Reinsone 2021)

Nepieciešams arī atzīmēt, ka ne vienmēr notiek atsevišķa resursa vai rīka izveide. Daudzos gadījumos pētnieks dodas pie IT speciālista (kas var būt arī IT zinātnieks) ar konkrētu vajadzību, ko vēlas atrisināt, lai varētu iegūt pētnieciskos rezultātus konkrētai vajadzībai. Kamēr šādu vajadzību kāds nepauž otrreiz, izveidotais risinājums paliek tikai kā potenciāls, atsevišķs resursa fragments.

Tā kā pasūtītāja, izstrādātāja, pētnieka un lietotāja lomas nereti nav strikti nodalītas un nav arī eksplīcīti hierarhiskas (katra ir nepieciešama rezultātam un neatrodas pakļautības attiecībās), attiecības starp šo lomu īstenotājiem var raksturot kā sadarbību, lai arī to visai sistemātiski apgrūtina dažādu lomu pārstāvju atšķirīgā zināšanu bāze un "valoda". Īpaši tajos gadījumos, kad sastopas humanitāro un datorzinātņu (vai IT industrijas) pārstāvji. Kā jau minēts, zinātniskajā literatūrā šo sadarbības procesu mēdz apzīmēt kā "tulkosanu". Vairāk par šo procesu nākamajā sadaļā.

Kā redzams 1. grafikā, veidojot savus DH RR, sadarbība ar citu institūciju pārstāvjiem visbiežāk notiek “atsevišķos gadījumos” (52 %), “praktiski vienmēr” – trešdaļā gadījumu, bet 16 % respondentu nevarēja uz šo jautājumu atbildēt. Citā aptaujas jautājumā vaicāts, vai notikusi komunikācija ar līdzīgu resursu izstrādātājiem no citām institūcijām konkrētā resursa izstrādes procesā. Apstiprinošu atbildi sniegusi gandrīz puse respondentu.

Kopumā var teikt, ka iedomātās DH RR eksosistēmas aģentu attiecības un darbība notiek visai fragmentētā laukā, kur aģentu lomas mēdz arī pārklāties, finansētāji ietekmē izvēles, darbības strukturē projekta forma un noteikumi. Starpinstitucionālās attiecības raksturojamas kā sadarbība, bet tās intensitāte DH RR izstrādē atšķiras, ko varētu skaidrot ar katras konkrētās institūcijas resursiem, iespējām un infrastruktūru. Ja institūcija ir ar tiem visai bagāta, tad tā savus rīkus un resursus var veidot un veido viena pati. Skatot sadarbību mikrolīmenī, t. i., kā attiecības starp indivīdiem, to padara komplicētu vismaz divu “valodu” lietošana (humanitāro zinātņu pārstāvju un IT speciālistu). Tāpat būtiski uzsvērt, ka nereti institucionālā sadarbība sākas kā neformāla komunikācija un izvēle starp dažādu institūciju indivīdiem, kuriem ir svarīgas tieši viņu, nevis visas institūcijas kapitāli un resursi.

Idejas, saskaņošanas un izstrādes posms

Latvijā tieši pētnieki nereti ir bijuši jaunu DH resursu vai rīku veidošanas iniciatori, tāpēc svarīgi saprast, kā viņiem rodas tādas idejas un vajadzības. Gan aptaujā, gan intervijās jautāts par motivāciju un nepieciešamību, kāda bijusi, veidojot konkrētus DH rīkus un resursus.

Uz atvērtu aptaujas jautājumu “Kāda bija resursa N izveides nepieciešamība? Kāpēc tas tika izveidots?” saņemtas atbildes, kurās izteikti dominē pētnieciskais skatījums un pamatojums: vairāk nekā pusē atbilžu (60 %) respondenti norādīja, ka galvenā motivācija ir bijusi datu pieejamības nodrošināšana; pētniecības vajadzības tika minētas 27 % atbilžu, bet nepieciešamība pēc mūsdienīgiem risinājumiem – 14 % (sk. 2. grafiku). Viens no piemēriem iepriekš minētajai motivācijai nodrošināt datu pieejamību pausts tipiskā citātā no aptaujas atvērtā jautājuma atbildēm: *Milzu kvantumi ar dažāda veida datiem, kas alka pēc platformas, kurā tiem varētu piekļūt gan pētnieki, gan sabiedrība un rīkoties pēc vajadzības.*

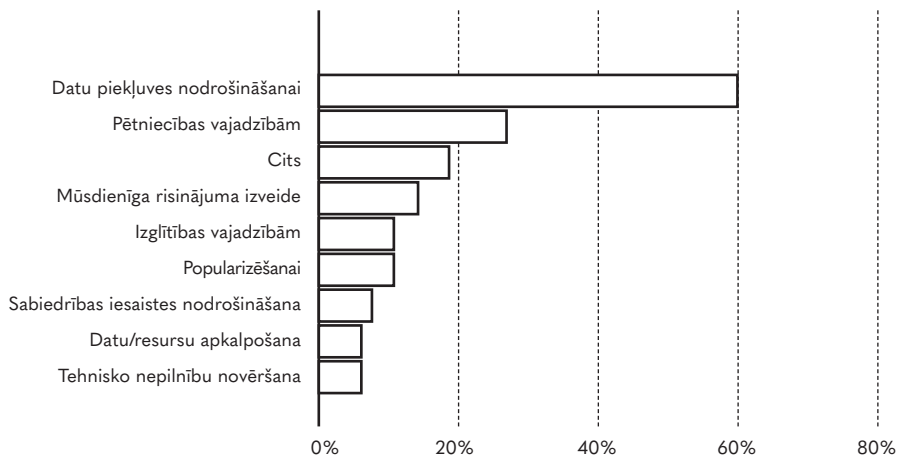
Intervijās atbildes bija personīgākas, kas ļauj izvirzīt apgalvojumu, ka, apstiprinot zinātniskajā literatūrā jau raksturoto modeli, šī motivācija pastāv spektrā no “misijas apziņas” līdz noteiktu kapitālu akumulēšanas vēlmei (simboliskais kapitāls – “atpazīstamība”, “prestīžs”; ekonomiskais kapitāls – sava izstrādātā koda autortiesību aizstāvēšana, vēlme komercializēt datus).

Bet faktiski jau to dara ne tikai sev, bet arī lai kopumā sabiedrībai būtu resursi, pilnvērtīga informācija. Reizē caur resursiem veidojas institūciju atpazīstamība.
(Reinsone 2021)

DH RR izveides pirmais impulss var nākt gan kā institucionāls pasūtījums, kas izriet, piemēram, no valsts zinātnes un izglītības politikas, no konkrētās zinātniskās institūcijas

“Kāda bija resursa izveides nepieciešamība? Kāpēc tas tika izveidots?”

n=63



2. grafiks. Resursa izveides nepieciešamība.

“Kāda bija resursa N izveides nepieciešamība? Kāpēc tas tika izveidots?”. Pārkodēts vairākatbilžu atvērtais jautājums.

stratēģiskajām un praktiskajām vajadzībām, no DH lauka tradīcijām un modes; gan nosacīti no “ierindas” humanitāro zinātņu pārstāvju privātās iniciatīvas – pētniekiem, kuru pētnieciskās vai sabiedriskās intereses rosinājušas ideju par noteikta veida datu ievākšanu, strukturēšanu, digitalizēšanu un pieejamību digitālai analīzei. Nereti šāda privātā iniciatīva ir neapmaksāts darbs, un tas bija īpaši raksturīgs tad, kad Latvijas humanitāro zinātņu akadēmiskajās institūcijās DH RR vēl tikai sāka veidoties, kad valsts pārvaldes digitalizācija vēl nebija politikas plānā un kad nauda šādiem mērķiem tika paredzēta vien specifiskām ar dator-tehnoloģijām saistītām akadēmiskajām institūcijām.

Nākamais procesa DH RR izveides posms ir idejas apspriešana – plašākā kolēģu lokā vai divatā ar citu pētnieku, kura intereses ir vistuvāk ieceres būtībai; tāpat var tikt pieaicināts arī viens vai vairāki IT speciālisti, lai gūtu priekšstatu par idejas realizējamību. Šis ir posms, kurā dalībniekiem vēl nav kopīgas “valodas”, vai arī tā ir ierobežota, īpaši gadījumos, ja humanitāro zinātņu pārstāvjiem iepriekš nav bijusi tāda pieredze, un tāpēc saprasties nav viegli.

Mums¹⁰ grūts bija sākuma posms, tas 2013./2014. gads, mēs pusgadu nevarējām saprasties – bija sajūta, ka mēs katrs runājam citā valodā, īpaši e-pastā (..) tās bija arī terminoloģijas atšķirības: kas kā ko sauc. (Reinsone 2021)

10 Runa ir par LU LFMI Latviešu folkloras krātuves digitālās versijas *garamantas.lv* izveidi.

Daži DH eksperti no humanitāro zinātņu puses norāda, ka viņiem būtu svarīgi arī pašiem būt spējīgiem novērtēt idejas īstenošanas iespējas (sarežģītību, ilgumu), lai nebūtu jāpaļaujas tikai uz IT speciālistu viedokli.

Vismaz tās pamatzināšanas vajadzētu – vai X ir viegli izdarāms vai nav viegli izdarāms. Tad varētu kļauvēt pie izstrādātāju durvīm un teikt – šo var ātri izdarīt, izdari ātri.
(Rapa 2021)

Rodas iespaids, ka reizēm humanitāro zinātņu pētnieki neuzticas IT speciālistu viedoklim un uzskata, ka to paustais viedoklis par uzdevuma sarežģītību nav gluži atbilstošs īstenībai. Tāpēc humanitāro zinātņu pētnieki izsakās par nepieciešamību abām pusēm apgūt otras puses “valodas” pamatus. Daži intervētie eksperti minējuši arī to, ka paši jau darbojas, lai iegūtu pamatzināšanas programmēšanā. Tāpat ticis pieminēts arī tas, ka laika gaitā viena puse (piemēram, matemātiķi, IT speciālisti) ir kļuvusi krietni zinošāka par otras puses (piemēram, valodnieku) vajadzībām un ierobežojumiem.

Turpmākajā idejas saskaņošanas un attīstīšanas posmā notiek savstarpējas konsultācijas, pēc vajadzības piesaistot (vai tikai informējot par nepieciešamo) arī citas puses (piemēram, grafiskā dizaina speciālistu). Apjomīgāka resursa izstrādē (pretstatā viena pētnieka vajadzības risināšanai) kāds no iniciatīvā iesaistītajiem kļūst par projekta vadītāju, pretējā gadījumā process var “aiziet pašplūsmā”.

Par procesa ietekmētājiem eksperti nosaukuši vairākus faktoros. Pašsaprotami faktori ir laiks un finansējums. Nedaudz citā griezumā situāciju ietekmē vajadzīgo speciālistu pārslodze un aizņemtība citos projektos – kā atzīst DH eksperti, tā ir sistēmas iezīme, kas nav attiecināma tikai uz kādu konkrētu institūciju, jo Latvijas zinātniskajās institūcijās ikviens augsti kvalificēts pētnieks darbojas pārāk daudzās lomās vienlaicīgi.

Kā uzsver daži no ekspertiem, izšķirošs faktors ir spēja vienoties par precīzu mērķi (“kādam jābūt gala rezultātam”) – tā, lai visas puses runā par vienu un to pašu. Sākumā ideja un funkcionalitāte ir neskaidra, un, pēc vairāku ekspertu IT speciālistu domām, pētnieks parasti nevar definēt, kas viņam vajadzīgs, tāpēc nepieciešams IT speciālista atbalsts idejas pārvēršanai vajadzīgajos soļos.

Šobrīd pētnieks nāk un mēs mēģinām rast risinājumu kopīgās sarunās, jo ar precīziem uzstādījumiem neviens nenāk. (Freimanis 2021)

Galvenais ir vienotas izpratnes panākšana tās stratēģiskās grupas vidē, kas nodarbojas ar izstrādi. Šī izpratne jātestē pirms resursa radīšanas – kā tas varētu strādāt vai ne. Tā ir, manuprāt, lielākā problēma. Tas ir jautājums pirms kodēšanas. (Brūvelis 2021)

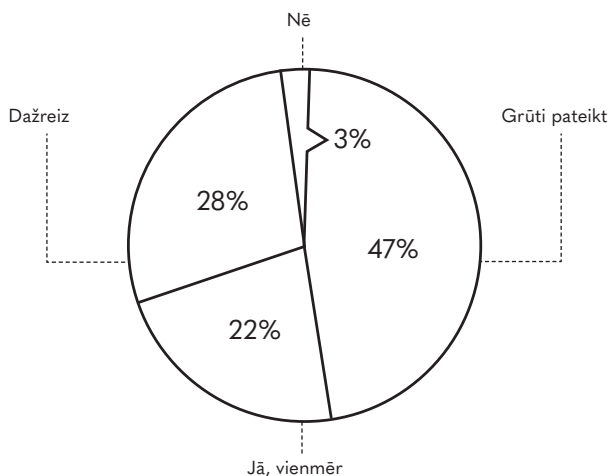
Daži DH eksperti runājuši par šī salāgošanas procesa sarežģītību, savukārt citi norādījuši pat to, ka process var novest arī pie “konflikta” – ja izrādās, ka iegūt vēlamo rezultātu nevar bez papildu darbībām (piemēram, datus salīdzināt var tikai tad, ja tie ir apstrādāti, klasificēti). Var

izrādīties arī, ka, tiekot līdz punktam X, kļūst skaidrs, ka risinājumam nepieciešams daudz vairāk resursu, nekā tie ir pieejami.

Par mērķi vienoties ir grūti, jo pieredze ir salīdzinoši maza; var darboties iteratīvi, bet tā var arī beigties laiks vai nauda. (Freimanis 2021)

Mūsdienās zinātniskajā darbībā datu pārvaldības plāna izstrāde, iesniedzot projekta pieteikumu, vai dažādas standartizētas projektu vadības prakses šķiet ne tikai pašsaprotamas, bet arī arvien biežāk obligātas. Latvijas zinātnes ekosistēmā tā vēl nenotiek, un to apliecina arī šis pētījums. Sistemātiska plānošana un dokumentēšana sekmē zināšanu pārmantošanu un pēctecību DH RR izstrādē un uzturēšanā, kā arī liecina par kontrolētu, vismaz daļai iesaistīto aģentu pārskatāmu procesu. Aptaujā puse respondentu uzskata, ka izstrādes un uzturēšanas procesā dokumentēšana notiek, savukārt otra puse izvēlas atbildes variantu “Grūti pateikt” (sk. 3. grafiku). Līdzīgā jautājumā par aktivitātēm, ko darba grupa veikusi resursa N izstrādes procesā, variantu “Bija izveidots dokuments ar resursa veidošanas, attīstības un uzturēšanas plānu, stratēģiju” izvēlējušies 39 % respondentu, bet “Resursa veidošanas process tika dokumentēts” – 31 % (sk. 7. grafiku).

“Vai [institūcija] dokumentē savu DH resursu izveidošanas un uzturēšanas procesus? Jautājums attiecas gan uz projekta, gan uz koda dokumentēšanu.”



n=85

3. grafiks. DH RR izveidošanas un uzturēšanas procesu dokumentēšana.

Apkopojot šajā sadaļā analizēto par DH RR idejas, saskaņošanas un izstrādes posmu, kā svarīgākos var izcelt dažus aspektus. Pirmkārt, visbiežāk tieši pētnieku privātiniciatīva un brīvprātīgais darbs ir jaunu DH resursu vai rīku veidošanas sākums. Tas notiek, domājot par savām vajadzībām, resursiem savā un savu institūciju rīcībā. Ir grūti identificēt DH RR pasūtītāju, kas būtu aģents ārpus pētnieciskās un atmiņas institūcijas. Valsts pasūtījums nereti tiek

īstenots netiešā veidā, izmantojot zinātnes vai citu politiku instrumentus. Otrkārt, DH RR izveides nepieciešamības centrā visbiežāk ir dati, kas vai nu ir jāievāc, izmantojot šos instrumentus, vai arī jāapstrādā (jādigitalizē, jāstrukturē, jānodrošina piekļuve). Treškārt, izstrādes posmā lielākās grūtības sagādā sadarbība starp DH laukā divu plašāk pārstāvēto nozaru – IT un humanitāro zinātņu – pārstāvjiem. Sapratne starp šo divu “nometņu” pārstāvjiem tiek panākta tikai ilgā sadarbības (mēģinājumu un kļūdu) procesā, jo vairuma zinātnieku un speciālistu izglītība un iepriekšējā darba pieredze ir šauri specializēta, un līdzšinējā pieredze oriģinālu Latvijas DH RR izstrādē ir bijusi salīdzinoši neilga. Ceturtkārt, ne tikai Latvijas DH laukam tipiska ir pastāvīga, nevis projektveida, finansējuma problēma. Piektkārt, tā ir neapmierinoša DH RR izstrādes un uzturēšanas dokumentēšanas prakse.

Mērķauditorija

Pētījumā īpaša uzmanība pievērsta lietotāju lomai DH RR izstrādes un uzturēšanas procesā. Mērķauditorijas pārstāvjus var iesaistīt dažādos veidos un posmos: pašā procesa sākumā (vēl tikai plānojot rīka/resursa izveidi), testējot topošu resursu vai tā prototipu, analizējot jau gatava resursa izmantošanu, komunicējot ar ieplānoto auditoriju, lai to piesaistītu izmantošanai. Reti tiek izmantoti visi veidi. Tāpat būtiski norādīt, ka ne vienmēr tā ir faktiska lietotāju iesaiste: resursa idejas izstrādātāji var darboties ar “darboties ar domu par noteiktu mērķgrupu”.

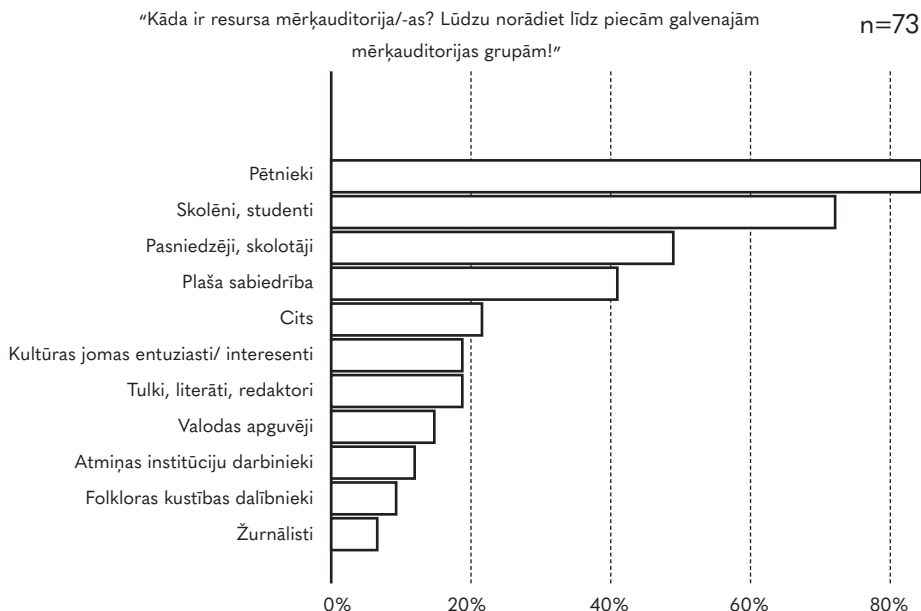
Jā, mēs izstrādājot nevaram nedomāt. Tad tā mērķauditorija, potenciālais adresāts – ja mēs to neredzam, tas mums traucē būvēt pašu resursu. (Martena 2021)

Kā jau vairākkārt norādīts, bieži vien pētnieki DH RR izstrādā sev vai saviem kolēģiem, tādējādi paši būdami mērķauditorija – izstrādātāji un lietotāji vienā personā. Tas ir viens no attaisnojumiem tajos gadījumos, kad praktiski netiek veikti īpaši pasākumi lietotāju izpētei vai komunikācijai. 85 % aptaujas respondentu atbildējuši, ka viņu resursa mērķauditorija ir pētnieki. Citas biežāk minētās grupas ir studenti, skolēni (73 %), pasniedzēji, skolotāji (41 %), plašāka sabiedrība (40 %) (sk. 4. grafiku).

Mērķauditorijas un to vajadzību apzināšanos eksperti norāda kā īpaši svarīgu aspektu resursa izstrādes posmā, lai to izveidotu lietotājam piemērotu. DH RR lietotāju izpētei pēc tā izveides bieži nav jēgas, jo pārveide ir daudz komplicētāka nekā tā izveide.

Pārtaisīt kaut ko nozīmē pārdefinēt datu struktūras, pārdefinēt klasifikācijas sistēmu – tas nozīmē roku darbu (...). Testēšana var radīt kaut kādus “piepucējumus” izveidotajam produktam, bet šai testēšanai būtu jābūt daudz vairāk ideju projekta līmenī, piesaistot arī ārējos lietotājus – pasakot, mums būs šāds interfeiss, vai jums patīk šī poga labajā pusē (...), ņemot līdztīgos piemērus un paskatoties, kā tie strādā.” (Brūvelis 2021)

Vienlaikus lietotāju izpēte sākumposmā ir retums. Kā norāda DH jomas eksperti, galvenais šķērslis ir laika un citu resursu trūkums.



4. grafiks. Resursa mērķauditorijas.

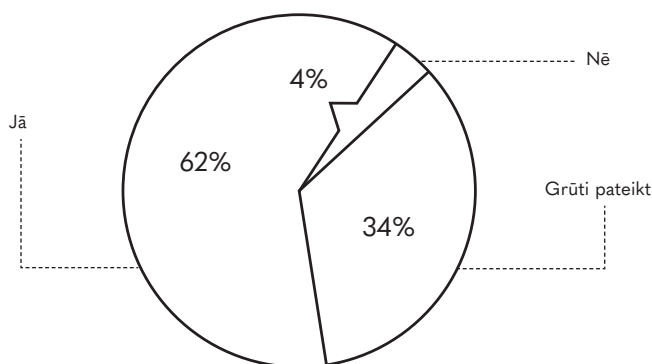
Atbilžu sadalījums jautājumā: "Kāda ir resursa N mērķauditorija/-as? Lūdzu, norādiet līdz piecām galvenajām mērķauditorijas grupām!". Pārkodēts vairākatbilžu atvērtais jautājums.

Varētu jau gribēt testēt bieži, bet laiks visiem ir dārgs. Tāpēc drīzāk tam jābūt integrētākam darba plūsmā – nevis par visu kopā veic aptaujāšanu, bet katrā konkrētā brīdī ir iespēja kaut kādu atgriezenisko saiti saņemt no lietotāja. (Grūzītis 2021)

It īpaši svarīgi tas ir lielajiem resursiem. Katrā ziņā mēs esam diezgan daudz domājuši par to, kā to pielāgot dažādām mērķauditorijām. Respektīvi, kā vienkāršot. Jo mēs paši varbūt labi orientējamies un jūtamies ērti, bet aptaujas rāda, ka cilvēki maldās un īsti nesaprot, utt. Un tas mums liek domāt, kā to pārveidot. (Reinsone 2021)

Arī vēlākos DH RR dzīves posmos lietotāju vajadzību zināšana ietekmē resursa pilnveidošanas procesu. Resursu popularitāte ir īpaši svarīga tad, ja tie tiek uzturēti, pilnveidoti ar digitālās talkošanas (*crowdsourcing*) palīdzību – tad ir svarīgs mērķtiecīgs darbs pie mērķauditorijas uzrunāšanas, kā arī attiecību uzturēšanas, gūstot atgriezenisko saiti. Digitālā talkošana ir īpašs gadījums, kad par sabiedriskajiem zinātniekiem (*citizen scientist*) kļūst brīvprātīgie sabiedrības pārstāvji – vācot, apstrādājot, labojot vai papildinot datus (sk. Causer, Terras 2014; Hedges, Dunn 2017; Reinsone 2020 u. c.). Tomēr Latvijas DH RR izstrādē iesaistītajiem ne vienmēr ir gana zināšanu, tāpēc tiek izmantots arī citu institūciju atbalsts.

“Vai bija/tiek veiktas kādas aktivitātes, lai piesaistītu potenciālo mērķauditoriju [konkrētajam resursam]?”



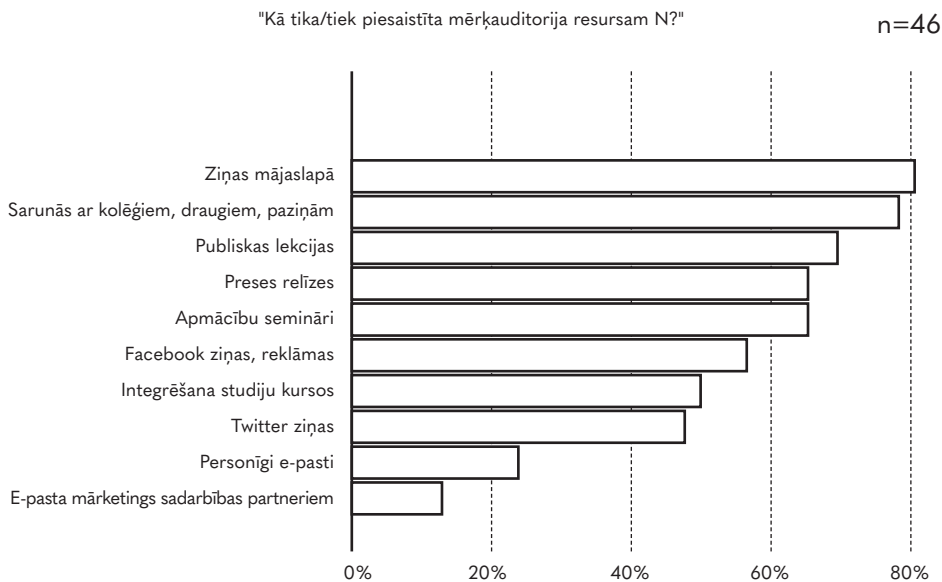
n=78

5. grafiks. “Vai bija/tiek veiktas aktivitātes mērķauditorijas piesaistīšanai?”.

Bet pūlpakalpes [crowdsourcing] rīkiem gan ir vajadzīga popularizēšana, nežēlīgs darbs. “Vietvārdu talkā” mums daudz palīdzēja UNESCO Latvijas Nacionālā komisija; viņi daudz ko saprot no šīm lietām – vilka uz intervijām, teica, ka jāstāsta, citādi nekas nebūs (...), jā māca, jā rāda. Braukājām pa visu Latviju, radio, televīzijā (...), rīkojām visādas akcijas, jo vienu reizi pateikt nepietiek. Un cilvēkiem vajadzīga atgriezeniskā saite – vai viņu devums ir bijis vērtīgs. Mēs ap 1000 vēstulēm uzrakstījām, lai cilvēkus uzmundrinātu. Bez atgriezeniskās saites nekā nebūs.
(Rapa 2021)

Lai arī kāds būtu bijis konkrētais DH RR, aptaujā tikai 4 % norādījuši, ka nav veikuši aktivitātes mērķauditorijas piesaistīšanai konkrētajam resursam, bet 62 % atbildējuši apstiprinīgi. Tomēr 35 % nav varējuši uz šo jautājumu atbildēt. Tas var liecināt ne tikai par to, ka DH RR izstrādē un uzturēšanā parasti piedalās vairāki cilvēki, kuru uzdevumi un kompetences būtiski atšķiras, bet arī par to, ka pat ievērojamai daļai darbinieku, kas strādā vai strādājuši pie konkrētā rīka vai resursa izveides un/vai uzturēšanas, nav pārliecības, ka šādas aktivitātes ir notikušas vai notiek. Jāpiezīmē, ka šis nav vienīgais aptaujas jautājums, kurā liela respondentu daļa izvēlējusies atbildes variantu “Grūti pateikt”.

Nākamais aptaujas jautājums tiem, kas uz iepriekšējo jautājumu atbildējuši apstiprinīgi, ir par mērķauditoriju piesaistes veidiem. Interesanti, ka grupējot visbiežāk minēta mērķauditorijas piesaiste, izmantojot strukturētas klātienē aktivitātes – publiskās lekcijas, apmācību, sarunas ar kolēģiem, draugiem un paziņām, integrāciju studijuursos. Protams, arī tīmekļvietņu, sociālo mediju un preses rīku izmantošana ir ļoti populāri veidi.



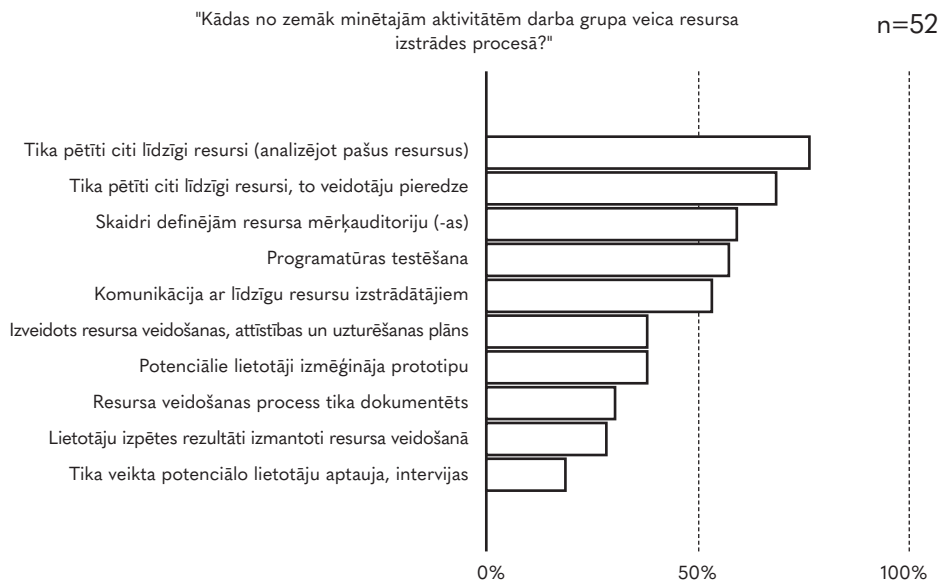
6. grafiks. Mērķauditorijas piesaistīšanas veidi.

Lai arī eksperti norāda, ka lietotāju vajadzību apzināšanas metodes būtu izmantojamas gan resursa izstrādes, gan pilnveidošanas procesā, tiek atzīmēts, ka Latvijā tam nereti nepietiek ne tikai laika un finansējuma, bet arī zināšanu. Tiek aprakstīti gadījumi, kad lietotāju vajadzības un vērtējums tiek apzināts nevis ar mērķtiecīgu testēšanu vai aptaujāšanu, bet gan pastarpinātā ikdienas komunikācijā, kad lietotāji sniedz atgriezenisko saiti.

Vajadzību apzināšana tomēr notiek, bet tā nenotiek tā mērķtiecīgi, izveidojot aptauju un prasot cilvēkiem (..), bet tas tīri dabiski mums notiek, mums raksta vēstules, zvana, stāsta, pat nāca un institūtu un teica, ko viņi gribētu labāk... bet tas nenotika tā no mūsu puses. (Rapa 2021)

Analizējot respondentu atbildes uz jautājumu par aktivitātēm, ko darba grupa veikusi konkrētā resursa izstrādes procesā, redzams, ka potenciālās mērķauditorijas pārstāvju iesaiste šajā posmā ir salīdzinoši reta prakse (sk. 7. grafiku). Izteikti dominē nevis faktiska (primāra), bet gan sekundāros datos balstīta mērķauditorijas izpēte: tikusi skaidri definēta resursa mērķauditorija (-as); tikuši pētīti citi līdzīgi resursi, to veidotāju pieredze (lasot publikācijas); tikuši pētīti citi līdzīgi resursi (analizējot pašus resursus). Kā jau norādīts iepriekš, iemesli varētu būt saistāmi ar dažādu resursu un kapitālu – laika, atsevišķa finansējuma, arī zināšanu – trūkumu DH RR komandās. Vienlaikus eksperti norāda arī uz atgriezeniskās saites pastāvēšanu citās formās.

Parasti nav resursu lietotāju izpētei. Bet cenšas visur norādīt kontaktus, lai var sazināties, ja lietotājs novēro kādu problēmu. Ja vien nav sensitīvu datu, cenšas pieiet no atvērto datu



7. grafiks. Aktivitātes resursa izstrādes posmā. Atbilžu sadalījums jautājumā: "Kādas no zemāk minētajām aktivitātēm darba grupa veica resursa N izstrādes procesā?". Vairākatbilžu jautājums.

perspektīvas, kur katrs lietotājs var piedāvāt savas korekcijas, kļūdu ziņojumus. Ja ir vienots formāts, tad arī citu valstu pētnieku komandas var ko komentēt – tāpēc svarīgi to ievērot. (Grūzītis 2021)

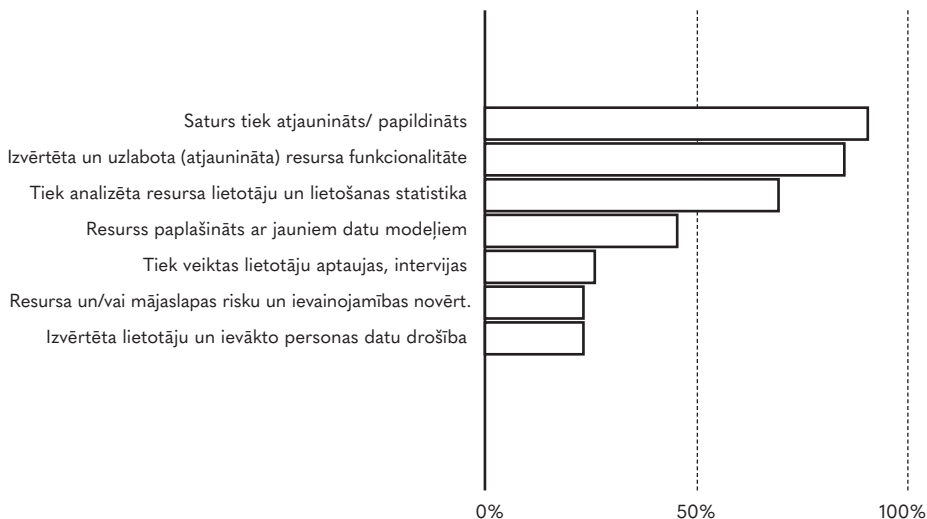
DH RR lietotāju iesaiste resursu veidošanā ir prakse, par kuras nepieciešamību nav šaubu, taču projektu realitātē tā bieži notiek samērā ierobežotā veidā. Daļēji to kompensē resursa izstrādes darba grupas un citu pētnieku iesaiste resursa testēšanā, kā arī atgriezeniskā saite no lietotājiem ikdienas darbību līmenī, tomēr eksperti norāda, ka svarīga būtu ideju un prototipu testēšana jau pirms kodēšanas, jo sistēmas pārveidošana ir ļoti apjomīgs darbs.

Uzturēšana, pilnveide un ilgtspēja

DH RR ilgtspēja, kā norādīts DH teorētiskajā literatūrā, ir augstākas pakāpes uzdevums un joprojām salīdzinoši jauna prasība. Tas atbilst gan vispārīgajam ilgtspējīgas attīstības politiskajam diskursam, gan specifiskajiem atvērtās zinātnes principiem, kas arī Latvijas valsts pārvaldē arvien vairāk kļūst par dominējošu ideoloģiju ar tai sekojošajām izmaiņām zinātnes politikā, finansējumā un zinātniskās darbības izvērtējumā.

"Vai pēc resursa izveides tiek veiktas zemāk minētās darbības?"

n=76



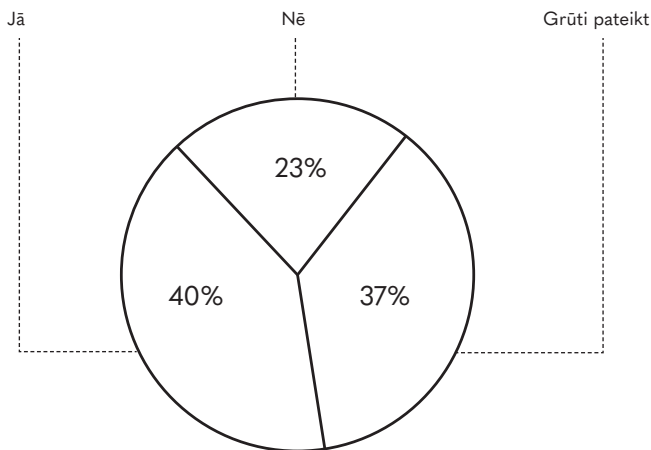
8. grafiks. Darbības pēc resursa izveides.
Atbilžu sadalījums jautājumā: "Vai pēc resursa
N izveides tiek veiktas zemāk minētās darbības?"
Vairākatbilžu jautājums.

DH RR izstrādātāju un uzturētāju aptaujas dati liecina, ka darbs pie RR turpinās arī pēc to izveides. Vienlaikus, kā parādīts vēlāk, ne vienmēr satura un funkcionalitātes papildināšana var turpināties ilgstoši – to nosaka resursu atkarība no projektu finansējuma, kas ir pretrunā ar nepieciešamību DH RR pastāvīgi attīstīties.

Pie rīka nemitīgi jāstrādā. Dati jāintegrē citos resursos tā, lai tie tiek joprojām izmantoti. Arī tāpēc, ka rīku pašlaik jau ir par daudz, nevar saprast, kur kas atrodas. Piemēram, vietvārdus var atrast četrās vietās. Rīka integrēšana/uzturēšana ir svarīga, jo piemēram, "Mūsdienu latviešu valodas vārdnīcā" katru nedēļu, katru mēnesi rodas jauni vārdi. Arī slenga vārdi būtu jāiekļauj kādā apakšsadaļā, lai jaunieši zina, ka to var pateikt arī latviski. Visam vajadzētu satecēt vienā laukā.
(Rapa 2021)

Lai arī eksperti norāda, ka to izveidotāji no vienas puses ir motivēti, ka resursi sasniedz plānoto auditoriju, resursa popularizēšanai nereti netiek atvēlēti resursi, jo primārā nepieciešamība ir RR izveide. Eksperti iezīmē tendenci, ka pie resursa popularizēšanas mēdz atgriezties jau pēc resursa izstrādes pabeigšanas citos projektos, kur tā izvirzīta kā prioritāte.

“Vai ir bijuši gadījumi, kad [Jūsu institūcijas] izveidots un funkcionējošs DH resurss tiek pamests novārtā (netiek atjaunots/papildināts saturs un programmatūra), beidzoties finansējumam vai citu iemeslu dēļ?”



n=84

9. grafiks. Izveidots resurss pamests novārtā.

Ja tu koncentrējies uz abiem [izstrādi un popularizēšanu], tad tu neizdarīsi neko.
(Martena 2021)

Ja finansējuma apjomā rodas kāda problēma, tad šī aktivitāte ir viena no pirmajām, kas uzreiz cieš (...). (Freimanis 2021)

Būtiska nianse ir arī tā, ka projektu budžetā parasti ir ierobežojumi iespējamā līdzekļu izlietojumā, un tāpēc to veidotāji var izvēlēties neieklāut nepieciešamo finansējumu kādai DH resursa izstrādes vajadzībai (piemēram, datu vākšanai). Īpaši šie ierobežojumi attiecas uz uzturēšanas fāzi, kas atbilst arī citur īstenoto pētījumu secinājumiem par to, ka finansētājus interesē unikāli, inovatīvi projekti, nevis jau pastāvošo resursu darbības turpināšana. Šī tēma spilgti izskanējusi DH ekspertu komentāros par resursiem, kuru uzturēšanai nav līdzekļu vai kuru izstrāde paliek pusceļā finansiālu ierobežojumu dēļ. Šādi veidojas “mirušo” vai “pusdzīvo” resursu slānis, kas ir nopietna ekosistēmas problēma vismaz divu iemeslu dēļ: līdzekļu neefektīvs izlietojums un apgrūtināta aktuālo (nevis novecojušu, nepapildinātu) datu meklēšana.

Uz jautājumu, vai respondentu pieredzē ir bijuši gadījumi, kad viņu institūcijas izveidots un funkcionējošs DH RR ir pamests novārtā, 41 % respondentu atbildējuši apstiprinoši. 24 % respondentu nav saskārušies ar šādu situāciju, taču lielai daļai (35 %) bijis grūti atbildēt uz šo jautājumu. Ja neņem vērā tos, kuriem bijis grūti atbildēt, tad gandrīz ¾ atbild, ka viņu pieredzē ir izveidoti, bet vēlāk pamesti DH RR.

Kopumā izstrādāto DH RR ilgtspējas nodrošināšana ir resursietilpīga. Resursu izstrādātāji parasti ir ieinteresēti, lai resurss sasniedz savu mērķauditoriju un turpina funkcionēt, bet projektos balstītā zinātnes pārvaldes sistēma un Latvijas zinātnes mazais finansējums¹¹ nozīmē to, ka ir visai liels risks, ka pēc izstrādes DH RR var tikt pamests novārtā.

Izstrādātāju priekšstati par labu resursu

DH rīki un resursi sākas ar noteiktu ieceri, kas kalpo par atskaites punktu tālākajām darbībām izstrādes posmā. Protams, rezultāts var pilnībā nesakrist ar šo sākotnējo ieceri, bet pētījumā bijis svarīgi noskaidrot, ko paši izstrādātāji uzskata par sava veida ideālo DH RR, lai pēc tam salīdzinātu, kā šie priekšstati par labu rīku materializējušies gala produktos un cik tie ir tuvu teorētiskajā literatūrā paustajām atziņām. Aptaujā un intervijās, jautājot par galvenajām pazīmēm, kas veido labu DH RR, nošķirti pētniekiem un plašakai sabiedrībai paredzēti rīki/resursi.

Aptaujā no sākuma lūgts nosaukt līdz trim labākajiem Latvijas DH resursiem, kas ir paredzēti pētniekiem. Visbiežāk nosaukti tieši DH VPP iesaistīto institūciju veidotie rīki un resursi: *garamantas.lv* (48 %), *korpus.lv* (ieskaitot vairākus atsevišķi nosauktos korpusus) (47 %), *periodika.lv* (29 %), *tezaurus.lv* (25 %), *literatura.lv* (16 %), LNB Digitālā bibliotēka (15 %). Populārākie citu institūciju resursi, piemēram, *raduraksti.lv*, *eresursi.arhivi.gov.lv*, *termini.gov.lv*, *data.gov.lv*, *hugo.lv*, nosaukti vien dažas reizes.

Nākamais aptaujas jautājums “Kāds, Jūsaprāt, ir labs digitālais resurss, kas ir paredzēts pētniekiem?” uzdots, lai precizētu laba DH RR iezīmes. Biežākā atbilde liecina par to, ka respondentiem svarīgākais ir plašs, aktuāls un arī drošticams saturs (52 %). Nedaudz retāk norādīts ērtums un vienkāršība (44 %), labi izstrādāts meklēšanas rīks, kas ļauj atlasīt datus, izmantojot dažādus filtrus (36 %), atvērtās zinātnes principu ievērošana un brīvpieejas nodrošinājums (21 %), responsīvs, patikams dizains (18 %) un iespēja apstrādāt un lejupielādēt datus (17 %), precīzi, skaidri metadati (12 %).

Plašs un aktuāls saturs, kas aptaujā tiek norādīts kā īpaši svarīga laba DH RR pazīme, retāk tiek minēts ekspertu intervijās. Taču tiek norādīts, ka ir svarīgi regulāri atjaunot saturu un iegūt pēc iespējas detalizētāku informāciju/datus interesējošajā jautājumā.

Man, valodniecei, būtu svarīgi, ka tas resurss ir bagāts ar valodas materiālu un “apdeitots”, atjaunots pēdējo triju gadu laikā ar jaunām valodas vienībām, ar vārdnīcu materiāliem. (Martena 2021)

Pētniekiem ir svarīgi, ka ir daudz dažādu sadaļu; nu, piemēram, ja vārdam mēs (valodas resursā galvenais ir vārds) dodam pēc iespējas dažādas apakšsadaļas: cilme,

11 Tas ilgus gadus ir viens no vismazākajiem ES (sk. Research and development expenditure, by sectors of performance (online data code: TSC00001); Eurostat; OECD. Pieejams: <https://ec.europa.eu/eurostat/databrowser/view/tsc00001>).

locījumi, paradigma, apraksti, avoti, kur tas ir minēts; valodniekiem ir arī svarīgi meklēt pēc vārda daļām. Katru daļu aprakstīt. Arī pēc nozīmēm meklēt, arī pēc cilmes. (Rapa 2021)

Attiecībā uz meklēšanas rīkiem, kas aptaujā identificēta kā viena no nozīmīgākajām iezīmēm labam DH RR tieši pētniekiem, eksperti komentē, ka pētniekiem ir svarīgi, lai datus var atlasīt pēc detalizētām, skaidrām kategorijām un ka atlases rezultāti ir uzticami. Tomēr eksperti arī piebilst, ka tieši rezultātu uzticamība ir vēl pilnveidojams aspekts Latvijas DH laikā, jo viņi pētniecības darbā ir pieredzējuši situācijas, kad jūt vajadzību pārliecināties par atlases rezultātu precizitāti. Turklāt ekspertu intervijās minēti pretēji gadījumi: gan tādi, kad pētnieks jūtas zaudētājs, kad kāds labi strādājošs meklētājs pazūd vai vairs nefunkcionē, gan tādi, kad izstrādāts komplicēts meklētājs netiek praktiski izmantots.

Pētniekiem parasti gribas ļoti sadrumstaloti, ļoti specifiski norādīt katru parametru atsevišķi (...). Nu, kad tu uztaisi tādu plašu [meklētāju], un tad vienā brīdī tu saproti, ka to nemaz neizmanto. Tu pamani, ka ir kaut kāda kļūda, ka sen jau kādam vajadzēja teikt, ka, re, šis nestrādā, bet neviens pat nav pamanījis, ka ir tāda problēma. (Ķirsis 2021)

Pētnieku darbā īpaši svarīga ir iespēja lejupielādēt un apstrādāt datus ārpus DH RR IT sistēmas, kā arī tas, lai šie dati tiktu regulāri papildināti un būtu pieejamas jaunākās to versijas. Šī ir viena no DH RR ilgtspējas pazīmēm, kas Latvijas DH laikā ir tipiski problemātiska. Nereti pēc projekta beigām, kurā RR tiek izstrādāts, netiek ielānoti resursi regulārai datu papildināšanas funkcijas nodrošināšanai.

Savukārt pētniecībā daudz svarīgāks ir tas, lai tev būtu piekļuve arī ārpus tā, ko piedāvā sistēma. Respektīvi, svarīgi ir tas, lai varētu lejupielādēt informāciju dažādos veidos. Piemēram, veic kaut kādu materiālu atlasīšanu, un tad tu šo indeksu izveidoti ar visām saitēm vari saglabāt kaut vai pie sevis, lai tu vari tālāk strādāt. Vai arī lejupielādēt jau veselus korpusus. (Reinsone 2021)

Kā atšķirību no resursiem, kas tiek veidoti plašākai sabiedrībai, eksperti iezīmē, ka pētniekiem svarīgāk ir piekļūt neagregētiem datiem, veikt analīzi pašiem. Savukārt plašākai sabiedrībai paredzētajiem resursiem jādomā par risinājumiem, kas veiktu datu strukturizēšanu, apstrādi viņu vietā.

Parastais lietotājs darbojas ar datiem, kas ir jau pilnībā apstrādāti, un viņam tiek piedāvāts rezultāts, kas ir praktiski ļoti līdzīgs rezultātam, kādu viņš redzētu fiziskajā pasaulē. (...) Nišas pētnieciskajā daļā būtu pictuvinājums "jēlajiem" datiem. Ja šie dati ir strukturēti, tad saskare ar tiem ir kā ar tabulām, ne caur dizaina pastarpinājumu. (Baklāne 2021)

Aspekts, ko eksperti uzsvēruši tikai pētniekiem paredzēto resursu kontekstā, ir to savietojamība/sadarbspējība (*interoperable*), starptautiski atzītu metadatu shēmu izmantošana un precīza metadatu aprakstīšana, tā vēlreiz akcentējot datu pieejamības nozīmi pētniecības procesā.

Ļoti būtiska ir šī “pārrobežas” savietojamība – vai ir marķēti teksta korpusi, vai tie ir marķēti standartizētā veidā, vai pētnieki citās valstīs var veidot pētniecību, balstoties uz šiem datiem. Īstenībā līdzīgs princips ir arī citiem resursiem, kas ir datubāzes. Pat ja tur ir kaut kādas muzeja vai mākslas kolekcijas digitalizētas, tiem arī vajadzētu būt ar metadatiem aprakstītiem standartizētā veidā, lai tie būtu pamanāmi un izmantojami pētniekiem citās valstīs. (Blūms 2021)

Pārejot pie Latvijas DH RR vērtēšanas, kas paredzēti plašam lietotāju lokam, no sākuma atkal lūgts nosaukt labākos. Kā visbiežākās atbildes ir norādīts LU LFMI resurss *garamantas.lv* (45 %) un LU MII veidotā vārdnīca *tezaurs.lv* (44 %). Trešdaļā gadījumu norādīts LNB resurss *periodika.lv* (32 %), bet 13 % atbilžu minēts LU LFMI *literatura.lv*. Jāņem vērā, ka aptaujas respondenti bijuši zinātnieki vai zinātniskie asistenti, un viņu viedoklis varētu atšķirties no “plašāka lietotāju loka” vajadzībām, turklāt šeit viņi faktiski vērtē arī savas institūcijas izstrādātos resursus.

Jautājumā par labu DH RR, kas paredzēts plašam lietotāju lokam, visbiežāk minēta vienkāršība un ērtums (73 % gadījumu). Daudz retāk norādīts plašs, aktuāls saturs (30 %), patīkams dizains (30 %), izstrādāts meklēšanas rīks (19 %) un datu brīvpieeja (18 %). Tātad faktiski apstiprinās teorētiskajā literatūrā bieži minētā nepieciešamība pēc vienkāršības un “intuitīvi saprotama” dizaina.

Ekspertu intervijās parādās līdzīgas tendences – tiek uzsvērtā resursa vienkāršības un saprotamības nozīme. Neveiksmīga pieredze, nespējot atrast veidu, kā izpildīt kādu no vēlamajām darbībām konkrētajā resursā, lietotāju var atturēt no atgriešanās tajā atkārtoti.

Samazināt to sliekšni, lai vispār sāktu izmantot resursu, jo katram resursam ir savs sarežģītības sliekšnis. Vārdnīcai sliekšnis parasti ir zems, jo vārdnīcas mēs visi saprotam, vienmēr esam lietojuši. Korpusiem sliekšnis ir augstāks, bet to mēs arī cenšamies pazemināt. (Grūzītis 2021)

Ekspertu vērtējumā viens no veidiem, kā samazināt šo iesaistes sliekšni, ir ne tikai veidot resursus vienkāršus un viegli saprotamus, bet arī atvērtus un brīvi pieejamus, jo līdzšinējā pieredze rāda, ka pat reģistrēšanās potenciālajam lietotājam var sagādāt grūtības vai nepatiku. Īpaši aktuāli tas ir resursiem, kas tiek papildināti ar digitālās talkošanas (*crowdsourcing*) palīdzību, tādējādi aicinot uz aktīvu līdzdalību, kas savukārt paredz to, ka RR jābūt saprotamam lietotājiem ar dažādiem RR izmantošanas prasmju līmeņiem.

Lietotājam ir ļoti svarīgi, lai rīks ir vienkāršs un saprotams, lai nav ļoti daudz sadaļu, kuras viņš vienkārši nesaprot; tas vecuma posms, kad gribas iesaistīties un valodai palīdzēt bagātināties, īsti nepārvalda datorus. Vēl ir svarīgi, lai ir atvērtā piekļuve. Atceros, ka bija grūtības vienkārši reģistrēties, lai krātu un lai pierakstītu šos vietvārdus. (Rapa 2021)

Līdzīgi kā tas parādās aptaujas datos, arī intervijās eksperti akcentē patīkama, saprotama resursa dizaina nozīmi, uzsverot tā lomu arī lietotāja pieredzes (*user experience*) veidošanā.

(..) lai telpa, kur es strādāju, man būtu patīkama. Es virtuālo telpu uztveru lielā mērā tāpat kā fizisko, un es domāju, ka manai darba telpai, manam arhīvam, lasītavai, kurā es strādāju, būtu jābūt mājīgai. (Baklāne 2021)

Diezgan svarīgi lietot labo praksi produktu dizainā – izmantot dizaina domāšanas metodes, user experience, user interface un pašu dizainu uztaisīt glītu. Man liekas, ka tādas lietas ir ļoti būtiskas. (Blūms 2021)

Savukārt attiecībā uz meklēšanas rīkiem tiek uzsvērts ne tikai resursa iekšējās meklēšanas rīku nozīme, bet arī datu un satura indeksēšanas nozīme, kas cita starpā DH RR saturu “savieno” ar tādiem globāliem meklētājiem kā *Google*. Tas ir viens no izplatītiem ceļiem, kā lietotāji nokļūst līdz resursam, nereti arī tādi, kas līdz šim par tā esību nav zinājuši.

Ar globālajiem lietotājiem mūsdienās, man liekas, ir tā, ka viņi pat arī neizmanto [konkrētā resursa] meklētāju, viņi izmanto globālos meklētājus – Google, piemēram. Un tev ir jāpanāk, ka tie dati ir veiksmīgi indeksēti [globālajos meklētājos]. Viņi no meklētāja ienāk, paskatās, ko viņi meklē, un praktiski uzreiz iziet. (Kīrsis 2021)

Lai salīdzinātu DH RR izstrādātāju uzskatus par laba DH RR iezīmēm un to iestrādāšanu viņu pašu veidotajos resursos, aptaujā viņi tikuši lūgti izvēlēties un atzīmēt visas atbilstošās funkcijas, kuras apsvēruši iekļaut savos DH RR izstrādes posmā. Visbiežāk atzīmētā atbilde uz šo jautājumu ir intuitīvi saprotamas saskarnes izveidošanas nepieciešamība (74 %), nedaudz retāk apsvērta estētiska dizaina izveide (65 %). Nedaudz vairāk kā pusē (57 %) gadījumu atzīmēts adaptīvs, reaģējošs dizains un lietotājam pieejami detalizēti metadati. Uz to, ka visbiežāk tiek apsvērtas tieši ar dizainu saistītās funkcijas, norāda arī atbildes atvērtajās atbildēs aptaujā un intervijās. Retāk apsvērtas tādas funkcijas kā iespēja sazināties ar uzturētājiem (48 %), datu eksports (46 %) un imports (39 %), pieejamas un saprotamas instrukcijas (32 %), iespēja lietotājam iesaistīties resursa uzlabošanā un papildināšanā (31 %), bet visretāk norādīta funkcija lietotājam pašam izveidot savas krātuves vai grāmatzīmes (26 %) un resursa mērogojamība (24 %).

Gan teorētiskie avoti priekšizpētē, gan izstrādātāju aptaujas rezultāti norāda uz atvērtās zinātnes principu ievērošanu DH RR izveidē kā vienu no svarīgākajiem elementiem un priekšnosacījumiem laba, ilgtspējīga DH RR izveidē. Tieši tādēļ pētījuma autori vēlējušies pievērst lielāku uzmanību šo principu piemērošanai respondentu pašu izstrādāto rīku praksēs, aptaujā lūdzot atzīmēt DH RR izstrādē ievērotos atvērtās zinātnes principus.

Visbiežāk respondenti centušies nodrošināt, lai DH RR saturs un tajā atrodamie dati ir viegli atrodamā (72 %), nedaudz retāk (67 %) domāts, lai saturs un dati būtu viegli pieejami un izgūstami, nedaudz mazāk par pusi (48 %) gadījumu nodrošināts, lai saturs un dati ir viegli atkārtojami izmantotā, 41 % gadījumu izmantoti atvērtā koda rīki vai programmatūra, 35 % gadījumu iesaistīti amatierpētnieki, 38 % atbilžu norādīts, ka domāts par to, lai dati būtu viegli savietojami, bet 28 % gadījumu izmantots atvērtais kods. Tikai 2 % atbilžu norādīts, ka DH RR izstrādē nav ticis ņemts vērā neviens no iepriekš minētajiem atvērtās zinātnes

principiem. Cik veiksmīgi bijuši šie centieni ievērot attiecīgos atvērtās zinātnes principus, varētu noteikt vien audita ceļā, t. i., izvērtējot konkrēto DH rīku un resursu atbilstību šiem principiem un standartiem.

Ja raugās no atvērtās zinātnes politikas viedokļa, kas gan Latvijā vēl nav ieviesta un adaptēta, bet ir obligāta prasība vairākos pēdējos lielākos ES projektu konkursos, tad, spriežot pēc aptaujas datiem, līdzšinējās Latvijas DH RR veidošanas prakses vismaz dažos rādītājos uzskatāmas par neapmierinošām. Ekspertu intervijās nereti teikts, ka atvērtā zinātne ir diezgan jauna ideoloģija, un zinātnieku kopiena, īpaši no humanitārajām disciplīnām, to tikai tagad sāk apgūt.

Senāk pazīstamais un izmantotais atvērtās zinātnes elements ir atvērtais kods un atvērtā koda programmatūra. Tai pašā laikā valsts pārvaldē un akadēmiskajā vidē tā nebūt vēl nav pašsaprotama prakse un kultūra. Par vienu no iemesliem, kāpēc Latvijas DH laukā izveidota rīka vai resursa kods reti kad tiek atvērts, tiek nosaukta drošība. Oriģināli izstrādāta DH RR kodu varētu publicēt tikai tad, kad tas kārtīgi notestēts pret iespējamajiem “caurumiem sistēmā”, bet tam praktiski nekad nepietiek laika un finanšu resursu (Brūvelis 2021).

Respondentu atbildes būtiski neatšķiras, kad runa ir par pētniekiem un plašai sabiedrībai paredzētiem DH RR, taču ir novērojamas atsevišķas nianšes izpratnē par atšķirībām auditorijas vajadzībās. Tas apstiprina pieņēmumu, ka pirms DH RR izveides svarīgi definēt tā potenciālo mērķauditoriju, kas attiecīgi noteiks daudzus ar izstrādi saistītos jautājumus, piemēram, RR vienkāršību/sarežģītību, datu apstrādes, importa/eksporta funkcijas, satura un formas attiecības u. tml. Tomēr ļoti bieži vienus un tos pašus resursus izmanto gan pētnieki, gan plašāka sabiedrība, kas nozīmē, ka izveidotājiem jāspēj izveidot RR, kas reizē ir gan vienkāršs un saprotams plašai sabiedrībai, gan spēj nodrošināt specifiskas funkcijas, kas nepieciešamas pētniekiem. To parāda arī aptaujas rezultāti – par labākajiem Latvijas DH RR, kas paredzēti pētniekiem un plašai sabiedrībai, tiek minēti praktiski vieni un tie paši resursi. Lai arī tiek izceltas funkcionālās atšķirības, kas tiek sagaidītas atkarībā no tās vai citas lietotāju grupas, nedrīkst automātiski pieņemt, ka zinātnieki vienmēr ir daudz kvalificētāki un izsmalcinātāki lietotāji nekā nezinātnieki. “Profesionāli lietotāji” un “pilnīgi iesācēji” ir abās grupās.

Diskusija un secinājumi

Raksta mērķis ir Latvijas DH rīku un resursu izstrādes un uzturēšanas praksi izvērtējums no to veidotāju perspektīvas un pieredzes, vienlaikus to kritiski salīdzinot ar akadēmiskajā literatūrā aprakstīto. Izpētes objekts ir vairāku lielu valsts akadēmisko institūciju veidotie DH resursi. Šis ir viens no pirmajiem pētījumiem par DH lauku Latvijā, tāpēc izvirzīti samērā vispārīgi pētnieciskie jautājumi, kas pamatā fokusēti uz izstrādātāju un uzturētāju pieredzi resursu veidošanā, kā arī viņu priekšstatiem par labu resursu. Lai īstenotu uzdevumus, veikta Latvijas DH RR izstrādātāju un satura veidotāju interneta aptauja un daļēji strukturētas intervijas, kā arī analizēta šīs jomas teorētiskā un lietišķā literatūra.

Latvijas DH ir salīdzinoši jauns zinātniskās darbības lauks. Aptaujā atklājas, ka vidējais gadu skaits, kas norādīts atbildē uz jautājumu par pavadīto laiku šajā jomā, ir astoņi gadi (mediāna – 5, bet moda – 1 gads), bet apmēram 1/10 atbildējuši, ka šajā jomā strādā 20–30 gadus. Tas apliecina, ka darbs šajā jomā noticis vismaz kopš 80. gadu beigām, bet tas ir intensificējies un paplašinājies tieši pēdējos dažos gados. Ja raugāmies uz šajā rakstā analizētajiem un biežāk minētajiem DH resursiem, tad tie tādā veidolā, kādā funkcionē šobrīd, darbojas tikai 12 vai mazāk gadus: *periodika.lv* (no 2008. g.), *tezaurs.lv* (2009), *corpuss.lv* (2009), *zudusilatvija.lv* (2011), *garamantas.lv* (2014), *literatura.lv* (2017), *enciklopedija.lv* (2018).

Latvijas mērogā kopīgas DH kopienas veidošanas pirmie mēģinājumi sākās tikai pirms dažiem gadiem, līdz ar tīmekļvietnes *digitalhumanities.lv* (2018) izveidi un praktisku kopienas veidošanas pasākumu rīkošanu. Tāpat jāmin CLARIN-LV aktivitātes, īpaši 2020. gadā izveidotais repozitorijs, kurā tiek reģistrēti latviešu valodas resursi un rīki¹². Savukārt institūciju līmenī DH darbos iesaistīto komunikācija varētu būt nepietiekama, un viens no galvenajiem iemesliem ir strikts darba dalījums starp humanitāro zinātņu un IT pētniekiem (vai ārpalpojuma nolīgtajiem IT uzņēmumiem vai speciālistiem), kā arī joprojām problemātiska sapratne un sadarbība šo “nometņu” starpā.

Viens no pētījuma uzdevumiem ir teorētiska DH resursa modeļa izveide un tā empīriskā aprobācija. Sākotnēji pieņemts, ka viena DH resursa ekosistēmā svarīgākie elementi ir trīs veida aģentu grupas jeb lomas: pasūtītāji, izstrādātāji un lietotāji. Protams, arī pats resurss kā IT sistēma. Vienu no pirmajām korekcijām ievieš literatūrā un empīriskajā izpētē bieži norādītais, ka šīs lomas ļoti bieži pārklājas, DH pētnieki paši ir gan resursu izstrādātāji, gan to lietotāji. Ekspertu intervijās atkārtotas stāsti par to, ka DH resursu vai rīku izveide Latvijā ir bijusi pētnieku vai industrijas pārstāvju privātā iniciatīva, kam ziedots neapmaksāts darba un brīvais laiks.

Arī mūsdienās, kad Latvijas DH un vispār zinātnes lauks ir institucionalizētāks (ar skaidrākiem spēles noteikumiem), grūti apgalvot, ka DH resursu izveides pasūtītājs ir kāds ārējs spēks vai aģents. Protams, valsts zinātnes institūcijas savas DH aktivitātes pamatā īsteno, izmantojot valsts finansējumu, pildot valsts pasūtījumu. Tomēr retos gadījumos tas ir tik konkrēts, lai varētu teikt, ka DH rīku vai resursu pasūtītājs ir kāds valsts ierēdnis vai ministrija, un tas joprojām atstāj lielas manevrēšanas iespējas pašu izstrādātāju rīcībā.

Īpaša DH lauka un atsevišķu resursu ekosistēmas dimensija ir attiecības starp dažādiem šajos laukos iesaistītajiem aģentiem, kā indivīdiem, tā to grupām, organizācijām un institūcijām. Pētot sadarbību, skatīts gan mikrolīmenis, gan mezolīmenis, kā noteicošos rīcības faktorus izvirzot aģentu intereses un aģentu kapitālus. Pētījums apliecina, ka šī dažādo mērogu pieeja ir pamatota, jo komunikācija galu galā notiek starp indivīdiem un nereti tik mazā DH kopienā kā Latvija resursa izveidē ir svarīgāks konkrēts cilvēks ar viņa zināšanām, nevis institūcija ar tās statusu vai finansējumu. Tai pašā laikā bez institucionālas sadarbības valsts uzturētie DH resursi nespētu ne rasties, ne pastāvēt. Savukārt jautājums par to, kā kompetencē ir uzraudzīt

12 Pieejams: <https://repository.clarin.lv>

un sekot, lai par sabiedriskajiem līdzekļiem veidoto DH rīku un resursu funkcijas un saturs nepārklātos, drīzāk ir valsts pārvaldes uzdevums.

Analizējot konkrētu Latvijas DH resursu veidošanas un uzturēšanas procesu jeb ciklu, lēmumu pieņemšana un tiem sekojošās darbības raksturojamas kā inkrementālas, t. i., tādas, kas pakāpeniski veic izmaiņas jau pastāvošajā sistēmā, atsaucoties uz attiecīgā brīža situāciju un blakusfaktoru ietekmi. Reti ir gadījumi, kad DH rīks vai resurss tiek izveidots pēc iepriekš izsvērtā un akceptēta plāna. Plānošana, izstrāde, testēšana, papildināšana ar saturu nereti notiek paralēli, un DH resursu gadījumā tas varbūt ir vienīgais iespējamais veids. Tai pašā laikā jāpieņem, ka sākotnēji izstrādāts kvalitatīvs plāns un ilgtermiņa stratēģija, kā arī visu izstrādes procesu dokumentēšana atvieglotu darbu.

Tomēr Latvijas DH ekosistēmā, tāpat kā zinātnē vispār, būtiskākais organizatoriskais izkārtojums ir projekts, savukārt finansējuma iegūšanas veids – konkursu kārtībā. Tas neveicina apstākļus ilgtermiņīgu DH resursu veidošanai, jo konkursos augstāk tiek vērtētas inovācijas, un ilgtermiņa “kapitālieguldījumi” jau pastāvošās struktūrās īsti neatbilst zinātniskās izcilības vai akadēmiskās konkurences vairošanai. Tāpēc DH resursu veidošana un papildināšana bieži notiek sporādiski, atkarībā no konkrētās institūcijas vai pētnieka veiksmēm vai neveiksmēm grantu konkursos.

Tas pats sakāms par darbu pie konkrētu DH resursu veidošanas. Vairāki procesa posmi, kuriem ideālajā gadījumā būtu jābūt secīgiem, notiek vienlaikus vai vispār tiek izlaisti, piemēram, mērķauditorijas izpēte vai prototipa testēšana. Resursu izstrādātāji apzinās, ka tas būtu jādara, bet dažādu iemeslu dēļ šie un citi mikroposmi vai uzdevumi tiek izlaisti vai īstenoti formāli, pavirši, nepietiekami. Allaž attaisnojums ir finansējuma un laika trūkums, arī tas, ka praktiski vienmēr resursa mērķauditorija ir pētnieki, tātad izstrādātāji paši ir sava mērķauditorija. Tomēr par būtisku faktoru var atzīmēt zināšanu trūkumu, kas daļēji saistīts gan ar salīdzinoši nelielo pieredzi DH laukā, gan šīs ekosistēmas (ieskaitot katras institūcijas iekšējās ekosistēmas un digitālās infrastruktūras) nepietiekošo institucionālo rutinizāciju.

Konceptuāli DH resursus un rīkus var dalīt divās lielās grupās pēc to mērķauditorijas: pētnieki un visi pārējie. Tam ir praktisks pamatojums, jo vajadzības, kā arī kompetences šīm divām grupām būtiski atšķiras, tomēr jāņem vērā, ka šīs plašās grupas iekšēji ir neviendabīgas. Visus pētījumā biežāk minētos Latvijas DH resursus, kuru mērķauditorija ir plašākā sabiedrība, savām akadēmiskajām vajadzībām izmanto arī pētnieki. Tas liek secināt, ka, izstrādājot un uzturot šādus resursus, jādomā dažādu funkciju, ne dažādu DH rīku vai resursu kategorijās. Tātad jāspēj izveidot tāds resurss, kas reizē ir gan vienkāršs un saprotams visdažādākajām sabiedrības grupām, gan spēj nodrošināt specifiskās funkcijas un vajadzības, kas nepieciešamas pētniekiem, piemēram, drošticams saturs, datu eksporta iespējas.

DH rīku un resursu veidošana ir ne vien DH laukā strādājošo pētnieku ikdienas darbs un mērķis, bet arī līdzeklis savu pētniecisko darbu uzdevumu veikšanai, tāpēc tas jāskata arī kā zinātnes digitālās infrastruktūras daļa. Vērtējot Latvijas DH rīkus un resursus no šobrīd ES zinātnes politikā dominējošās atvērtās zinātnes paradigmas viedokļa, stāvoklis vairākās jomās

ir neapmierinošs. Šajā pētījumā nav tikusi veikta pētāmo institūciju DH rīku un resursu revīzija, par pamatu ņemot konkrētus atvērtās zinātnes kritērijus, principus un prasības, bet par to liecina aptaujas un interviju atbildes – rūpes par metadatiem, dokumentāciju, datu importēšanas un eksportēšanas funkciju nodrošināšanu, koda atvēršana u. c. paliek otrajā plānā. Tas pamatā izriet no Latvijas zinātnes politikā vēl praktiski neiestrādātajām atvērtās zinātnes prasībām. Tai pašā laikā ir skaidrs, ka agrāk vai vēlāk tas notiks un būs saistoši arī jau izveidotiem DH rīkiem un resursiem, kā šauri specializētiem akadēmiskajiem, tā arī paredzētiem visplašākajai sabiedrībai. Domājams, atvērto datu FAIR modelis varētu kalpot par labu standartu, kam līdzināties, gan veidojot, gan pārveidojot Latvijas DH rīkus un resursus.

Pētot atsevišķu institūciju un indivīdu prakses, izstrādājot un uzturot DH RR, vienlaikus analizēta visu rīku un resursu ekosistēma, ko veido pētāmo institūciju RR. Raugoties uz šāda mēroga kopainu, var izteikt divus galvenos secinājumus. Pirmkārt, projektu finansējuma sistēma īpaši ietekmē par valsts finansējumu uzturētās pētnieciskās institūcijas. Projektu konkursos priekšroka tiek dota inovācijām un jaunu zināšanu (jaunu produktu vai to prototipu) radīšanai, nevis jau pastāvošu resursu ilglaicīgai attīstīšanai un uzturēšanai. Otrkārt, Latvijas DH lauks ir salīdzinoši jauns, tāpēc sadarbības struktūra un sistēma starp DH rīku un resursu veidotājiem un šo aktivitāšu finansētājiem (valsts zinātnes un kultūras pārvaldes institūcijām) tikai veidojas. Mūsu pētīto Latvijas DH RR ekosistēma, kas pamatā aprobežojas ar valsts subsidētajām nacionālā līmeņa pētnieciskajām un atmiņas institūcijām, joprojām lielā mērā funkcionē, balstoties uz pašu institūciju un to individuālo pētnieku iniciatīvu, vajadzībām, kompetenci un pieejamajiem resursiem, un daudz mazākā mērā to ietekmē valsts zinātnes politika un citas regulējošās sviras.

- Baklāne, Anda (2021). Eksperta intervija projekta "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" ietvaros. Intervēja Jānis Daugavietis. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6623689>
- Baldunčiks, Juris, Ilze Auziņa (eds.) (2012). *Latviešu valoda digitālajā vidē: datorlingvistika: informatīvi izglītojoša semināru cikla materiāli: rakstu krājums*. Rīga: Latviešu valodas aģentūra.
- Berry, David M. (2012). *Understanding digital humanities*. Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan.
- Bērziņš, Andris (1988). Bibliotekārs, informators, tekstu glabātājs... *Zinātne un Tehnika*, Nr. 5, 9. lpp.
- Bērziņš, Andris, Irēna Grauda (1987). Kompjuters informatora vietā. *Padomju Jaunatne*, 03.01., 4. lpp.
- Bite, Kitija, Daugavietis, Jānis, Kampars, Jānis, Kreicbergs, Jānis, Kuchma, Iryna, Ločmele, Elina, Ostrovska, Daina, Vecpuise, Elza, Veisa, Kristīne, Zelve, Margarita (2020). *Pētījums par atvērto zinātni un rīcībpolitikas ceļa kartes izstrādi*. Rīga: Latvijas Nacionālā bibliotēka.
- Blumberg, Matthew (2013). Foundations in Human Computation. Michelucci, Pietro (ed.). *Handbook of Human Computation*. New York: Springer, pp. 3–4.
- Blūms, Aleksandrs (2021). Eksperta intervija projekta "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" ietvaros. Intervēja Ilona Kunda. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624025>
- Bourdieu, Pierre (1986). The Forms of Capital. Richardson, John G. (ed.). *Handbook of theory and research for the sociology of education*. Westport: Greenwood Press, pp. 46–58.
- Boyd, Douglas A., Larson, Mary A. (eds.) (2014). *Oral History and Digital Humanities: Voice, Access, and Engagement*. New York: Palgrave Macmillan US.
- Briedis, Raimonds (1992). Neliela inventarizācija. *Literatūra un Māksla*, 30.10., 5. lpp.
- Brūvelis, Ainars (2021). Eksperta intervija projekta "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" ietvaros. Intervēja Jānis Daugavietis. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624059>
- Callison-Burch, Chris, Dredze, Mark (2010). Creating speech and language data with Amazon's Mechanical Turk., *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics, pp. 1–12.
- Causser, Tim, Terras, Melissa (2014). 'Many hands make light work. Many hands together make merry work': Transcribe Bentham and crowdsourcing manuscript collections. Ridge, Mia (ed.). *Crowdsourcing our cultural heritage*. Farnham: Ashgate, pp. 57–88.
- Collins, Sandra, Harrower, Natalie, Haug, Dag Trygve Truslew, Immenhauser, Beat, Lauer, Gerhard, Orlandi, Tito, Romary, Laurent, Wandl-Vogt, Eveline (2015). *Going digital: Creating change in the Humanities*.
- Dallas, Costis, Chatzidiakou, Nephelie, Benardou, Agiatis, Bender, Michael, Berra, Aurélien, Clivaz, Claire, Cunningham, John, Dabek, Meredith, Garrido, Patricia, Gonzalez-Blanco, Elena, Hadalin, Jurij, Hughes, Lorna, Immenhauser, Beat, Joly, Anne, Kelpšienė, Ingrida, Kozak, Michał, Kuzman, Koraljka, Lukin, Marko, Marinski, Irena, Maryl, Maciej, Owain, Robert, Papaki, Eliza, Schneider, Gerlinde, Scholger, Walter, Schreibman, Susan, Schubert, Zoe, Tasovac, Toma, Thaller, Manfred, Wcislik, Piotr, Werla, Marcin, Zebec, Tvrtko (2017). *European survey on scholarly practices and digital needs in the arts and humanities – Highlights Report*. Zenodo.
- Daugavietis, Jānis (2015). *Amatiermāksla Latvijā: kopienas attīstība un kultūrpolitika*. Npublicēts. Promocijas darbs doktora grāda iegūšanai socioloģijā, lietišķās socioloģijas apakšnozarē. Darba vadītājs Dr. soc. prof. Aivars Tabuns. Rīga: Latvijas Universitāte.
- Daugavietis, Jānis (2021). Motivation to engage in crowdsourcing: Towards the synthetic psychological–sociological model. *Digital Scholarship in the Humanities*, No. 36(4), pp. 858–870.
- Edmond, Jennifer, Morselli, Francesca (2020). Sustainability of digital humanities projects as a publication and documentation challenge. *Journal of Documentation*, No. 76(5), pp. 1019–1031.
- Freimanis, Jānis (2021). Eksperta intervija projekta "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" ietvaros. Intervēja Ance Kristāla. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624089>

- Gibbs, Fred, Owens, Trevor (2012). Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs. *Digital Humanities Quarterly*, No. 6(2), pp. 1–36.
- Given, Lisa M., Willson, Rebekah (2018). Information technology and the humanities scholar: Documenting digital research practices. *Journal of the Association for Information Science and Technology*, No. 69(6), pp. 807–819.
- Golub, Koraljka, Kamal, Ahmad M., Vekseliu, Johan (2021). Knowledge organisation for digital humanities: An introduction. Koraljka, Golub, Ying-Hsang, Liu (eds.). *Information and Knowledge Organisation in Digital Humanities*. Routledge, pp. 1–22.
- Green, Harriett E, Courtney, Angela (2015). Beyond the Scanned Image: A Needs Assessment of Scholarly Users of Digital Collections. College & Research Libraries. Available: doi:10.5860/crl.76.5.690 College & Research Libraries.
- Grūzītis, Normunds (2021). Eksperta intervija projekta “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” ietvaros. Intervēja Ance Kristāla. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624097>
- Hai-Jew, Shalin (ed.) (2017). *Data Analytics in Digital Humanities*. Multimedia Systems and Applications. Cham: Springer International Publishing.
- Hedges, Mark, Dunn, Stuart (2017). *Academic crowdsourcing in the humanities: Crowds, communities and co-production*. Oxford: Chandos Publishing.
- Klein, Julie T. (2018). *Interdisciplining Digital Humanities*. *Boundary Work in an Emerging Field*. University of Michigan Press.
- Klotiņš, Arnolds, Muktupāvels, Valdis (1985). Latviešu tautas mūzikas instrumenti un to lietošanas semantika K. Barona “Latvju dainās”. Darkevics, A., Mūrniece, L. (sast.). *Latviešu mūzika* '85, Nr. 17. Rīga: Liesma, 53.–82. lpp.
- Ķīrsis, Uldis (2021). Eksperta intervija projekta “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” ietvaros. Intervēja Jānis Daugavietis. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624130>
- Levenberg, Lewis, Neilson, Tai, Rheams, David (eds.) (2018). *Research Methods for the Digital Humanities*. Cham: Springer International Publishing.
- Liu, Alan (2004). *The Laws of Cool: Knowledge Work and the Culture of Information*. Chicago: University of Chicago Press.
- Martena, Sanita (2021). Eksperta intervija projekta “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” ietvaros. Intervēja Ilona Kunda. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624157>
- Maxwell, John W., Hanson, Erik, Desai, Leena, Tiampo, Carmen, O'Donnell, Kim, Ketheeswaran, Avvai, Sun, Melody, Walter, Emma, Michelle, Ellen (2019). Prospects. *Mind the Gap: A Landscape Analysis of Open Source Publishing Tools and Platforms*.
- Poole, Alex H. (2017a). “A greatly unexplored area”: Digital curation and innovation in digital humanities. *Journal of the Association for Information Science and Technology*, No. 68(7), pp. 1772–1781.
- Poole, Alex H. (2017b). The conceptual ecology of digital humanities. *Journal of Documentation*, No. 73(1), pp. 91–122.
- Prikulis, Juris (1982). Lai uzzina katrs. *Cīņa*, 02.05., 2. lpp.
- Pūtelis, Aldis (2021). Eksperta intervija projekta “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” ietvaros. Intervēja Jānis Daugavietis. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624170>
- Rapa, Sanda (2021). Eksperta intervija projekta “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” ietvaros. Intervēja Agnese Karlsona. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624194>
- Reinsone, Sanita (2020). Searching for Deeper Meanings in Cultural Heritage Crowdsourcing. Hetland, Per, Pierroux, Palmyre, Esborg, Line (eds.). *A History of Participation in Museums and Archives: Traversing Citizen Science and Citizen Humanities*. Routledge, pp. 186–207.
- Reinsone, Sanita (2021). Eksperta intervija projekta “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” ietvaros. Intervēja Jānis Daugavietis. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624222>
- Rožkalne, Anita (1992a). Latvijas institūts. Idejas projekcija. *Diena*, 07.05., 4. lpp.
- Rožkalne, Anita (1992b). Datu banka *Literāti* meklē. *Literatūra un Māksla*, 12.06., 2. lpp.

- Rožkalne, Anita (2021). Eksperta intervija projekta "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" ietvaros. Intervēja Jānis Daugavietis. Digitāls audiofails. Pieejams: <https://doi.org/10.5281/zenodo.6624257>
- Ryan, Richard M., Deci, Edward L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- Sabharwal, Arjun (2015). *Digital curation in the digital humanities: Preserving and promoting archival and special collections*. Chandos Publishing.
- Schmitz, Carsten (2018). *Limesurvey GmbH*. Hamburg, Germany: LimeSurvey GmbH.
- Schreibman, Susan, Hanlon, Ann M. (2010). Determining Value for Digital Humanities Tools: Report on a Survey of Tool Developers. *Digital Humanities Quarterly*, No. 4(2).
- Skadiņa, Inguna (2019). Some Highlights of Human Language Technology in Baltic Countries. Lupeikiene, A. et al. (eds.). *Databases and Information Systems X*, pp. 18–30.
- Skadiņa, Inguna (2021). Datorlingvistika Latvijā. *Nacionālā enciklopēdija*. Rīga: Latvijas Nacionālā bibliotēka.
- Skadiņa, Inguna, Spektors, Andrejs (2009). Latviešu valoda datorā – pētījumi, resursi, tehnoloģijas. Prezentācija zinātniskajā konferencē "Matemātika un informātika pēc 50..." Rīga, 09.11.
- Spektors, Andrejs (1999). Latviešu valodas datorlingvistikas resursi. *Baltu filoloģija VIII*, 53–59. lpp.
- Spektors, Andrejs (2022). Eksperta intervija projekta "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" ietvaros. Intervēja Jānis Daugavietis. Digitāls audiofails (divās daļās). Pieejams: <https://doi.org/10.5281/zenodo.6624716>
- Steiner, Christina, Agosti, Maristella, Sweetnam, Mark, Hillemann, Eva-C, Orio, Nicola, Ponchia, Chiara, Hampson, Cormac, Munnely, Gary, Nussbaumer, Alexander, Albert, Dietrich, Conlan, Owen (2014). Evaluating a digital humanities research environment: the CULTURA approach. *International Journal on Digital Libraries*, No. 15(1), pp. 53–70.
- Thoden, Klaus, Stiller, Juliane, Bulatovic, Natasa, Meiners, Hanna-Lena, Boukhelifa, Nadia (2017). User-Centered Design Practices in Digital Humanities – Experiences from DARIAH and CENDARI. *ABI Technik*, No. 37(1), pp. 2–11.
- Vikis-Freibergs, Vaira, Freibergs, Imants (1978). Formulaic analysis of the computer-accessible corpus of Latvian sun-songs. *Computers and the Humanities*, No. 12(4), pp. 329–339.
- Walsh, David, Clough, Paul, Foster, Jonathan (2016). User categories for digital cultural heritage. *First International Workshop on Accessing Cultural Heritage at Scale*, pp. 3–9.
- Warwick, Claire (2012). Studying users in digital humanities. *Digital Humanities in Practice*, No. 17(5), pp. 1–21.
- Warwick, Claire, Terras, Melissa, Galina, Isabel, Huntington, Paul, Pappa, Nikoleta (2007). Evaluating digital humanities resources: the LAIRAH project checklist and the Internet Shakespeare Editions project. *Openness in digital publishing: awareness, discovery, and access: proceedings of the 11th International Conference on Electronic Publishing*. Vienna, June 13–15, 2007: ELPUB, pp. 297–306.
- Wilkinson, Mark D., Michel, Dumontier, Aalbersberg, IJsbrand Jan, Appleton, Gabrielle, Axton, Myles, Baak, Arie, Blomberg, Niklas, Boiten, Jan-Willem, Santos, Luiz Bonino da Silva, Bourne, Philip E., Bouwman, Jildau, Brookes, Anthony J., Clark, Tim, Crosas, Mercè, Dillo, Ingrid, Dumon, Olivier, Edmunds, Scott, Evelo, Chris T., Finkers, Richard, Gonzalez-Beltran, Alejandra, Gray, Alasdair JG, Groth, Paul, Goble, Carole, Grethe, Jeffrey S, Heringa, Jaap, Hoen, Peter AC 't, Hooft, Rob, Kuhn, Tobias, Kok, Ruben, Kok, Joost, Lusher, Scott J., Martone, Maryann E., Mons, Albert, Packer, Abel L., Persson, Bengt, Rocca-Serra, Philippe, Roos, Marco, Schaik, Rene van, Sansone, Susanna-Assunta, Schultes, Erik, Sengstag, Thierry, Slater, Ted, Strawn, George, Swertz, Morris A., Thompson, Mark, Lei, Johan van der, Mulligen, Erik van, Velterop, Jan, Waagmeester, Andra, Wittenburg, Peter, Wolstencroft, Katherine, Zhao, Jun, Mons, Barend (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, No. 3(1), pp. 1–9.
- Woodcock, J. C., Greenhill, A., Holmes, K., Graham, G., Cox, J., Oh, E. Y., Masters, K. (2017). Crowdsourcing citizen science: exploring the tensions between paid professionals and users. *Journal of Peer Production*, No. 10, pp. 1–10.
- Zorich, Diane (2008). *A survey of digital humanities centers in the United States*. Council on Library and Information Resources Washington, DC.

Pielikums: intervētie eksperti.

Vārds	Uzvārds	Institūcija	Amats
Anda	Baklāne	LNB	LNB digitālās pētniecības vadītāja, <i>C. Phil.</i>
Aleksandrs	Blūms	IZM	IZM zinātnes nozares padomnieks, <i>Mg. soc.</i>
Ainars	Brūvelis	Lursoft, Timenote, iepriekš – LU LFMI	IT projektu vadītājs, <i>Mg. phys.</i>
Jānis	Freimanis	LNB	Tehnoloģiju departamenta direktors, <i>Mg. sc. comp.</i>
Normunds	Grūzītis	LU MII AiLab	Laboratorijas vadītājs, <i>Dr. sc. comp.</i>
Uldis	Ķirsis	LU LFMI	Programmētājs, IT projektu vadītājs
Sanita	Martena	RTA	Profesore, <i>Dr. philol.</i>
Aldis	Pūtelis	iepriekš – LU LFMI	Pētnieks, <i>Mg. philol.</i>
Sanda	Rapa	LU LaVI	Institūta direktore, <i>Dr. philol.</i>
Sanita	Reinsone	LU LFMI	Vadošā pētniece, <i>Dr. philol.</i>
Anita	Rožkalne	LU LFMI	Vadošā pētniece, <i>Dr. philol.</i>
Inguna	Skadiņa	LU MII AiLab	Vadošā pētniece, <i>Dr. sc. comp.</i>
Andrejs	Spektors	LU MII AiLab	Vadošais pētnieks, <i>Dr. phys.</i>

Tools and Resource Development Practices in Latvian Digital Humanities

Jānis Daugavietis, Agnese Karlsonē,
Ilona Kunda, Ance Kristāla

Keywords: Latvian DH field, DH tools and resources ecosystem,
user experience research, interdisciplinarity, open science

The study was developed within the framework of a state-commissioned research project, the overarching goal of which is the development of digital humanities resources and tools (DH RR), as well as facilitation of the entire DH field in Latvia, which is a practically under-explored topic. As the object of research, we have chosen the practices of tool and resource developers, comparing them with the prevalent findings in theoretical literature and requirements of the current science policy. We use two main methods to obtain data – online survey of DH RR developers and maintainers, as well as expert interviews. Respondents were asked not only about their practices in developing and maintaining DH RR, but also surveyed to find out their ideas about what the ideal situation would be. The results of the study are in line with the conclusions made in the foreign academic literature on the main problems in the DH ecosystem. Firstly, it is a project funding system that has a particular impact on publicly funded research institutions. In project competitions, priority is given to innovation and the creation of new knowledge (new products or prototypes) over the long-term development and maintenance of existing resources. Secondly, the field of DH in Latvia is relatively new, so the structure and system of cooperation between the creators of DH tools and resources and the financiers of these activities (state scientific and cultural administration institutions) are only being formed. To a large extent, Latvian DH RR ecosystem, which in our research is mainly limited to state-subsidized national research and memory institutions, still functions on the basis of the initiative, needs, competence and available resources of the institutions themselves and their individual researchers.

Latviešu folkloras krātuves digitālais arhīvs *garamantas.lv*: priekšvēsture un attīstība

Sanita Reinsone, Sandis Laime

Raksts sagatavots Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi” projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (VPP-IZM-DH-2020/1-0001).

Ievads

Latviešu folkloras krātuves (turpmāk – LFK) mērķis kopš dibināšanas 1924. gadā līdz mūsdienām ir dažādos formātos dokumentēt Latvijas iedzīvotāju, tostarp mazākumtautību, nemateriālo kultūras mantojumu, veikt tā izpēti un publicēt gan pētījumu rezultātus, gan zinātniski strukturētas pētniecisko avotu izlases. Šobrīd LFK ir institucionāli piederīga LU Literatūras, folkloras un mākslas institūtam (turpmāk – LFMI), tā aktīvi sadarbojas ar institūcijām, kas arhivē tradicionālo kultūru ārpus Latvijas, apmainoties ar pieredzi un īstenojot kopīgus projektus, un darbojas Ziemeļvalstu un Baltijas valstu folkloras arhīvu tīklā un folkloras arhīvu darba grupā Starptautiskās Folkloras un etnoloģijas biedrības (SIEF, *Société Internationale d'Ethnologie et de Folklore*) paspārnē.

Apmēram gadsimta laikā tapušais LFK arhīvs ir viena no plašākajām NKM dokumentu krātuvēm Eiropā, kurā glabājas vairāk nekā 3 000 000 folkloras vienību. Kopš 2014. gada LFK digitālā arhīva *garamantas.lv* ilgtermiņa mērķis ir nodrošināt brīvu piekļuvi apjomīgajam LFK krājumam, kā arī iesaistīt sabiedrību jaunu materiālu iesniegšanā un apstrādē. Šī raksta mērķis ir sniegt pārskatu par LFK digitālā arhīva un citu digitālo resursu tapšanas vēsturi, datu digitalizēšanas un apstrādes procesu, kā arī par digitālā arhīva struktūru un saturu.

Folkloras materiālu digitalizācijas vēsture

Latviešu humanitāro zinātņu materiālu digitalizācijas agrākais posms aizsākās tieši ar folkloras materiālu datorizēšanas centieniem, kas norisējuši jau kopš 20. gadsimta 60. gadu beigām. 1976. gadā Imants Freibergs un Vaira Viķe-Freiberga izveidoja pirmo tautasdziesmu tekstu korpusu ar nosaukumu “Monreālas latviešu saules dziesmu korpus”, kas ietver vairāk nekā 4000 tautasdziesmu tekstu un kura veidošana aizsākās jau kopš 1967. gada (Vikis-Freibergs & Freibergs 1978a, 1978b). Tas 80. gadu beigās papildināts.

1982. gadā tapa pirmā plašākā latviešu tautasdziesmu datubāze, kurai par pamatu ņemti Arveda Švābes, Kārļa Strauberga un Edītes Hauzenbergas-Šturmas sastādītie 12 “Latviešu tautas dziesmu” sējumi (Švābe et al. 1952–1956). Šo projektu Monreālā un Bostonā īstenoja Valdis Bērziņš, Imants Freibergs, Kristīne Konrāde, Guntis Strazds un Vaira Viķe-Freiberga. Datubāzē iekļauts vairāk nekā

71 000 tautasdziesmu tekstu (Viķis-Freibergs 1999). Uz šo būtisko iestrāžu pamata vēlāk tapa Krišjāņa Barona Dainu skapja digitalizācijas projekts.

Iedvesmojoties no trimdas latviešu veikuma, tautasdziesmu tekstu digitālā korpusa izveidē iesaistījās arī Latvijas pētnieki. Sadarbojoties LU LFMI Latviešu folkloras krātuvei un LVU Fizikas un matemātikas fakultātes Datorikas nodaļai, starp 1986. un 1994. gadu tika realizēts tā saucamais “Rīgas dainu projekts”. Tā laikā tika datorizēti Krišjāņa Barona sastādītajās “Latvju dainās” iekļautie 217 996 tautasdziesmu teksti, kā arī LFK sagatavotais “Latvju dainu substantīvu rādītājs”, kurā iekļauti ~120 000 dažādu vārdformu un vārdlietojumu. No LVU Fizikas un matemātikas fakultātes puses projektā piedalījās Harijs Bondars, A. Niedrītis, I. Rāta un Ā. Sproģe, savukārt LFK darba grupā iesaistījās Kārlis Arājs, Silvija Kokina, Aldis Pūtelis un Beatrise Reidzāne. Tekstu pārrakstīšanu datorlasāmā formātā veica LVU Fizikas un matemātikas fakultātes studenti, savukārt datortehnikas iegāde LFK tika veikta ar Sorosa fonda atbalstu. 1994. gadā tika plānots uzsākt arī LFK sistematiskā rādītāja datorizēšanu (Kokina 1993). Rīgas dainu projekta noslēgumam tika veltīta 1993. gada Krišjāņa Barona konference “Man bij dziesmu vēcelīte: no rokrakstiem līdz skaitļotājiem”¹.

20. gadsimta 80. gadu beigās LU Matemātikas un informātikas institūta (turpmāk – MII) Mākslīgā intelekta laboratorijā (turpmāk – AiLab) tika aizsākta fundamentālu folkloras publikāciju digitalizācija, kā rezultātā 1997. gadā tiešsaistē tika publicēti Pētera Šmita apkopotie “Latviešu tautas ticējumi” 4 sējumos², kas ir līdz šim lielākais latviešu ticējumu publicējums. 1998. gadā tiešsaistē tika publicēti Pētera Šmita sastādīto “Latviešu pasaku un teiku” 15 sējumi (Šmits 1925–1937), kas ir lielākais latviešu vēstītājfolkloras publicējums. Teksti publicēti *html* formātā kopā ar sastādītāja ievadu un komentāriem³. Šos projektus atbalstīja Sorosa fonds – Latvija.

2001. gadā Maruta Lietiņa-Reja (*Lietiņa-Ray*) XML formātā sagatavoja un Virdžīnijas Universitātes tīmekļvietnē padarīja pieejamas “Latviešu tautasdziesmas” jeb t. s. Kopenhāgenas dainas⁴ (Švābe et al. 1952–1956).

2002. gadā Latviešu folkloras krātuve Alda Pūteļa vadībā, sadarbojoties ar Hariju Bondaru (Latvijas Universitāte), Ainaru Brūveli (SIA “Lursoft”) un Imantu Freibergu, ar Latviešu fonda atbalstu tiešsaistē publicēja Krišjāņa Barona Dainu skapja digitālo versiju *dainuska-pis.lv*, kuras izveide tika uzsākta jau 1994. gadā⁵. Dublējot Krišjāņa Barona sastādīto “Latvju dainu” saturu, vietnē iekļauti 119 000 tautasdziesmu tekstu. Tautasdziesmu teksti sastatīti ar ieskenētiem dziesmu oriģinālpierakstiem, kas glabājas Dainu skapī, tiem pievienoti arī metadati – dziesmu pierakstītāji un pierakstīšanas vietas. Šis aizvien ir viens no

1 <http://lfk.lv/krisjana-barona-konference/1993#apraksts> [skatīts 12.06.2022.]

2 <http://valoda.ailab.lv/folkloraticejumi> [skatīts 29.04.2022.]

3 <http://valoda.ailab.lv/folklorapasakas> [skatīts 29.04.2022.]

4 <https://latviandainas.lib.virginia.edu> [skatīts 29.04.2022.]

5 Par atvērto Dainu skapja elektronisko versiju. *Latvijas Vēstnesis*, 2002. 07.11., 14. lpp.

visatpazīstamākajiem humanitāro zinātņu digitālajiem resursiem Latvijā, kaut arī kopš 2005. gada nav mainīta nedz tā saskarne, nedz funkcionalitāte.

2003. gadā Latviešu folkloras krātuve Alda Pūteļa vadībā ar UNESCO atbalstu aizsāka fotoattēlu, zīmējumu un audio ierakstu digitalizāciju. Turpmākajos gados sadarbībā ar MII Mākslīgā intelekta laboratoriju tika izveidota datubāze “Krātuve”, kurā pakāpeniski tika ievadīts LFK sistemātiskais rādītājs un kas darbojās lokālajā tīklā. Paralēli tapa ilustrāciju (fotoattēlu un zīmējumu) un audio un video ierakstu rādītāju datubāzes, kas ar Valsts kultūrkapitāla fonda atbalstu tika padarītas publiski pieejamas LFK tīmekļvietnes sadaļā “Materiālu interneta lasītava” kopš 2006. gada. 2003. gadā tika aizsākts arī projekts “Latviešu pasaku un teiku paralēlais korpuss latviešu un vācu valodā”, kur LFK (Guntis Pakalns), LU MII AiLab (Sanita Reinsone, Ilmārs Poikāns) un Getingenes Zinātņu akadēmijas “Pasaku enciklopēdijas” arhīva sadarbības rezultātā ar Baltijas–Vācijas Augstskolu biroja atbalstu tika digitalizēts trīs Pēterā Šmita “Latviešu pasaku un teiku” teiku sējumu tulkojums vācu valodā un *Wiki* vidē izveidots paralēli sastatīts tekstu korpuss latviešu un vācu valodā⁶ (Pakalns 2005).

2003. gadā ar Valsts kultūrkapitāla atbalstu tika izveidots “Latviešu sakāmvārdu datorfonds” – tiešsaistē *html* formātā publicēta LFK sakāmvārdu kartotēka⁷. Šo projektu īstenoja Gunita Ģeģere (LFK) un Sanita Reinsone (AiLab). Šiem digitālajiem resursiem ir būtiska nozīme LFK digitālā arhīva izveidošanā un arī attīstībā. Pirmkārt, to izmantošana pētniecībā un vispārējā izziņā bija liecinājums tam, ka pētniekiem un plašākai sabiedrībai digitālie folkloras resursi ir noderīgi un tie tiek plaši izmantoti. Otrkārt, šo resursu veidotāji ir guvuši nozīmīgu pieredzi digitālo resursu veidošanā, kas vairākos gadījumos tika tālāk izmantota LFK digitālā arhīva izstrādē. Treškārt, liela daļa no šiem resursiem veidojuši *garamantas.lv* sākotnējo pamatu vai arī tikuši integrēti *garamantas.lv* struktūrā vēlāk.

LFK digitālā arhīva *garamantas.lv* izveidošana

Pēdējās desmitgadēs norisošā vispārējā digitalizācija kultūras mantojuma jomā un tīmekļa tehnoloģiju straujā attīstība ir būtiski ietekmējusi kultūrvēsturisko kolekciju pieejamību un lietojuma funkcionalitāti. LFK digitālā arhīva pamata koncepcija saistās ar t. s. līdzdalības pieeju (*participatory approach*) digitālo kolekciju veidošanā, kur resursa lietotāji netiek uztverti par pasīviem materiālu skatītājiem un informācijas uztvērējiem, bet gan par dalībniekiem resursa veidošanā un sabiedrotajiem zināšanu radīšanā. Diskusijas par attieksmes maiņu pret arhīvu u. c. krājumu glabātāju lietotājiem pētnieciskajā literatūrā aizsākušās jau 20. gadsimta 90. gados (Ketelaar 1992; Bruebach 2003), tomēr būtiskākās izmaiņas arhīvu praksē, tostarp digitālo arhīvu veidošanā, ienāk pēc *Web2.0* tehnoloģiju attīstīšanās 21. gadsimta pirmās desmitgades vidū (Huvila 2008).

6 <http://pasakas.lfk.lv> [skatīts 29.04.2022.]

7 <http://valoda.ailab.lv/folkloras/sakamvardi> [skatīts 29.04.2022.]

Latviešu folkloras krātuves digitālā arhīva sākums un turpmākā attīstība iekļaujas šajā jaunajā vilnī, kur tehnoloģiju jaunievedumi tiek izmantoti, lai atklātu krājumu un satiktos ar tā lietotāju konceptuāli jaunā veidā.

Darbs pie Latviešu folkloras krātuves digitālā arhīva izveidošanas aizsākās 2013. gada sākumā ar vēlmi izveidot digitālu resursu, kur ikviens varētu brīvi piekļūt Latviešu folkloras krātuvē glabātajiem materiāliem. Digitālā arhīva izveidošanas pamatgrupa bija vizionārs, mecenāts un IT speciālists Ainars Brūvelis, kuram bijusi liela loma Krišjāņa Barona Dainu skapja digitālās versijas izveidošanā, programmētājs Uldis Ķīrsis un LFK pētniece Sanita Reinsone. Šai nelielajai grupai pēc nepieciešamības un atbilstoši kompetencei pievienojās arī citi Latviešu folkloras krātuves darbinieki. *Garamantas.lv* vadītāji: Sanita Reinsone (2014–2017, 2019–2020), Kristīne Pokratniece (2018) un Sandis Laime (kopš 2021).

Par LFK digitālā arhīva sākotnējo nosaukumu tika izmantots *folklore.lv*, kas digitālā arhīva pirmajos pastāvēšanas gados funkcionēja kā paralēlais nosaukums starptautiskajā komunikācijā un darbojas kā *alias* domēns. Nosaukums *garamantas.lv* radās īsi pirms arhīva pirmās publiskās prezentācijas 2014. gada decembrī Latviešu folkloras krātuves 90 gadu jubilejā. Kaut arī eventuālo nosaukumu saraksts bija garāks, *garamantas.lv* tika izvēlēts tāpēc, ka vārdā nav garumzīmju, kam ir nozīme domēnvārda labskanībā, un *garamantu* nozīme ir folkloras – šī digitālā resursa centrālā tematiskā fokusa – nosaukuma metaforisks līdzinieks kopš 19. gadsimta vidus (Bula 2018).

Garamantas.lv saskarne ir izveidota astoņās valodās (latviešu, latgaliešu, angļu, baltkrievu, igauņu, krievu, lietuviešu, vācu), taču dinamiska sākumlapa, kuras saturs tiek papildināts, ir tikai latviešu un angļu valodā. Digitālajā arhīvā integrētais metadatu tulkošanas modulis un lietotāju tiesību regulējums, kas ļauj piešķirt tiesības metadatu tulkošanai jebkuram lietotājam, atvieglo sistēmas tulkošanas darbu.

LFK digitālā arhīva *garamantas.lv* attīstībā var iezīmēt trīs posmus.

1. Sistēmas izveidošana un agrāk radīto datu un datubāžu integrācija.

Pirmais *garamantas.lv* attīstības posms datējams ar 2014. gada otro pusi līdz 2016. gada beigām, kad tika izveidota pamata struktūra un integrēti iepriekš digitalizētie arhīva materiāli un to metadati. No LFK fotogrāfiju un ierakstu datubāzes jaunajā sistēmā tika importētas visas digitalizētās fotogrāfijas un zīmējumi, kā arī audio ieraksti. Tika importēta arī datubāze “Krātuve”, kurā bija ievadīta daļa no LFK sistēmātiskā rādītāja. Importēto datu sakārtošana un tīrīšana turpmākos divus gadus prasīja lielāko daļu redakcionālā darba, jo importēto vienību skaits bija salīdzinoši liels, taču dati nebija pietiekami labi sistematizēti, bija daudz drukas kļūdu un arī dublikātu. Problemātiskākās datubāzes daļas bija saistītas ar personām

un vietām. Piemēram, personu datubāzē bija iekļauts personas vārds, uzvārds un vecums (atbilstoši tam, kā norādīts LFK manuskriptos), taču cita informācija lielākoties pieejama nebija. Turklāt informācija par vienu un to pašu personu varēja būt iekļauta gan attēlu datubāzē, gan arī audio ierakstu datubāzē, kas iepriekš funkcionēja atsevišķi, un personu vārdi tika ievadīti brīva teksta laukā, kas pieļāva dažādas variācijas. Lai sameklētu personu dublikātus un apvienotu tām piesaistītos datus, tika izstrādāts īpašs rīks, kas palīdz *garamantas.lv* redaktoriem salīdzināt divos personu profilos ievadīto informāciju un, ja nepieciešams, to apvienot, izvēloties, kurš no šiem profiliem tiek dzēsts, bet kurš paliek datubāzē.

Savukārt vietas iepriekš digitalizētajos LFK resursos nebija sistematizētas un piesaistītas koordinātēm. Dažādās datubāzēs un pat vienas datubāzes daļās vienai un tai pašai vietai varēja būt norādīti dažādi administratīvie iedalījumi (atkarībā no tā, kurā laikā folkloras fiksējums veikts), nesistemātiski tika lietoti vietu nosaukumi, reizēm pierakstot senāku nosaukuma variantu, citreiz – mūsdienās eksistējošo nosaukumu, dažkārt norādot tikai apriņķi vai aprakstošu nosaukumu. Tā, protams, ir situācija, kāda sastopama folkloras manuskriptos, taču vietu datubāzes vienādošana un sistematizēšana bija nepieciešama, lai folkloras materiālos fiksētās vietas varētu tikt attēlotas uz kartes sistemātiskā veidā. Šobrīd izstrādāts rīks, ar kuru sākotnējā sistēmā izmantotās vietas tiek pārkartētas saskaņā ar šobrīd digitālajā arhīvā pieņemto kartēšanas kārtību.

2015. gada sākumā *garamantas.lv* sistēmā tika importēts arī “Sakāmvārdu datorfonds” jeb LFK sakāmvārdu kartotēka, kas bija digitalizēta un pieejama tīmeklī *html* formātā.

Šī attīstības posma noslēgumā – 2016. gadā – tika īstenotas divas sadarbības iniciatīvas, kas sniedza LFK digitālajam arhīvam lielāku atpazīstamību, būtiski palielināja apmeklētāju skaitu un arī deva lielu ieguldījumu brīvprātīgo kopienas izveidošanā. 2016. gada pavasarī sadarbībā ar UNESCO Latvijas nacionālo komisiju tika izstrādāta manuskriptu atšifrēšanas platforma “Valodas talka”⁸, kas bāzējās *garamantas.lv* sistēmā un bija sinhronizēta ar to. Šī jaunā platforma bija paredzēta tikai LFK manuskriptu atšifrēšanai. Iniciatīva ilga 71 dienu. Lielākā daļa no iesaistītajiem bija skolēni, kopā ap 1500 dalībnieku (Reinsone 2018). 2016. gada maija beigās, sadarbojoties ar Latvijas simtgades biroju, uz izstrādātās tehniskās bāzes tika izveidota līdzīga manuskriptu atšifrēšanas platforma ar nosaukumu “Simgades burtnieki”. Arī tā bija sasaistīta un sinhronizēta ar *garamantas.lv* platformu, taču uzreiz veidota kā ilgtermiņa resurss un kā tāds arī turpina darboties (Reinsone 2020). Tā kā “Valodas talka” ilga vien dažus mēnešus un izstrādātās platformas uzturēšana prasīja ieguldījumus, šī lapa vairākus gadus pēc iniciatīvas noslēguma tika slēgta, pāradresējot domēnu *talka.garamantas.lv* uz “Simgades burtnieku” platformu.

2015. gadā *garamantas.lv* sistēmā tika integrēti arī biogrāfiskie un bibliogrāfiskie literatūras dati no LFMI Anitas Rožkalnes vadībā tapušās literatūras datubāzes un izveidots saistītais resurss *eliterature.lv*, kas kopš 2017. gada darbojas ar nosaukumu “Latvijas un latviešu literatūras vortāls” domēnā *literatura.lv* (Eglāja-Kristsone 2021). Ar šo paplašināšanos tika likts pamats vienotajam redakcionālajam moduļim visiem LFMI digitālajiem resursiem, kas top uz *garamantas.lv* bāzes.

8 <https://web.archive.org/web/20160426120256/http://talka.garamantas.lv/>

2. Līdzdalības moduļu un kartēšanas rīku attīstīšana.

Kaut arī tādas līdzdalības funkcijas kā manuskriptu atšifrēšana, atslēgvārdu pievienošana, komentēšana un tekstu tulkošana bija iestrādāta *garamantas.lv* sistēmā jau kopš paša tās izveides sākuma, šajā attīstības posmā, kuru var datēt no 2017. līdz 2020. gadam, piesaistot projektu finansējumu⁹, tika izstrādāti papildu digitālās līdzdalības moduļi, kā arī attīstīta viena ģeogrāfisko datu sistēma¹⁰, kas turpmāk nodrošina sistemātisku un korektu ģeogrāfisko datu attēlošanu uz kartes un arī atvieglo tālākizmantošanu.

Šajā attīstības posmā *garamantas.lv* digitālajā platformā tika izstrādāts (1) radošās līdzdalības modulis, kura funkcionalitāte nodrošina balsis ierakstīšanu un saglabāšanu digitālajā platformā īpaši izveidotā kolekcijā. Šis modulis vispirms tika izmēģināts, organizējot dzejas lasīšanas publiskās akcijas¹¹ *literatura.lv* (Eglāja-Kristsone, Raudive 2020) un LFK skaņu arhīva popularizēšanas iniciatīvu “Dziedi ar arhīvu”¹². (2) LFK aptauju modulis ar nosaukumu “LFK jautā”¹³, kurš turpmāk tiek attīstīts kā galvenais LFK digitālo materiālu vākšanas instruments. Aptaujās iegūtie dati tiek automātiski integrēti atbilstošā LFK digitālā arhīva kolekcijā. ERAF projekta laikā uz “LFK jautā” tehniskās bāzes tika izstrādāts arī resurss “Mūsdienu kalendārs”¹⁴, kura mērķis bija uzkrāt informāciju par sabiedrības svinēšanas paradumiem, taču pēc ERAF projekta beigām “Mūsdienu kalendārs” izstrāde apstājās.

Lai pievērstu sabiedrības uzmanību līdzdarbošanās iespējām LFMI digitālajos resursos un jo sevišķi LFK digitālajā arhīvā *garamantas.lv*, ERAF projektā tika izveidota jauna sabiedriskās zinātnes platforma *iesaisties.lv*¹⁵, kas darbojas kā LFMI digitālās līdzdalības iniciatīvu informācijas punkts un apkopo informāciju par jau minētajām un citām sabiedrības līdzdalības iniciatīvām.

Šajā posmā *garamantas.lv* tika izstrādāta konceptuāli jauna vietu datubāze, būtiski uzlaboti kartēšanas rīki un karšu vizualizācijas iespējas, kā arī tika izstrādāts tekstu anotēšanas rīks, kas sniedz iespēju veidot ne tikai metadatos balstītās kartes, kā tas bijis līdz šim, bet arī tādas kartes, kurās tiek attēlotas tekstos minētās vietas, un ir iespējams arī veidot saiknes starp dažādām datubāzes entitītijām.

9 “Stiprinot zināšanu sabiedrību: starpdisciplināras pieejas sabiedrības iesaistei digitālā kultūras mantojuma radīšanā” (Nr. 1.1.1.1/16/A/040, 2017–2019, ERAF, vad. Sanita Reinsone).

10 Sadarbībā ar pēcdoktorantūras projektu “Latviešu tautas naratīvu izpēte: ģeotelpiskās informācijas analīzes rīka un tiešsaistes teikumu motīvu un tipu rādītāja izstrāde” (Nr. 1.1.1.2/VIAA/1/16/193, 2017–2020, līdzfinansē ERAF, vad. S. Laime).

11 “Skandē Veidenbaumu” <https://lasi.literatura.lv> (2017), “Lasīsim dzejiņas” <https://berni.literatura.lv> (2018)

12 “Dziedi ar arhīvu” <https://dziedi.garamantas.lv> (2019)

13 <https://jauta.garamantas.lv> (2018)

14 <https://kalendars.garamantas.lv> (2019)

15 <https://iesaisties.lv> (2019)

3. Funkcionalitātes paplašināšana un starpinstitucionāla sadarbība.

Šī posma sākums datējams ar 2020. gada otro pusi. Tas ir laiks, kad aizvien ciešāk tiek īstenota sadarbība ar citām humanitāro zinātņu institūcijām, jo sevišķi tām, kas veido digitālos humanitāro zinātņu resursus. Sadarbības intensificēšanās gan norisējusi arī iepriekš, īstenojot starpinstitūciju iniciatīvu *digitalhumanities.lv* ar mērķi popularizēt digitālās humanitārās zinātnes Latvijā, tomēr 2020. gadā uzsāktais Valsts pētījumu programmas projekts “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (DH VPP), kura konsorcijs darbojas 8 partneri¹⁶, top par nākamo pagrieziena punktu digitālā arhīva *garamantas.lv* un LFMI digitālo resursu attīstībā kopumā.

Projekta laikā pilnībā digitalizēts un LFK digitālajā arhīvā *garamantas.lv* integrēts Liepājas Universitātes folkloras materiālu krājums, kas ir būtisks Kurzemes folkloras materiālu avots. Sadarbība nodibināta arī ar Rēzeknes Tehnoloģiju akadēmiju ar mērķi *garamantas.lv* iekļaut studentu folkloras vākumus. Šāda sadarbība jau vairākus gadus ir īstenota ar Latvijas Kultūras akadēmiju, kuras studentu savāktie lauka pētījumu materiāli arī tiek nodoti glabāšanai LFK un iekļauti digitālajā arhīvā.

DH VPP projekts dod iespēju uzlabot *garamantas.lv* administratīvo funkcionalitāti saistībā ar digitālā arhīva lietotāju lomu un tiesību definēšanu, sistēmas efektīvizēšanu un ātrdarbības uzlabošanu, no iepriekšējiem resursiem importēto datu pielāgošanu, rediģēšanu, tīrīšanu utt. Tomēr būtiskākais solis ir jaunas platformas izstrāde ar pagaidu nosaukumu *dati.lulfmi.lv*¹⁷, kur pirmo reizi publiski vienkopus ir pieejami visi LFMI digitālajos resursos esošie dati. Šis resurss ir pamatā orientēts uz pētnieku un studentu vajadzībām, tajā tiek izstrādāta detalizētas meklēšanas un datu atlasē funkcionalitāte, datu kopu, tostarp tekstu korpusu, lejupielādes iespējas. Arī šī datu platforma ir sinhronizēta ar *garamantas.lv*, kā arī ar *literatura.lv* un pārējiem apakšprojektiem.

LFK arhīva digitalizēšana

LFK digitālā arhīva *garamantas.lv* saturs top vairākos digitalizēšanas un digitalizētā materiāla apstrādes posmos. Katrs no tiem paredz specifisku darbību veikšanu un dažādas iesaistīto personu grupas, darbību realizēšanai izmantojot to realizēšanai digitālā arhīva administrēšanas un publiskajā vidē ieviestos specifiskos rīkus.

16 Projektu NR. VPP-IZM-DH-2020/1-0001 koordinē LU Literatūras, folkloras un mākslas institūts. Partneri: LU Matemātikas un informātikas institūts, LU Lībiešu institūts un Latviešu valodas institūts, Rēzeknes Tehnoloģiju akadēmija, Liepājas Universitāte, Latvijas Kultūras akadēmija, Rīgas Tehniskā universitāte, Latvijas Nacionālā bibliotēka. Projekta vadītāja: Sanita Reinsone. Projekta īstenošanas laiks: 05.10.2020.–04.10.2022.

17 <https://dati.lulfmi.lv> (2022)

1. *Analogā arhīva struktūras dublēšana digitālajā vidē*: šis darba posms attiecībā uz vēsturisko arhīva daļu ir īstenots, izveidojot digitālā arhīva pamata infrastruktūru (LFMI programmētāji) un dublējot sistēmā analogā arhīva struktūru – krājumus un kolekcijas (LFK digitālā arhīva redaktori), tomēr sistēmas optimizēšana un jaunu krājumu un kolekciju izveide turpinās joprojām.
2. *Materiālu digitālo kopiju sagatavošana*: manuskriptu un fotogrāfiju skenēšana (LFK digitālā arhīva skenētāji), audio un video ierakstu digitalizēšana (LFK digitālā arhīva redaktori un ārpakalpojumā piesaistīti nozares profesionāļi). Digitalizētais materiāls tiek publicēts digitālajā arhīvā, dodot iespēju brīvpieejā veikt materiāla izpēti, kā arī iesaistīt materiālu transkribēšanā brīvprātīgo kopienu. Šobrīd digitalizēta un publicēta lielākā daļa LFK manuskriptu arhīva, kā arī liela daļa fotogrāfiju, audio un video ierakstu arhīva. LFK vidēja termiņa mērķis ir līdz 2025./2026. gadam pabeigt LFK manuskriptu krājuma skenēšanas un publicēšanas etapu.
3. *Datorlasāma teksta ieguve*: manuskriptu tekstu transkribēšana, audio un video ierakstu tekstuāla atšifrēšana. Manuskriptu tekstu atšifrēšanai digitālajā arhīvā izveidots īpašs transkribēšanas rīks, ar kuru tekstu pārrakstīšanu datorlasāmā formātā var veikt ne tikai LFK darbinieki, bet jebkurš sistēmā reģistrēts lietotājs tam paredzētajās publicējām vietnēs (<http://garamantas.lv>, <http://lv100.garamantas.lv>). Pateicoties veiksmīgām pūles saistes akcijām un stabilai brīvprātīgo manuskriptu šifrētāju kopienai (Reinsone 2020), no LFK digitālajā arhīvā ievietotajiem ieskenētajiem manuskriptu attēliem datorlasāmā formātā transkribēti jau vairāk nekā 180 000 manuskriptu lapu.
4. *Vienību profilu izveide un metadatu ievade*: digitālā arhīva ilgtermiņa mērķis ir visu kolekciju vienību metadatu ievadišana digitālajā arhīvā, kas nodrošinās daudzpusīgas pētniecības iespējas, kā arī strukturētā veidā ļaus veikt LFK arhīva iepazīšanu jebkuram interesentam. Šī posma īstenošana ir ļoti darbietilpīga, tāpēc tā uzskatāma par LFK digitālā arhīva ilgtermiņa mērķi. Ņemot vērā to, ka LFK arhīva saturs nemītīgi tiek papildināts, šis arhīva materiālu apstrādes posms nekad netiks pabeigts pilnībā.
5. *Tekstu marķēšana*: datorlasāmo tekstu apstrāde, marķējot tajos minētās vietas, personas, darbus un organizācijas. Šis digitalizētā materiāla apstrādes posms pagaidām tiek veikts tikai tām vienībām, kurās marķēto objektu analīze nepieciešama konkrētu pētniecisko mērķu sasniegšanai.
6. *Klasifikatoru izveide*: analogā satura digitalizēšana, iekļaujot digitālajā arhīvā arvien jaunas vienības un to metadatus, būtiski paplašina pētniecībai pieejamo datu apjomu, kas savukārt nodrošina iespējas pētniekiem veidot tematiski daudzveidīgus pētniecisko datu korpusus. Pētnieku sagatavotie klasifikatori nodrošinās šaurāk specializētu datu kopu pieejamību gan vietējiem interesentiem, gan starptautiskajai pētniecībai.

Avotu struktūra LFK digitālajā arhīvā

Folkloras avoti LFK arhīvā glabājas dažādos formātos – rokrakstos, attēlos, audio un video ierakstos. Vissistemātiskāk un ar vienotu numerāciju no LFK pirmsākumiem līdz 21. gs. sākumam apstrādāti folkloras pieraksti rokrakstos jeb manuskripti (atsevišķos gadījumos arī attēli, ja tie iekļauti manuskriptos), savukārt lielākā daļa no pārējiem folkloras dokumentēšanas formātiem sistemātiski nav tikuši apstrādāti. Piemēram, audio un video ieraksti nereti izmantoti tikai materiāla ērtākai dokumentēšanai lauka apstākļos un vēlāk – tajos dzirdamo/redzamo norišu atšifrēšanai manuskriptā, tādējādi par vienību uzskatīts norises atšifrējuma teksts manuskriptā, savukārt fotogrāfijas, audio un video ieraksti labākajā gadījumā ievietoti papildkrājumos ar savu patstāvīgu numerāciju (piemēram, LFK valču kolekcija, LFK lenšu kolekcija), bet sliktākajā gadījumā šis materiāls nav reģistrēts vispār. Ņemot vērā to, ka dažādu norišu dokumentācija attēlos, bet jo īpaši audio un video ierakstos mūsdienās kļūst arvien pieprasītāka, LFK digitālā arhīva izveide, no vienas puses, ir stimulējusi gan attēlu, gan audio, gan video ierakstu digitalizēšanu un, no otras puses, mudinājusi izstrādāt papildinājumus materiālu reģistrēšanas metodoloģijā un numurēšanā, tādējādi iekļaujot šo materiālu kopīgajā LFK arhīva datu struktūrā.

Pētnieciskie avoti digitālajā arhīvā strukturēti trīs līmeņos pēc LFK arhīva analogās versijas parauga: arhīvs pamatā iedalīts krājumos, kas sastāv no vienas vai vairākām kolekcijām, kurās savukārt iekļautas vienības jeb dažādos formātos dokumentētas visdažādākās nemateriālā kultūras mantojuma izpausmes. Šie termini turpinājumā skaidroti detalizētāk.

-) Krājums¹⁸ – institucionāli vai tematiski noteikta LFK kolekciju kopa. Krājumā var tik apkopotas: a) vienas institūcijas folkloras kolekcijas, piemēram, Latviešu folkloras krātuves pamatkrājuma kolekcijas; LFK Bb jeb Krišjāņa Barona biedrības 1924.–1940. gadā savāktās folkloras kolekcijas, kas tika nodotas Latviešu folkloras krātuvei; LFK LiepU jeb Liepājas Universitātes pētnieku un studentu folkloras vākumi, kuru digitālās kopijas deponētas LFK digitālajā arhīvā, u. tml.; b) tematiski veidotas kolekcijas, piemēram, LFK Ak jeb Latviešu folkloras krātuves autobiogrāfiju krājums, kurā apkopotas rakstītās dienasgrāmatas, dzīvesstāsti un atmiņas; SEEM jeb Latvijas Universitātes un Sanktpēterburgas Valsts universitātes organizēto Sibīrijas etnolingvistisko ekspedīciju materiāli, kuru digitālās kopijas deponētas LFK digitālajā arhīvā, u. tml. Digitālā arhīva sadaļā “Krājumi” pieejami krājumu apraksti, ko papildina informācija par kolekcijām krājumos, pievienotajiem failiem, teksta vienībām, attēliem un ierakstiem – šie dati aplūkojami indeksa formā¹⁹. LFK digitālā arhīva krājumi detalizētāk raksturoti pēdējā apakšnodaļā.
-) Kolekcija – pēc dažādiem kritērijiem veidots vienību kopums, piemēram: a) LFK un LFK Bb krājumos – viena individuālā folkloras vācēja noteiktā laika periodā iesniegto materiālu kopums; b) LFK krājumā – vienas organizācijas (skolas, augstskolas, retāk arī cita veida iestādes) noteiktā laika periodā iesniegto materiālu kopums; c) LFK krājumā – konkrētā gada

18 LFK digitālā arhīva krājumu apraksti ievietoti atsevišķā apakšnodaļā “LFK digitālā arhīva krājumu raksturojums” raksta beigās.

19 <http://garamantas.lv/lv/repository/index> [skatīts 15.12.2021.]

LFK zinātniskās ekspedīcijas laikā savāktie materiāli; d) LFK LiepU krājumā – konkrētā gada Liepājas Universitātes studentu prakšu materiālu kopums; e) LFK Ak krājumā atsevišķa kolekcija tiek veidota katrai autobiogrāfijai vai dienasgrāmatai; f) FMP krājumā kolekciju veido konkrēta publikācija (daudzskējumu publikāciju gadījumā – konkrēts publikācijas sējums) u. tml. LFK folkloras kolekcijās var būt iekļauta arī tikai viena vienība, tomēr parasti – no dažiem simtiem līdz dažiem tūkstošiem, retāk arī vairāki desmiti tūkstoši vienību. Kolekcijas tiek numurētas krājuma ietvaros, kolekcijas numurs norādīts vienības numurā aiz krājuma abreviatūras (piemēram, LFK Bb 13 – Barona biedrības krājuma 13. kolekcija).

Informācija par kolekcijām LFK digitālajā arhīvā pieejama digitālā arhīva sadaļā “Kolekcijas”²⁰. Indeksa skatā aplūkojamās kolekcijas iespējams atlasīt pēc dažādiem kritērijiem (pēc kolekcijas numura, nosaukuma, krājuma, iesūtīšanas gada, iesniedzēja – personas vai iestādes, pēc kolekcijas vienību pieraksta valodas). Tāpat indeksa skatā redzams, vai kolekcijas materiāli ir digitalizēti (norādīts kolekcijai pievienoto failu skaits) un kāds ir failu transkribēšanas progress tajā. Šī informācija pieejama, atverot arī interesējošās kolekcijas profila skatu, kur papildus izlasāms arī detalizētāks kolekcijas satura apraksts (pieejams daļai kolekciju). Ņemot vērā to, ka digitālais arhīvs ir brīvi pieejams tiešsaistē un orientēšanās šajā materiālā jāveic pašiem lietotājiem, digitālā arhīva izveide ir stimulējusi arhīva kolekciju aprakstu sagatavošanu. Kolekciju aprakstus gatavo LFK pētnieki, tajos detalizētāk raksturojot kolekcijā iekļauto folkloras materiālu – kolekcijas tapšanas vēsturisko kontekstu, folkloras pierakstītāju(s) un teicēju(s), pieraksta vietas, žanru pārstāvēniecību un citas interesantas kolekcijas nianšes. Tāpat kolekcijas profilā indeksa veidā iespējams detalizētāk aplūkot kolekcijai pievienotos failus (ja kolekcija ir digitalizēta), ja kolekcijai ir pievienoti metadati – detalizētu kolekcijās reģistrēto vienību sarakstu, pa lomām sadalītu (kolekcijas iesūtītājs, vienību teicēji un pierakstītāji) ar kolekciju saistīto personu sarakstu, savukārt kartes pārlūkā – ar kolekciju saistītās kartētās vietas (lielākoties – kolekcijas vienību dokumentēšanas vietas; ja veikta tekstu marķēšana – arī tekstos minētās vietas)²¹.

-) Vienība – sīkāka satura elements LFK analogā un digitālā arhīva struktūrā. Saskaņā ar vēsturiski pieņemto arhivēšanas praksi LFK iesniegtais materiāls klasificēts pēc žanriskā principa, t. i., vienu vienību veido, piemēram, viena pasaka, viena teika, tautasdziesma, rotaļas vai dejas apraksts, sakāmvārds u. tml. Katrai vienībai LFK arhīvā tiek piešķirts unikāls numurs, kas sastāv no krājuma abreviatūras, kolekcijas numura un vienības numura (piemēram, LFK Bb 13, 1689 nozīmē Barona biedrības krājuma 13. kolekcijas 1689. vienību).

Gan krājuma, gan kolekcijas profila skatā pieejams šobrīd digitālajā arhīvā iekļauto attiecīgā krājuma vai kolekcijas vienību indekss, no kura iespējams nokļūt interesējošās vienības profilā. Vienības apraksta cilnē apkopota visa zināmā informācija par konkrēto vienību – tās numurs, nosaukums, žanrs, pierakstītājs, iesūtītājs, teicējs, pierakstīšanas gads un vieta, pieraksta valoda un dialekts (jāņem vērā, ka reizēm vienības metadati var būt norādīti arī fragmentāri).

20 <http://garamantas.lv/lv/collection/index> [skatīts 15.12.2021.]

21 Piemēram, <http://garamantas.lv/lv/collection/885704/Baunu-Milites-pamatskolas-vakums> [skatīts 15.12.2021.]

Ja attiecīgais materiāls ir digitalizēts, vienības profila faila cilnē ir pieejama avota digitālās kopijas datne – manuskripta lapa, attēls, klausāms audio vai skatāms video fails. Ja šī faila saturs ir atšifrēts datorlasāmā formā, faila cilnē ir pieejama precīza, nerediģēta satura transkripcija, savukārt teksta cilnē pieejams digitālā arhīva redaktoru rediģēts vienības teksts (tekstā piemērotas mūsdienu latviešu vai latgaliešu rakstu valodas pareizrakstības un interpunkcijas normas, saglabājot dialektu iezīmes). Daļai vienību tekstu veikta arī tekstā minēto vietu, personu un organizāciju marķēšana, attēliem – tajos redzamo vietu atzīmēšana; šīs darbības tiek veiktas izlases kārtībā konkrētām pētnieciskām vajadzībām. Vienības profila vietu cilnē kartes pārlūkā iespējams aplūkot vienības pieraksta vietu (ja tā ir zināma un ir kartēta), kā arī vienības tekstā minētās vietas (ja tās ir marķētas). No vienības profila iespējams turpināt plašāku ar vienību saistītā konteksta iepazīšanu, ar profilā iestrādātajām saitēm vajadzības gadījumā nokļūstot saistīto kolekciju, personu, organizāciju un vietu profilos.

LFK digitālā arhīva avotu metadati

LFK digitālā arhīva metadati jeb informācija, kas vairāk vai mazāk kontekstualizē arhīvā iekļautās vienības, iedalāma vairākās savstarpēji saistītās datu kategorijās, no kurām četras veido metadatu struktūras pamatu – personas, organizācijas, vietas un notikumi –, un katrai no tām, izņemot notikumus, digitālajā arhīvā atvēlēta atsevišķa datubāze, dodot iespēju arhīva datus iepazīt no dažādiem skatpunktiem.

Personas LFK digitālā arhīva kontekstā var būt dažādās lomās, tomēr visbiežāk tie ir folkloras teicēji, vācēji, iesūtītāji un pētnieki. Tās var būt arī personas, kas dažādos kontekstos minētas folkloras un citu vienību saturā (piemēram, vēsturiskas personas, par kurām stāstīts nostāstos un teikās; cilvēki, kas minēti dienasgrāmatās; personas, kas minētas dzimtu pētniecības materiālos; cilvēki, kas redzami fotogrāfijās u. tml.). LFK pieejamā informācija par personām apkopota Personu datubāzē (publiski redzamas tikai tās personas, kuru datus neaizsargā Eiropas Savienības personu datu aizsardzības regula). Datubāze pārlūkojama indeksa skatā²², kurā iespējams veikt meklēšanu pēc personas vārda, dzīvesvietas, dzimšanas gada un nodarbes, kā arī aplūkot kopsavilkumu par LFK arhīva avotiem, kas ar konkrēto personu saistīti (saistīto vienību, ierakstu un attēlu saraksts, saistītās kolekcijas). Detalizētāka informācija atrodama konkrēto personu profilā: personas apraksta cilnē iekļauta pamatinformācija (personas vārds, pseidonīmi, nodarbes un tekstuāls apraksts), notikumu cilnē iekļauti LFK pieejamie personas biogrāfijas dati (dzimšanas, miršanas un apbedīšanas laiks un vieta, izglītotošanās un darba vietas, dalība organizācijās, apbalvojumi utt.), veikuma cilnē ievietotas saites uz indeksiem, kuros apkopotas ar personu saistītās kolekcijas, teksta, audio un video vienības, attēli un darbi, savukārt vietu cilnē aplūkojama kartes pārlūkā vizualizēta ar personu saistīto notikumu ģeogrāfija. Jāatzīmē, ka lielākoties LFK rīcībā esošā informācija par folkloras dokumentēšanas procesā iesaistītajām personām ir ļoti fragmentāra.

22 <http://garamantas.lv/lv/person/index> [skatīts 23.04.2022.]

Organizācijas LFK digitālā arhīva kontekstā lielākoties figurē divos kontekstos – 1) organizācijas kā LFK partneri, kas iesaistījušās folkloras vākšanas procesā, un 2) organizācijas, kas caur notikumiem saistītas ar personu biogrāfijām. Organizācijas var būt minētas un marķētas arī LFK avotu saturā. Informācija par LFK digitālajā arhīvā iekļautajām iestādēm apkopota Organizāciju datubāzē²³, kas pārlūkojama indeksa skatā. Indeksā iespējams atlasīt organizācijas pēc nosaukuma, veida un atrašanās vietas, savukārt plašāka informācija par konkrētu iestādi aplūkojama tās profilā. Organizācijas profila apraksta cilnē minēts organizācijas aktuālais un alternatīvie (vēsturiskie, tautā lietotie utt.) nosaukumi, detalizētāks apraksts (ja pievienots), dibināšanas un likvidēšanas datums. Personu cilnē apvienotas ar organizāciju saistītās personas, vietu cilnē karšu pārlūkā apskatāma(s) organizācijas atrašanās vieta(s), savukārt kolekciju, vienību, attēlu, ierakstu un izdevumu cilnēs indeksa skatā aplūkojami un pēc dažādiem kritērijiem atlasāmi ar konkrēto iestādi saistītie avotu materiāli.

Vietas LFK digitālajā arhīvā figurē dažādos kontekstos. Tās var būt: 1) vienību dokumentēšanas vietas; 2) ar personu biogrāfijām saistīto notikumu vietas; 3) organizāciju atrašanās vietas; 4) darbu publicēšanas vietas un 5) avotos minētās vietas. Šīs vietas pēc kartēšanas tiek apkopotas Vietu datubāzē²⁴, kas pārlūkojama indeksa skatā, kurā iespējams veikt datu atlasīšanu pēc nosaukuma, adreses, veida un apraksta. Plašāka informācija par vietām pieejama vietas profilā: apraksta cilnē norādīts vietas standartizētais un alternatīvie (vēsturiskie, paralelie un tautā lietotie) nosaukumi, vietas veids, adrese, ģeogrāfiskās koordinātes, kartēšanas precizitāte, statuss (eksistējoša vai neeksistējoša vieta), kā arī karšu pārlūkā apskatāma(s) tās atrašanās vieta(s). Pārējās cilnēs iekļauti ar konkrēto vietu saistīto datu indeksi – šajā vietā dokumentētie NKM avoti, ar vietu saistītās personas, organizācijas, vietas attēli utt. Informāciju par konkrētām vietām iespējams aplūkot arī *OpenStreetMap* karšu pārlūkā²⁵, vajadzīgo vietu atlasot pēc tās atrašanās vietas kartē. Nospiežot uz konkrētās vietas atzīmes kartē, iespējams aplūkot ar šo vietu saistīto relāciju skaitu (ar vietu saistīto teksta vienību, ierakstu, attēlu, personu, organizāciju un darbu skaitu), kā arī nokļūt vietas profilā.

LFK digitālā arhīva krājumu raksturojums

Apakšnodaļā sniegts konspektīvs pārskats par LFK digitālajā arhīvā iekļautajiem krājumiem un to digitalizēšanas progresu.

LFK – Latviešu folkloras krātuves pamatkrājums, kurā iekļauti visu folkloras žanru pieraksti un ieraksti. Pirmās folkloras vienības LFK krājumā reģistrētas vien gadu pēc LFK dibināšanas, 1925. gada 3. septembrī, un materiāla papildināšana turpinās arī mūsdienās. Krājumā iekļautas LFK pētnieku individuālās kolekcijas un ikgadējo LFK zinātnisko ekspedīciju

23 <http://garamantas.lv/lv/organization/index> [skatīts 23.04.2022.]

24 <http://garamantas.lv/lv/place/index> [skatīts 23.04.2022.]

25 <http://garamantas.lv/lv/map/index> [skatīts 23.04.2022.]

materiāli, skolu un augstskolu audzēkņu, organizāciju dalībnieku, kā arī individuālo folkloras vācēju kolekcijas. Tā kā pēc LFK dibināšanas tās rīcībā nonāca arī senāki folkloras vākumi, krājums hronoloģiski aptver laika posmu no 19. gs. otrās puses līdz mūsdienām. Krājumu šobrīd veido 2246 kolekcijas, kurās kopumā reģistrēti vairāk nekā 3 miljoni folkloras vienību. Krājumā ietilpst manuskriptu, fotogrāfiju, audio un video ierakstu kolekcijas. Apjoma ziņā plašāko segmentu krājumā veido manuskripti, kuru sistematiska digitalizēšana, pateicoties Latvijas Republikas Izglītības un zinātnes ministrijas apakšprogrammas finansējumam²⁶, notiek kopš 2015. gada, gadā ieskenējot vairāk nekā 70 000 manuskriptu lapu. Turpinot digitalizēšanu šādā apjomā, prognozējams, ka LFK pamata krājums tiks ieskenēts līdz 2025. gadam. Nelielai daļai kolekciju pievienoti metadati – 2021. gadā tika uzsākta sistematiska LFK bijušo darbinieku – folkloristu – kolekciju apstrāde. Tāpat metadati izlases kārtībā tiek pievienoti vienībām, kas izmantotas atsevišķu pētniecisko projektu vajadzībām. Būtisku LFK krājuma daļu veido arī cita veida audiovizuālie materiāli. Nozīmīga daļa no audio un video materiāliem krājumā tika digitalizēta 2021. gadā Valsts kultūrkapitāla fonda finansētajā mērķprogrammā “Kultūrelpa”²⁷. Tāpat digitālajā arhīvā ievietoti pēdējo gadu LFK zinātnisko ekspedīciju digitālie audio, video un foto materiāli, tomēr, ņemot vērā fizisko personu datu aizsardzības tiesisko regulējumu, šis materiāls publiski nav pieejams.

LFK Bd – Bērnu dziesmu krājums, kurā iekļauta tikai viena kolekcija – LFK Bd [1]. Krājums tapis no 1922. līdz 1930. gadam pēc Rīgas 2. vidusskolas skolotājas, vēlāk LFK dibinātājas Annas Bērkalnes ierosinājuma. Viņa izstrādāja aptaujas lapu un materiāla vākšanā iesaistīja skolēnus. Krājumā iekļautas 42 358 folkloras vienības. Manuskripti pilnībā ieskenēti, metadati krājumam nav pievienoti.

LFK Bb – Barona biedrības krājums. Viens no Barona biedrības (1924–1940) uzdevumiem bija turpināt Krišjāņa Barona folkloras vākšanas darbu. Lielu daļu no krājuma veido Rīgas Skolotāju institūta audzēkņu vākumi. Krājumā iekļautas 49 kolekcijas, kurās kopumā reģistrēta 53 471 vienība. Krājums sastāv no manuskriptiem un tajos iekļautajiem attēliem (zīmējumiem). Manuskripti pilnībā ieskenēti, lielākajai daļai kolekciju pievienoti metadati.

LFK LiepU – Liepājas Universitātes krājums. Liepājas Universitātes Kurzemes Humanitārajā institūtā glabātais un LFK digitālajā arhīvā deponētais manuskriptu krājums, kurā apkopotu pētnieku organizēto etnogrāfisko ekspedīciju materiāli, kā arī laika posmā no 1987. līdz 2012. gadam iesniegtie 300 studentu folkloras prakšu vākumi. Krājumā līdz 2021. gadam iekļautas 27 kolekcijas, kurās reģistrētas 12 503 vienības. Krājuma digitalizēšana un metadatu apstrāde notiek Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi”

26 LR Izglītības un zinātnes ministrijas budžeta apakšprogramma 05.04.00. “Krišjāņa Barona Dainu skapis” (programmas vadītāji: Rita Grīnvalde, 2015–2020; Sandis Laime, 2021–).

27 LU LFMI īstenotais Valsts kultūrkapitāla fonda mērķprogrammas “KultūrELPA” projekts “Daudzveidīga tradicionālās kultūras satura nodrošinājums digitālajā vidē” (vad. Ieva Tihovska).

īstenotajā projektā²⁸ un turpinās arī 2022. gadā. Projektam noslēdzoties, krājums būs pilnībā ieskenēts un tiks pabeigta krājuma metadatu apstrāde.

LFK RTA – Rēzeknes Tehnoloģiju akadēmijas krājums. Rēzeknes Tehnoloģiju akadēmijā glabātais un LFK digitālajā arhīvā deponētais manuskriptu krājums, kurā apkopoti 1993.–1995. gada Rēzeknes Tehnoloģiju akadēmijas (tobrīd – Rēzeknes Augstskolas) filoloģijas studentu folkloras prakšu vākumi. Krājuma digitalizēšana un metadatu apstrāde notiek Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi” īstenotajā projektā.

LFK Ak – Autobiogrāfiju krājums. Krājuma veidošana aizsāka 2018. gadā LFK digitālajā arhīvā ar mērķi saglabāt un dokumentēt dzīves pierakstīšanas tradīciju Latvijā, kā arī veicināt šīs tradīcijas turpināšanos. Krājumu pamatā veido rakstītas dienasgrāmatas, dzīvesstāsti un atmiņas, kā arī dažādi materiāli, kas sniedz papildinošu informāciju – fotogrāfijas, intervijas ar rakstītājiem un radnieku stāsti par viņiem. Krājums dažādu projektu gaitā²⁹ tiek aktīvi papildināts, veicot jauniegūto materiālu digitalizēšanu – skenēšanu, manuskriptu tekstu transkribēšanu un metadatu pievienošanu. Līdz 2021. gada beigām krājumā ir iekļautas 197 pilnībā ieskenētas un apstrādātas kolekcijas.

NMV – Nacionālās mutvārdu vēstures krājums. Krājums kopš 1992. gada veidots mutvārdu vēstures pētījumu gaitā LU Filozofijas un socioloģijas institūtā sadarbībā ar Latvijas mutvārdu vēstures pētnieku asociāciju “Dzīvesstāsts”. Krājuma 57 kolekcijās iekļauti 4556 dzīvesstāstu un mutvārdu vēstures interviju skaņu ieraksti. Ņemot vērā šo avotu sensitīvo raksturu, digitālajā arhīvā publiski pieejami tikai dati par intervijām, savukārt interviju ieraksti pieejami tikai pēc pamatota intervētās personas radu vai pētnieku pieprasījuma iesniegšanas.

SEEM – Sibīrijas etnolingvistisko ekspedīciju materiāli. Krājumā apkopoti un LFK digitālajā arhīvā deponēti Latvijas Universitātes un Sanktpēterburgas Valsts universitātes desmit Sibīrijas etnolingvistisko ekspedīciju (2004–2009) nozīmīgākie audio, video un foto materiāli. Ekspedīcijas lielākoties notikušas Krievijas Federācijas Krasnojarskas apgabalā ar mērķi dokumentēt austrumu diasporas (Sibīrijas latviešu) vēstures, valodas un kultūras liecības. Krājumu veido 11 kolekcijas, kurās iekļauti audio, video un foto materiāli, kā arī to metadati. Krājuma digitalizācija veikta 2021. gadā ar Valsts kultūrkapitāla fonda mērķprogrammas “Kultūrelpa” finansiālu atbalstu³⁰.

LFK Dig – LFK digitāli radītu materiālu krājums izveidots 2020. gadā ar mērķi apkopot materiālus, kas LFK iesniegti, izmantojot LFK digitālā arhīva tiešsaistes aptauju rīku. 2021. gada

- 28 LU LFMI un sadarbības partneru īstenotais Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi” projekts “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001) (vad. Sanita Reinsons).
- 29 Latvijas Zinātnes padomes fundamentālo un lietišķo pētījumu projekts “Sevis dokumentēšana: inovatīvi modeļi autobiogrāfiskā naratīva interpretācijā un izpratnē” (2018–2021; vad. Sanita Reinsons).
- 30 Biedrības “LgSC” īstenotais Valsts kultūrkapitāla fonda mērķprogrammas “Kultūrelpa” projekts “Sibīrijas etnolingvistisko ekspedīciju materiālu publiskošana” (vad. Lidijs Leikums).

beigās šajā krājumā ir divas kolekcijas, kurās apkopoti divu projektu gaitā veikto tiešsaistes aptauju rezultāti.

LFK FMP – Folkloras materiālu publicējumi. Krājums izveidots LFK digitālajā arhīvā ar mērķi digitalizēt un iekļaut plašākā pētnieciskā un izziņas aprītē nozīmīgākās folkloras materiālu publikācijas. Krājumā līdz šim iekļautas nozīmīgas Friča Brīvēznieka, Edītes Kurcas un Zvaigžņu Andža buramvārdu publikācijas, Emīļa Melngaiļa un Jēkaba Vītoliņa tautas mūzikas materiālu publikācijas, kā arī 2021. gadā ar Valsts kultūrkapitāla fonda mērķprogrammas “Kultūrelpa” finansiālu atbalstu digitalizētie akadēmiskā izdevuma “Latviešu tautsdziesmas” pirmie deviņi sējumi³¹.

LFK Nmk – Novadpētniecības materiālu krājums. Krājums izveidots 2018. gadā ar mērķi publicēt LFK iesniegtos novadpētnieciskos materiālus, tostarp ģeoloģiskus dzimtu pētījumus, kultūrvēsturisku vietu un pagastu vēstures aprakstus, vēstures avotus un tamlīdzīgus materiālus. Līdz 2021. gada beigām krājumā iekļautas deviņas kolekcijas, kas pilnībā digitalizētas.

Secinājumi

Folkloras publicējumi ir vieni no pirmajiem latviešu valodas materiāliem, kas tikuši digitalizēti jau kopš 20. gadsimta 60. gadu beigām. Līdz 2014. gadam folkloras materiālu digitalizācijā un pieejamības nodrošināšanā raksturīgā tendence ir veidot savrupas, salīdzinoši nelielas digitālās kolekcijas vai datubāzes. LFK digitālais arhīvs *garamantas.lv* nāk ar konceptuāli jaunu pieeju, veidojot daudzpusīgu digitālo resursu, kura dažādās daļas un apakšprojekti ir sasaistīti vienā sistēmā. Digitālā arhīva izveide ir veicinājusi ne tikai LFK arhīva materiālu pieejamību daudz plašākam interesentu lokam, bet arī būtiski stimulējusi attēlu, audio un video arhīvu sakārtošanu un iekļaušanu LFK krājumā, arhīva papildināšanu ar jauniem, digitāli radītiem materiāliem, nodrošinājusi plašākas sabiedrības iesaisti arhīva materiālu apstrādē, kā arī, izmantojot dažādas digitālo humanitāro zinātņu pieejas, pavērusi daudz plašākas iespējas materiāla izpētei un iekļaušanai starptautiskajā pētnieciskajā aprītē. Astoņu gadu laikā *garamantas.lv* ir tapušas par centrālo Latvijas tradicionālās kultūras materiālu digitālo krātuvi un turpina attīstīties sadarbībā ar citām institūcijām, integrējot aizvien jaunas folkloras un cita veida materiālu kolekcijas.

31 LU LFMI īstenotais Valsts kultūrkapitāla fonda mērķprogrammas “KultūrELPA” projekts “Daudzveidīga tradicionālās kultūras satura nodrošinājums digitālajā vidē” (projekta vad. Ieva Tihovska, aktivitātes vad. Ginta Pērle-Stīle).

- Bruebach, Nils (2003). Archival science in Germany – Traditions, Developments and Perspectives. *Archival Science*, No. 3(4), pp. 379–399.
- Bula, Dace (2018). Īpašuma metaforas folkloras izpratnē: ieskats jēdziena *garamantas* vēsturē. *Letonica*, Nr. 38, 31.–43. lpp.
- Eglāja-Kristsons, Eva (2021). The Digital Resource Literatura.lv: An Introduction. *Colloquia*, No. 47, pp. 149–159. Pieejams: <http://www.llti.lt/failai/Publikacija.pdf> [skatīts 10.09.2022.].
- Eglāja-Kristsons, Eva, Raudive, Signe (2020). Sabiedrības iesaistes akcija “Lasi skaļi”: estētiskie un izglītojošie dzejas ieskaņošanas aspekti. *Letonica*, Nr. 42, 66.–83. lpp.
- Huvila, Isto (2008). Participatory archive: towards decentralised curation, radical user orientation, and broader contextualisation of records management. *Arch Sci*, No. 8, pp. 15–36.
- Ketelaar, Eric (1992). Archives of the people, by the people, for the people. *South African Archives Journal*, No. 34, pp. 5–16.
- Kokina, Silvija (1993). “Latvju dainu” datormasīva izveide. *Latvijas Vēstnesis*, 30.10., 4. lpp.
- Pakalns, Guntis (2005). P. Šmita “Latviešu pasaku un teikņu” tulkojums vāciski – vēsture un digitālā publikācija. *Meklējumi un atradumi: 2005*. Rīga: Zinātne, 145–171. lpp.
- Reinsone, Sanita (2018). Participatory practices and tradition archives. Harvilahti, Lauri, Kjus, Audun, O’Carroll, Cliona, Österlund-Pötzsch, Susanne, Skott, Fredrik, Treija, Rita (eds.). *Visions and Traditions: Knowledge Production and Tradition Archives*. Helsinki: Academia Scientiarum Fennica.
- Reinsone, Sanita (2020). Searching for deeper meanings in cultural heritage crowdsourcing. Hetland, Per, Pierroux, Palmyre, Esborg, Line (eds.). *A History of Participation in Museums and Archives: Traversing Citizen Science and Citizen Humanities*. Routledge, pp. 186–217.
- Vikis-Freibergs, Vaira, Freibergs, Imants (1978). The Sun-songs of Latvian folklore: A computer-accessible corpus. *Journal of Baltic Studies*, No. 9(1), pp. 20–31.
- Vikis-Freibergs, Vaira, Freibergs, Imants (1978). Formulaic Analysis of the Computer-Accessible Corpus of Latvian Sun-Songs. *Computers and the Humanities*, No. 12(4), pp. 329–339.
- Vikis-Freibergs, Vaira (1999). Similarity and contrast as structuring principles in the Latvian Daina Quatrain. *Journal of Baltic Studies*, No. 30(3), pp. 201–226.

The Digital Archive of the Archives of Latvian Folklore *garamantas.lv*: Prehistory and Development

Sanita Reinsone, Sandis Laime

Keywords: digital archive, digitization of intangible cultural heritage, history of digitization, Archives of Latvian Folklore

The aim of this article is to provide an overview of the history of the digital archive of the Archive of Latvian Folklore (ALF) and other digital resources, the process of data digitization and processing, as well as the structure and content of the digital archive. The main task of the ALF since its establishment in 1924 until today is to document the intangible cultural heritage of the Latvian population, including ethnic minorities, in various formats, to conduct its research and to publish both research results and scientifically structured selections of research sources. The ALF archive is one of the largest repositories of intangible cultural heritage in Europe, containing more than three million folklore texts.

Folklore publications are among the first Latvian language materials that have been digitized since the late 1960s. Until 2014, the characteristic trend in the digitization and accessibility of folklore sources is to create separate, relatively small digital collections or databases. The ALF digital archive *garamantas.lv* comes with a conceptually new approach, creating a versatile digital resource, the various parts and sub-projects of which are connected in one system. The creation of the digital archive has not only promoted the availability of ALF archive materials to a much wider range of interested parties, but also significantly stimulated the organization and inclusion of illustration, audio and video archives in the ALF collections, supplementing the archive with new, digitally created materials, ensured the involvement of the wider public in the processing of archive materials, as has also opened much wider possibilities for the study of the material using various digital humanities approaches and for inclusion in the international research circuit. Over the course of eight years, *garamantas.lv* has become the central digital repository of Latvian traditional cultural materials and continues to develop in cooperation with other institutions, integrating more and more collections of folklore and other types of materials.

Metadatos balstīta dienasgrāmatu teksta korpusa analīze

Sanita Reinsone, Haralds Matulis,
Ilze Ļaksa-Timinska

Raksts ir tapis Latvijas Zinātnes padomes projektā Nr. lzp-2018/1-0073 veiktā pētījuma rezultātā.

Ievads

Personiskās dienasgrāmatas ir unikāls laikmeta naratīvais liecinājums, kas laikmetu, notikumus atklāj caur dziļi personisku skatījumu. Nešaubīgi katra dienasgrāmatā ir unikāla, jo rakstīta noteiktā vietā un laikā, pierakstos iekapsulējot daudzas autoru šodienas. Tomēr tās visas kopā rada īpašu personisko rakstījumu dimensiju, kur individuālo stilistiku veido autora gaume un rakstīšanas prasme, arī dienasgrāmatu rakstīšanas tradīciju, daiļliteratūras un citu kultūras parādību ietekmes. Atzīstot katras dienasgrāmatas unikalitāti un savpatību, šajā rakstā mēs pievēršamies personiskajām dienasgrāmatām kā kopumam un veicam metadatos balstītu dienasgrāmatu pilotkorpusa analīzi ar mērķi atklāt korpusa īpatnības dažādu kritēriju sastatījumā.

Šī raksta pamatā ir nesena pieredze, kas gūta, veidojot personīgo dienasgrāmatu tekstu korpusu no paša sākuma, proti, sākot ar dienasgrāmatu vākšanu, digitalizāciju un tālāk – materiālu apstrādi korpusa izveidošanai. Šajā rakstā analizējam dienasgrāmatu tekstu korpusa veidošanu, metodoloģiskos pārbaudījumus un sarežģījumus, ar kādiem nācies saskarties, īstenojot šādu iniciatīvu.

Personīgo dienasgrāmatu pētniecībā līdz šim gluži saprotamā kārtā dominējušas kvalitatīvās pieejas. Tās izmantotas arī tad, ja analizējamais materiāls ir bijis ļoti plašs un sarežģīti pārskatāms. Pētnieki apbrīnojama kārtā spējuši apgūt un sistematizēt pavisam apjomīgas dienasgrāmatas, atrodot un piefiksējot sev interesējošās tēmas un teksta elementus. Kā nesenā publikācijā norāda Džūlija Raka: “[Š]obrīd dzīves pierakstīšanas pētniecības metodes ir, šķiet, pretinā ar lielo datu metodēm.” (Rak 2019: 118) Kaut arī datormetozu izmantošana humanitāro zinātņu pētījumos ir pazīstama jau kopš 20. gadsimta 50. gadiem (Nyhan, Flinn 2016: 1–4; Sula, Hill 2019: 190–206), pēdējās desmitgadēs digitālās humanitārās zinātnes attīstījušās patiesi strauji. No sākotnēji dominējošās datorlingvistikas kā centrālās digitālo humanitāro zinātņu¹ nozares digitālās humanitārās zinātnes ir būtiski paplašinājušās un metodoloģiski sazarājušās, aizvien dziļāk ietiecoties literatūrzinātnes un citu humanitāro zinātņu diskursīvajā laukā.

1 Digitālās humanitārās zinātnes ir salīdzinoši jauns termins, kas tiecas aptvert dažādas humanitāro zinātņu apakšnozares un datormetodes. Iepriekš dominējis jēdziens gan “humanitāro zinātņu programmēšana” (*humanities computing*), gan citi vairāk vai mazāk iekļaujoši vai nozārēm specifiski nosaukumi (Hockey 2004; Unsworth 2004).

Lai gan par datormetožu lietojumu dzīves pierakstīšanas pētniecībā ir paustas zināmas bažas, jo sevišķi saistībā ar autoru privātumu un ētiku (Rak 2019: 117–118), mūsaprāt, autobiogrāfisko materiālu izpētei ar tāllasījuma (*distant reading*) metodēm (Moretti 2013), kas darbojas rokrakā ar kvalitatīvās izpētes paradigmu, ir vērā ņemama perspektīva un labs attīstības potenciāls dzīves pierakstīšanas fenomena izpētē. Jo sevišķi pētnieciski interesantas šādā perspektīvā ir personīgās dienasgrāmatas.

Formāli raugoties, dienasgrāmatas sadalītas daudzās sīkākās vienībās, kur katrai no tām ir atšķirīga metainformācija. Tās ir piesaistītas konkrētam autoram un laikam, kas gluži kā laikrakstu publikācijas, kuras veiksmīgi analizētas un pētītas ar tāllasījuma metodi (Koncar et al. 2020), ļauj laikā pārlūkot un distancēti vērot tēmu un konceptu attīstību, autora noskaņojumu pret noteiktām tēmām un, protams, arī valodas savdabību, variācijas un izmaiņas, ko var vērot diahroniskā attīstībā, ko jo sevišķi ļauj tās dienasgrāmatas, kas nav literāri rediģētas. Tomēr autobiogrāfisko tekstu korpusu veidošana nav vienkārša un ir saistīta ar daudziem sarežģījumiem, ar kādiem nesaskartos, piemēram, periodikas korpusu vai literāro tekstu korpusu veidotāji.

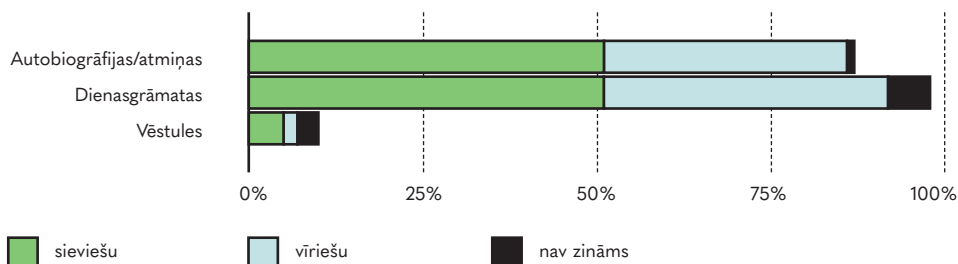
Ceļā uz korpusu – autobiogrāfiju kolekcijas veidošana

Iecere par autobiogrāfisko tekstu korpusu radusies vienlaikus ar Autobiogrāfiju krājuma izveidošanu. Līdz 2018. gadam Latvijā nebija atsevišķas kultūras mantojuma institūcijas, kuras mērķis būtu veidot vienotu un iekļaujošu autobiogrāfisko rakstījumu kolekciju. Tiesa, Latvijas muzeju un bibliotēku krājumos, arī arhīvos un speciālos pētniecisko institūciju krājumos glabājas ievērojama autobiogrāfiskā mantojuma daļa. Tomēr institūcijās glabāto materiālu apjomu un dažādību nav vienkārši precizēt, jo reģistri ir grūti pieejami (tie ne vienmēr ir publiski), tikai daļa no tiem ir digitalizēti un pieejami tiešsaistē², un nereti šādi materiāli netiek īpaši pierēģistrēti vispār, jo stāv ārpus institūcijas pamatdarbības (piemēram, reģionālajās bibliotēkās).

Turklāt ievērojama autobiogrāfiskā mantojuma daļa nepārtraukti top un glabājas arī privātajos un ģimeņu arhīvos. Šo materiālu autori, autoru pēcteči vai citi īpašnieki tos visbiežāk nenodod glabāšanai kultūras mantojuma institūcijās, jo to glabātājiem ne vienmēr ir skaidrs, kuru institūciju konkrētais materiāls varētu interesēt, kā arī daudzos gadījumos tie oriģinālu vēlas paturēt savā ģimenes arhīvā, nododot mantojumā nākamajām paaudzēm, bet reti kuru mantojuma institūciju interesē uzglabāt digitālu materiālu bez fiziskā oriģināla. Kopumā autobiogrāfiskais mantojums Latvijā vērtējams kā izkaisīts un sadrumstalots. Tomēr tā nav tikai Latvijai vien raksturīga situācija.

Lai līdzsvarotu šo situāciju un autobiogrāfiskajam mantojumam piešķirtu lielāku un redzamāku nozīmi, 2018. gada nogalē LU Literatūras, folkloras un mākslas institūta Latviešu

2 Piemēram, muzeju katalogs un arhīvu datubāze: <http://www.nmkk.lv>.



1. attēls. LFK Autobiogrāfiju krājuma kolekciju pārskaits pēc rakstījuma formas.

folkloras krātuvē³ tika izveidots Autobiogrāfiju krājums ar mērķi nodrošināt vietu, kur ilgtspējīgi un pieejami saglabāt dažāda veida dzīves rakstījumus neatkarīgi no to valodas, autora reģionālās piederības, reliģijas, politiskajiem uzskatiem, nopelniem politiskajā, sociālajā vai kultūras dzīvē u. tml.

Pirmajos krājuma pastāvēšanas gados lielākā uzmanība tika pievērsta tieši personīgajiem jeb mājas arhīviem, kuros autobiogrāfiskie materiāli ir visvairāk apdraudēti. Lai pievērstu sabiedrības uzmanību Autobiogrāfiju krājumam, veicinātu izpratni par autobiogrāfisko materiālu vērtību un nepieciešamību tos saglabāt, kā arī lai aicinātu sabiedrību iesaistīties un sniegt materiālus digitālai deponēšanai vai nodošanai arhīvam pavisam, tika īstenotas dažādas publicitātes iniciatīvas, tostarp manuskriptu atšifrēšanas digitālās talkas⁴, mediju kampaņas un atvērtās arhīva dienas⁵, kad ikviens tiek aicināts atnest digitalizēšanai autobiogrāfisku materiālu. Līdz ar to krājuma veidošanas sākumposmā saturu pamatā veido materiāli, ko iesnieguši privātie iesniedzēji. Salīdzinoši nelielu krājuma daļu veido sadarbībā ar mantojuma institūcijām veidotās kolekcijas, kas Autobiogrāfiju krājumam nodotas digitālai deponēšanai.

Pārlūkojot no 2018. gada otrās puses līdz 2021. gada beigām krājumam iesniegtos materiālus (skat. 1. attēlu), redzams, ka aina ir diezgan līdzvērtīga – retrospektīvo formu kolekciju (atmiņu, dzīvesstāstu) krājumā ir tikai nedaudz mazāk nekā dienasgrāmatu kolekciju. Vēstuļu kolekciju niecīgais daudzums izskaidrojams ar to, ka publiskajos uzsaukumos izskanējuši aicinājumi iesniegt dienasgrāmatas un dzīvesstāstus, vēstules nepieminot, lai kolekcijas

3 Latviesu folkloras krātuve ir dibināta 1924. gadā. Tā ir viens no lielākajiem un senākajiem folkloras arhīviem Eiropā, kas šobrīd atrodas Latvijas Universitātes Literatūras, folkloras un mākslas institūta paspārnē. Vairāk informācijas: <http://lfk.lv>.

4 Regulāri aicinājumi *garamantas.lv* un sociālajos tīklos atšifrēt konkrētus pētniekiem interesējošos materiālus, piemēram, aicinājums šifrēt garās dienasgrāmatas pa fragmentiem (<http://garamantas.lv/lv/post/1787/lfk-autobiografiju-krajums-aicina-talka>).

5 Piemēram, 2018. gada bija viena no pirmajām “Autobiogrāfiju dienām”, aicinot ikvienu atnest materiālus uz Latviesu folkloras krātuvi (<https://www.facebook.com/754708391535312/photos/a.754721168200701/761805567492261/>).

veidošanas sākumposmā neveidotos pārslodze. Kopskaitā 2021. gada beigās Autobiogrāfiju krājumā iekļautas 207 kolekcijas, kur katru veido viena autora (vai autoru kolektīva) rakstītais. Vairumā no tām dominē viena rakstījumu forma, piemēram, tā ir dienasgrāmata vai dzīvesstāsts. 4,84 % jeb 10 kolekcijās sastopamas vairākas formas.

Tekstuālo kvantitāti šis grafiks gan neparāda. Viena autobiogrāfiskā vienība var būt gan pāris lappušu garš dienasgrāmatas fragments⁶, gan desmitgadēm ilgi rakstīta dienasgrāmata⁷, kā arī atsevišķos gadījumos vienā kolekcijā var būt iekļauts dienasgrāmatu kopums, kas veidots kā atsevišķs projekts, piemēram, Pandēmijas dienasgrāmatu kolekcija Nr. 166⁸, kas ietver 238 autoru dienasgrāmatas. Līdztekus autobiogrāfiskajiem pierakstiem kolekcijās tiek iekļauti arī ar autoriem saistīti papildu materiāli, piemēram, fotogrāfijas. Daļēji publiska piekļuve krājumam tiek nodrošināta LFK digitālajā arhīvā *garamantas.lv*⁹.

Liela nozīme krājuma veidošanā ir sabiedrībai, kas tiek iesaistīta ne tikai krājuma veidošanā, bet arī manuskriptu pārveidošanā datorlasāmā formā, izmantojot LFK digitālā arhīva rīkus manuskriptu atšifrēšanai¹⁰. Digitalizēto manuskriptu pārveidošana datorizētā formā ir viens no svarīgākajiem un arī laikietilpīgākajiem posmiem korpusa un automatiski meklējamas kolekcijas veidošanā, tāpēc brīvprātīgo kopienas darbs ir ārkārtīgi nozīmīgs un varētu teikt – pat izšķirīgs korpusa veidošanā. Sabiedrības interesi var vērtēt atzinīgi, jo salīdzinoši īsā laikā visi pārrakstīšanai piedāvātie manuskripti arī tiek atšifrēti. Dienasgrāmatas un dzīvesstāsti folkloras digitālajā arhīvā ienes jaunu, personisku tematisko dimensiju, kas daudziem līdzstrādniekiem šķiet saistoša un uzmanības vērtā. Veiksmīgajā sadarbībā ar sabiedrību liela nozīme ir LFK digitālā arhīva *garamantas.lv* iedibinātajām tradīcijām un digitālo līdzstrādnieku kopienai, kas kopš 2016. gada ar pieaugošām sekmēm palīdz atšifrēt folkloras manuskriptus (Reinsone 2018: 279–296; Reinsone 2020: 186–207).

Korpusa veidošana

Informācija par dienasgrāmatu korpusu veidošanu un digitālo metožu izmantošanu tā analizē akadēmiskajā literatūrā ir visai skopa, un pētījumus par šo tēmu kopumā nav izdevies

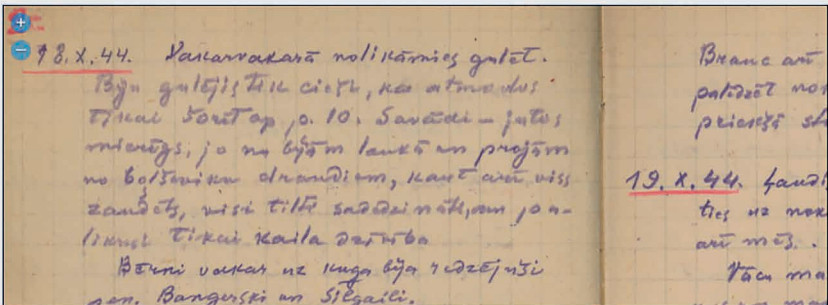
- 6 Piemēram, kolekcijā Nr. 12 ietilpst vien divas lapas no 12 gadu vecas meitenes dienasgrāmatas, kas rakstīta 1944. gadā (Jurča 1944).
- 7 Apjomīgākā dienasgrāmata Autobiogrāfiju krājumā ir Kaspara Aleksandra Irbes (1906–1996) no 1927. līdz 1996. gadam uz 8600 lappusēm sarakstīta dienasgrāmata (Lipša 2021: 415–442).
- 8 Pandēmijas dienasgrāmatu kolekcija tapusi, sākot ar 2020. gada martu: <https://garamantas.lv/lv/collection/1415829/Pandemijas-dienasgramatas-2020>.
- 9 Daļa materiālu, visvairāk dienasgrāmatas, nav publiski pieejami autortiesību, privātuma un citu iemeslu dēļ. Ar katru materiāla iesniedzēju ir noslēgta rakstiska vienošanās, kā šis materiāls tiek apstrādāts un kādas ir piekļuves tiesības.
- 10 Vairāk par Autobiogrāfiju krājumu sk. <http://autobiografijas.lv>.

GARAMANTAS.LV LATVIEŠU FOLKLORAS KRĀTUVES DIGITĀLAIS ARHĪVS LV - meklēt

Kolekcijas | Personas | Attēli | Audio | Organizācijas | Karte | Citi - info ienākt

Sākums / Krājumi / Autobiogrāfiju krājums / Kolekcijas / Alvīla Kalnieša (Klābaha) dienasgrāmatas / Falli / Ak26-Alvīla-Kalniesa-dienasgramata-12-0003

Atšifrēt Iepriekšējais Nākamais



Atšifrējums

Atšifrēja	Dina Ozola
Pabeigts	Jā
Pārbaudīts	Nē
Labots	2021-04-03 13:43:58
Atšifrējums	2. 18.X.44. Vakarvakarā nolikāmies gulēt. Biju gulējis tik cieši, ka atmodos tikai šorīt ap p.10. Savādi - jutos mierīgs, jo nu bijām laukā un projam no boļševiku draudiem, kaut arī viss zaudēts, visi tilti sadedzināti, un palikusi tikai kaila dzīvība. Bērni vakar uz kuģa bija redzējuši gan Bangerski un Silgaili. Ar šo pašu kuģi brauca Velta Toma- Skuja, vecā pliskavieša A-dra Skujas - Cēsu ģimn. skolotāja dēla D.Skujas sieva, pazīstama dzejniece, un Tamāra Anšberga kundze. Abas kundzes bijušas par darbiniecēm Džērbenes slimnīcā, kad tur atradies Džimītiņš. Viņas redzēju Kuldiģā pie Jansona kunga. Brauc arī DŽ. brāja meita Zenta un Melis ar dēlu.

2. attēls. Atšifrēšanas rīka saskarne *garamantas.lv*.

atrast pietiekamā skaitā. Pētnieciskajos rakstos uzsvars lielākoties likts uz dienasgrāmatu digitālu izdevumu izveidošanu. Piemēram, tiek reflektēts par dienasgrāmatu kodēšanu TEI XML formātā, anotējot atšifrējumā personas, vietas un citus konceptus (Thain 2016: 226–241). Tas digitālā izdevuma lasītājiem palīdz atklāt padziļinātu un bagātīgu kontekstu, kuru pētījuši un labi pārzina pētnieki, kas strādā pie šī projekta. Vēl kādā pētījumā datormetodes ir piemērotas (auto)biogrāfisko avotu pārveidošanai semantiski un lingvistiski anotētā korpusā, lai, izmantojot tīmekļa ontoloģijas valodu (*web ontology language*), varētu labāk izpētīt un interpretēt tekstā ietvertās zināšanu struktūras (Tóth 2013: 432–443). Savukārt kādā salīdzinoši nesenākā pētījumā ir izstrādāta jauna sistēma naratīvo pavadieņu datorizētai identificēšanai dienasgrāmatām līdzīgiem tiešsaistes blogiem, izmantojot vairākus dabiskās valodas apstrādes paņēmienus (Bandeli et al. 2020: 63–69). Adelaidas Universitātes pētnieku grupa veidojusi I pasaules kara dienasgrāmatu korpusu, kur iekļautas vairāk nekā 500 dienasgrāmatas, kas aptver laiku no 1914. gada augusta līdz 1918. gada novembrim. Piemērojot vairākas tāllasi-juma metodes, kā tēmu modelēšana (*topic modelling*) un sentimenta analīze, autori pētījumā

sniedz vispārīnātu ieskatu, parādot tematiskās tendences un atsevišķu jēdzienu izplatību korpusā (Dennis-Henderson et al. 2020: 90–104). Līdz šim nav izdevies atrast pētījumus, kuros būtu analizēta vispusīga dienasgrāmatu tekstu korpusa veidošana vai arī kur šāds – vispārīgs un eventuāli neviendabīgs korpus – būtu izmantots digitālo humanitāro zinātņu pētniecībā.

Pārlūkojot LFK Autobiogrāfiju krājumu, kas ir šī korpusa pamatā, redzam, ka iesniegtās un digitalizētās dienasgrāmatas ir stilistikas un formas ziņā samērā atšķirīgi materiāli, ko, protams, varēja paredzēt (Jackson 2010: 1–17; Lejeune 2009: 213–231). Autori kopumā dienasgrāmatās reflektējuši gandrīz visa 20. gadsimta garumā. Rakstītāji ir gan vīrieši, gan sievietes dažādos dzīves periodos, kā arī ir dienasgrāmatas, kuras rakstījuši vairāki autori, un to veikums nereti nav nošķirams. Dienasgrāmatām ir arī atšķirīga forma. Mēdz būt gan kalendāra veida dienasgrāmatas, kur līdztekus datumam fiksēts pavisam lakonisks dienas notikumu kopsavilkums, gan arī dienasgrāmatas, kur par ikdienu reflektēts daudz plašāk. Taču ir arī dienasgrāmatas, kurās šis nošķīrums nav tik skaidrs – autors mēdz rakstīt pavisam strupī, tad kādā brīdī ieraksti kļūst garāki, citkārt atkal pievienojas arī cita veida teksti, piemēram, dzeja, lugu vai scenāriju fragmenti, lūgšanas, darba piezīmes vai lasīto grāmatu citēšana.

Atšķirīga ir arī valoda, ko atrodam krājuma materiālos. Dienasgrāmatās uzskatāmi redzam valodas attīstību gadsimta garumā – gan leksikas, gan sintakses ziņā. Tā kā LFK Autobiogrāfiju krājumā tiek iekļautas npublicētas un nerediģētas dienasgrāmatas, tad dienasgrāmatās var sastapt nestandardizētu leksiku, netipiskas gramatiskās formas, senvārdus, izlokšņu vārdus, pierakstus no mūsdienu rakstu tradīcijas atšķirīgā ortogrāfijā, kā arī daudz gramatisko kļūdu un dažādu ortogrāfijas principu lietojumu. Dienasgrāmatas ir kultūrvēsturisks fenomens, kurš nav iegrožojams noteiktu formālu kritēriju standartos. Tas arī padara dienasgrāmatu korpusa izveidošanas procesu metodoloģisku sarežģītību bagātu.

Lai dienasgrāmatu tekstu korpusu tā veidošanas sākumposmā padarītu cik iespējami prognozējamu un pārskatāmu, un tādu, kam varētu piemērot vienotu metodoloģisko analīzi, izmēģinājuma korpusā iekļauta tikai daļa no LFK Autobiogrāfiju krājumā iesniegtajām dienasgrāmatu kolekcijām. Protī, raksts koncentrējas uz tā sauktajām stāstošajām dienasgrāmatām, kur autori par savu ikdienu reflektē izvērstākā formā. Pieņemot, ka viens dienasgrāmatas ieraksts ir relatīvi patstāvīga vienība, garāka naratīvā refleksija par ikdienu var sniegt dziļāku ieskatu ne tikai autoru personībās, bet arī izklāstītajās tēmās un naratīva veidošanas paņēmienos. Tomēr jāpiebilst, ka, lai saglabātu autoru dzīves pierakstu integritāti, korpusā dienasgrāmatas tiek iekļautas pilnībā. Ja kādā no tām kādā brīdī parādās lakoniski ieraksti, arī tie tiek iekļauti korpusā.

2021. gada februārī latviešu dienasgrāmatu pilotkorpusu veidoja 270¹¹ dienasgrāmatas, kas kopā veidoja 17 560 000 rakstzīmju jeb 2 830 000 vārdlietojumu. Vidējais dienasgrāmatu ieraksta garums korpusā ir 1079 rakstzīmes. Pilotkorpus, protams, aizvien nav viendabīgs. Atsevišķas dienasgrāmatas, kas rakstītas ļoti ilgu laiku jeb t. s. garās dienasgrāmatas,

11 Dienasgrāmatu skaits pārsniedz Autobiogrāfiju krājuma kolekciju skaitu, jo reizēm kolekcijās iekļauts vairāk par vienu dienasgrāmatu, piemēram, Pandēmijas dienasgrāmatu kolekcijā.

procentuāli aizņem lielu dienasgrāmatu korpusa daļu. Nenoteiktais un savstarpēji ļoti atšķirīgais apjoms ir dienasgrāmatu īpatnība, kas jāņem vērā, veidojot šādu korpusu. Līdzsvarotāka situācija varētu veidoties, ja tiek veidots dienasgrāmatu korpus vienam laika posmam vai izmantotas noteikta veida dienasgrāmatas, piemēram, kara dienasgrāmata (Dennis-Henderson et al. 2020: 90–104).

Lai izveidoto pilotkorpusu varētu izmantot dienasgrāmatu tekstu diahroniskai analīzei un veikt dažāda veida salīdzinošos mērījumus starp autoriem, dzimumiem vai vecumgrupām, nepieciešams dienasgrāmatu tekstus saskaldīt atsevišķu dienu ierakstu failos, kas vēlāk ļautu tos kombinēt pēc dažādiem kritērijiem, piemēram, pēc desmitgadēm vai autoru vecumposma utt. Proti, tas ir kā veidot vienu kopējo dienasgrāmatu, kur visu autoru rakstītais būtu vienkop sakārtots hronoloģiskā secībā. Šī uzdevuma veikšanu sarežģī dažāda datumu pieraksts dienasgrāmatās¹².

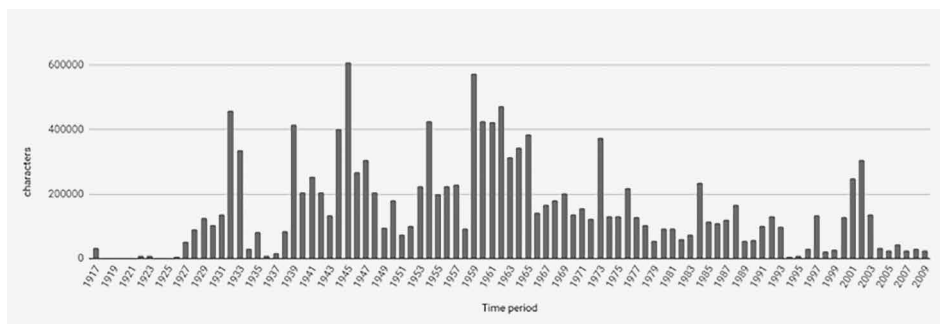
Veicot datumu pierakstu analīzi dienasgrāmatu korpusā ar mērķi noteikt datumu pieraksta sistēmu katrā atsevišķā dienasgrāmatā, atklājies, ka datumu pierakstīšana mēdz būt itin radošs process, kur liela nozīme ir autora gaumei, ieradumam un, iespējams, arī emocionālajam noskaņojumam. Datums var būt pierakstīts dienas ieraksta sākumā vai beigās, dažreiz arī pa vidu, bet citkārt ieraksts var būt arī bez datuma norādes, jo tas noprotams kontekstuāli no tekstā vēstītā, vai arī datuma vietā ir dota norāde uz svētkiem, kas tajā dienā svēti. Ir autori, kuri gadskaitli min tikai gada sākumā, bet mēnesi norāda mēneša sākumā, dienas numurējot ar skaitļiem. Tiek izmantoti arī ļoti dažādi gada un mēnešu saīsinājumi, arī pieturzīmju (punkts, komats, slīpsvītra, kols, semikols, defīse) lietojums ir daudzveidīgs, un visbeidzot mēdz būt arī kļūdaini norādīti mēneši, dienas vai gadi. Datumu ainava dienasgrāmatās ir patiešām daudzveidīgi krāšņa.

Piemēram, Dāvja Dauvarta dienasgrāmatā, kas rakstīta no 1932. līdz 1944. gadam, ir atrodamī 15 dažādi datuma pieraksti.

31. 12. 31.	3/I. 33.	2. jūlijā 1933. g.	7. septembra naktī	1933. g. 10. Septemb.	1933. g. 5. Septembrī
š. g. 11. septm.	33. g. 31. oktobrī	1933. gadā 1. janvārī	1934. 2. janvārī	1935. g.	31/IV. 36. g.
12. aug. 1938. g	6. jūlijā 1941. g.	1. 3. 44. g.			

1. tabula. Datumu pieraksta veidi Dāvja Dauvarta dienasgrāmatā (1889–1944),
LFK Ak145 (Dauvarts 1932–1944).

- 12 Protams, būtu bijis vēlams, ka datumu vienādošana notiek tajā brīdī, kad dienasgrāmatas tiek atšifrētas, piemēram, izmantojot kalendāra rīku datumu anotēšanai (*calendar tool for date annotation*), taču pilotkorpusa veidošanas brīdī šāda iespēja vēl nebija izstrādāta.



3. attēls. Pilotkorpūsā iekļauto dienasgrāmatu laika līnija pēc rakstzīmju skaita.

(Datu analīze veikta *Jupyter Notebook* lietojumprogrammā, izmantojot Python programmēšanas valodu, Pandas pakotni datu pārveidošanai un tīrīšanai, kā arī *Matplotlib* un *Plotly* pakotnes datu vizualizācijai. Daļa grafiku veidota, izmantojot *Google Sheets*.)

Datumu, kas norāda, kur sākas dienas ieraksts, varētu dēvēt par metadatumu. Tomēr datumi dienasgrāmatās izmantoti ne tikai kā metadatumi – gadi, mēneši, datumi, nedēļas, dienas un cita veida laika norādes (šodien, vakar, pērn utt.) ir bieži atrodamas arī pašā tekstā.

Tādējādi, lai izveidotu automatizētu rīku, kas atrastu visus metadatumus, tā ļaujot teksta dokumentu automātiski sadalīt pa dienu ierakstiem, bija veicami divi uzdevumi: (1) atrast pilnīgi visus datumus korpūsā; (2) efektīvi atšķirt un nošķirt metadatumus no cita veida datumu pieminējumiem tekstā, kā arī no citām ciparu un burtu kombinācijām tekstā, kas izskatās līdzīgi datumu pierakstam, bet nav datumi.

Lai labāk izprastu dienasgrāmatu tekstos atrodamo metadatumu pierakstu, tos nosacīti varētu iedalīt divās atšķirīgās grupās, proti, absolūtie un relatīvie metadatumi. Absolūtā metadatumā pieraksts būtu, piemēram, 14.02.1957. vai 14.II.57. Savukārt relatīvie metadatumi ir izsecināmi kontekstuāli un no attiecīgā ieraksta pozīcijas kopējā dienasgrāmatas struktūrā. Tas nesagādā grūtības pētnieka veiktā tuvlasījumā, taču krietni apgrūtina automatizētu datuma noteikšanu. Piemēram, ja dienasgrāmatas tekstā attiecīgās dienas ieraksts sākas ar *14*), tad automatizētajam datumu atpazīšanas rīkam ir jāspēj saprast vai atcerēties konteksts, ka vairākas lappuses iepriekš norādīts *februāris*, bet vēl iepriekš tekstā – *1957*. Tādējādi no kopējās teksta struktūras, datumu secības un konteksta ir secināms, ka attiecīgais metadatumus *14*) apzīmē 1957. gada 14. februāri.

Visbiežāk tomēr metadatumu izvietojums dienasgrāmatu tekstos ir šāds: tukša rinda pirms jaunas dienas ieraksta, metadatumus – jaunā rindā, attiecīgās dienas ieraksts – jaunā rindā. Lai pareizi veiktu atsevišķo dienasgrāmatu priekšapstrādi teksta sadalīšanai atsevišķu dienu ierakstos, būtiski ņemt vērā katrā izmantoto datēšanas sistēmu un attiecīgi pielāgot metadatumu atpazīšanas rīku. Taču kopumā rezultāts, veicot izlases veida manuālu pārbaudi, bija teicams – korpuss bija pareizi

sadalīts pa dienas ierakstiem. Kopumā pilot-korpusu veido 16 904 dienas pieraksti, kas aptver laika posmu no 1917. līdz 2021. gadam.

Iegūstot šādas – pa dienas ierakstiem sadalītas – dienasgrāmatas, paveras jaunas iespējas analizēt to, kā šis pēc nejaušības principa veidotais tekstu korpus izskatās dažādos šķērs griezumos. 3. attēlā redzams, kā dienasgrāmatu ieraksti pēc zīmju skaita izkārtojas cauri gadiem. Korpusā iekļautie teksti dokumentē laiku no 1917. gada līdz pat mūsdienām. Kaut arī dienasgrāmatu pierakstos nosepts viss redzamais laikposms, tomēr dokumentējums nav viendabīgs. Redzams, ka izteikti vairāk ir dienasgrāmatu ierakstu, kas tapuši ap Otrā pasaules kara beigām, un izceļas arī 20. gadsimta 60. gadu sākuma posms. Šajā grafikā gan nav integrētas 2020. un 2021. gadā pandēmijas laikā tapušās 239 dienasgrāmatas, jo to īpatsvars hronoloģiskā griezumā ir daudzkārt lielāks nekā pārējām dienasgrāmatām, kas šāda veida grafiku padarītu grūtāk uztveramu.

Sadalot dienasgrāmatu tekstu korpusu pa dienas ierakstiem, katram no tiem tiek piešķirti metadati, kuri norāda: (1) dienasgrāmatas kolekcijas numuru Autobiogrāfiju krājumā, kas ļauj atlasīt visus viena autora ierakstus; (2) ieraksta numuru pēc kārtas dienasgrāmatā, kas neļauj kļūdīties ar ierakstu secību, ja vairāki ieraksti tapuši vienā dienā; (3) metadatumu, kas norāda uz ieraksta tapšanas laiku – gan autora oriģinālajā pieraksta formā, gan pārveidots vienotā *sql* datubāzes datuma pierakstā; (4) dienasgrāmatas ieraksta rakstzīmju skaitu. Papildus tam ir izveidotas relācijas ar autoru metadatiem, kas savukārt ļauj veidot dienasgrāmatas ierakstu izlases jeb apakškorpusus pēc dažādiem kritērijiem, piemēram, pēc autora dzimumu vai/un vecuma posma, kā arī pēc noteikta vecuma posma noteiktā laika periodā.

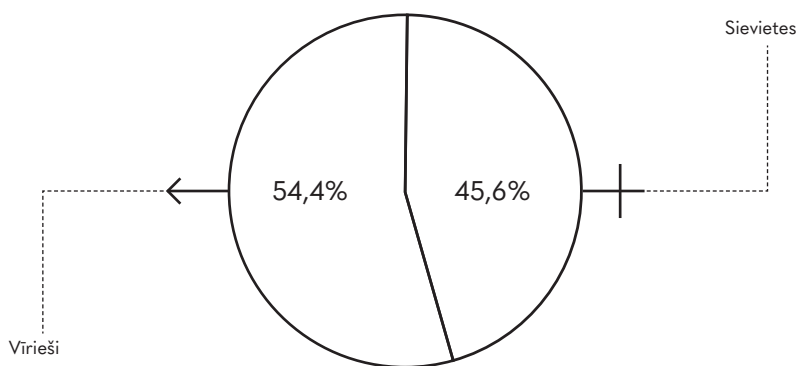
```
"lfk_numurs": "LFAK26ds",
"ieraksta_nr_pie_si_autora": 2,
"metadatum": "<08.08.1926.>",
"sql_datums_ierakstam": "1926-08-08",
"zīmju_skaitis": 658,
"dienas_teksts": "8.VIII.26. Tīrzas baznīcas pastāvēšanas 100 g. jubileja. Sprendīkoja bij. Tīrzas māc. tagad Pāvila dr. mc. Adolfs Kundziņš, prof. V. Maldonis, Gulbenes māc. J. Ozols, kas kā viskārs apkalpo Tīrzas draudzi. Satiku daudz vecu pazīstamu – Zēmiņu ļaudis, Rankas Ruti, Sebra Kārlis u. c. pēc dievkalpojuma bija garīgo koncerts. Dziedāja Tīrzas un Odulienes koris vecā skolotāja un ērģelnieka K. Dzelzkalna vadībā. Kā soliste piedalījās Amanda Liberte-Rebāne. Liberts palīdzējis gresnot Tīrzas baznīcu. I. ģimī Tīrzas muižā, kur tēvs bijis bruža pagraba meistars, māte vecā mežsarga Valvīna Rāča meita. No baznīcas mājā braucot salijām.",
```

4. attēls. Vienas dienas ieraksts datubāzē.

Vīriešu un sieviešu dienasgrāmatas

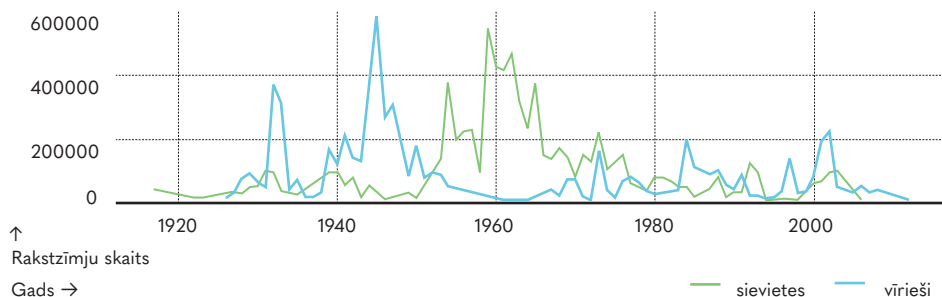
Dzimumu reprezentācija dienasgrāmatu korpusā varētu būt viens no svarīgākajiem jautājumiem, jo īpaši, lai noskaidrotu, vai ir iespējams izveidot divus līdzvērtīgus tekstu apakškorpusus – sieviešu un vīriešu –, lai tos salīdzinoši pētītu, izmantojot tāllasījuma metodes. Turklāt, tā kā dienasgrāmatu tekstu korpusi ir piesaistīts laikam, ir svarīgi arī noskaidrot, kā abi dzimumi ir pārstāvēti dažādos laika posmos. Mēs apzināmies, ka dzimumu jautājums nav viennozīmīgs un vienkāršs un ka eventuāli varētu tikt veidoti arī citi dalījumi apakškorpusos, ņemot vērā dienasgrāmatu autoru pašidentifikāciju. Šajā korpusā balstīto nākotnes pētījumu mērķis varētu būt noskaidrot, vai formālā ziņā ir pamats nošķirt sieviešu un vīriešu dienasgrāmatas. Varbūt tematiskā, emocionālā un stilistiskā līdzība nav dzimuma determinēta, bet būtisks ir laika periods, vecuma posms vai kāds cits kritērijs?

Analizējot korpusa metadatus dzimumu perspektīvā, no tā tika izņemtas ārā tās dienasgrāmatas, kuras ir rakstījuši kolektīvi dažādu dzimumu pārstāvji, un arī tās, kuru autoru dzimums nav zināms.



5. attēls. Sieviešu un vīriešu rakstījumu reprezentācija korpusā pēc rakstzīmju skaita.

Pārlūkojot korpusā iekļautos tekstus vispārīgi, ir redzama diezgan līdzsvarota aina. Vīriešu rakstītais tekstuālais apjoms ir tikai neliels lielāks nekā sievietēm (sk. 5. attēlu). Tomēr šāds vienkāršots skatījums neļauj apjaust to, kā korpusi ir veidojies dzimumu perspektīvā attiecībā pret laiku. Proti, cik līdzsvaroti vīriešu un sieviešu rakstītais ir pārstāvēts dažādos laika posmos. Vai varam runāt par kādu periodu, kurā pārliecinoši naratīvā dominē viens vai otrs dzimums? Ņemot vērā to, ka dienasgrāmatas nav standartizētas publikācijas, kuru izdošanas biežumu un apjomu varētu iepriekš prognozēt, tad detalizētāku ainu par to, kā izraudzīto dienasgrāmatu ieraksti izskatās laika līnijā, mēs ieraugām tikai tad, kad ir veikta padziļinātāka materiālu apstrāde un analīze.



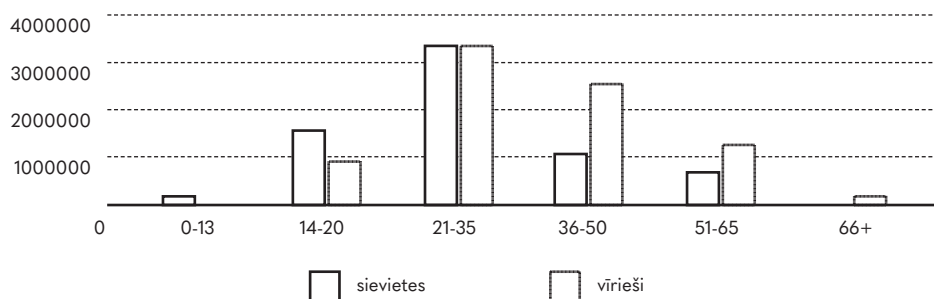
6. attēls. Sieviešu un vīriešu rakstīšanas intensitāte autobiogrāfisko tekstu korpusā.

6. attēls vizuāli attēlo to, kā latviešu dienasgrāmatu tekstu pilotkorpusā ir reprezentētas vīriešu un sieviešu dienasgrāmatas noteiktos laika periodos. Kaut arī, veidojot korpusu un strādājot ar kolekcijām, šķita, ka situācija šajā ziņā varētu būt samērā līdzsvarota, tomēr metadatu analīze parāda interesantu ainu. Uzskatāmi redzams, ka laiks no 1940. līdz pat 50. gadu sākumam dienasgrāmatās dokumentēts no vīriešu perspektīvas. Jāpiebilst, ka krājumā un arī tekstu korpusā nav mērķtiecīgi meklētas un iekļautas karavīru dienasgrāmatas, tādas atrodas tikai dažas (Sīlis 1943–1946; Upmalis 1944–1946).

Sieviešu naratīvās balsis sāk dominēt 50. gadu sākumā, un to dominance turpinās divas desmitgades. Vēlāk situācija kļūst līdzsvarotāka, bet būtiski samazinās arī pilotkorpusa tekstu apjoms. Šāda izteikti viena dzimuma dominēšana ilgākā laika periodā kopējai korpusa sabalansētībai ir nevēlama, un jebkuri vispārinoši secinājumi par korpusu kopumā vai noteiktā laika posmā ir jāizdara ar piesardzību un paturot prātā tā formālās īpatnības. Taču no pagaidu korpusa padziļinātākai pētniecībai pamatoti varētu izvēlēties vīriešu dienasgrāmatas, kas tapušas ap Otrā pasaules kara laiku, un sieviešu dienasgrāmatas padomju okupācijas pirmajās desmitgadēs.

Vecumgrupu rakstījumu pārstāvēniecība pilotkorpusā

Viens no vērtīgākajiem ieguvumiem korpusa metadatu analīzē ir iespēja izveidot relāciju starp autora vecumu un rakstīšanas vēsturisko laiku. Respektīvi, ja 1960. gadā autoram ir 10 gadu, tad viņš pārstāv vienu vecuma grupu. Bet, ja šim pašam autoram 1982. gadā ir 32 gadi, tad viņa rakstītais jau ir iekļauts citā vecuma grupā. Šāda informācija ļauj pārlūkot korpusu tādā perspektīvā, kādu bez datormetožu izmantojuma būtu ārkārtīgi sarežģīti īstenot. Balstoties uz šādu analīzi, ir iespējams pamatotāk izraudzīties tekstu analīzei piemērotākās vecuma grupas konkrētos laika periodos.



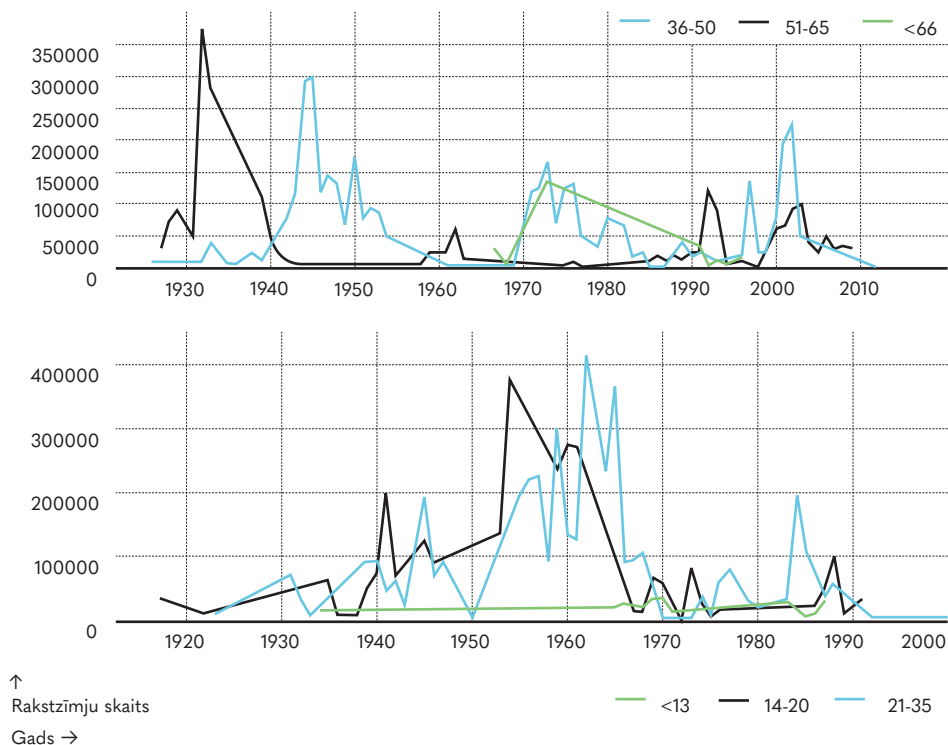
7. attēls. Sieviešu un vīriešu rakstījumu apjoma salīdzinājums pēc vecumgrupām.

Dienasgrāmatu tekstu pilotkorpusā sākotnēji izraudzīts sadalīt dienasgrāmatu ierakstus pa desmitgadēm. Proti, jauna vecuma grupa tiek veidota ikkatrus desmit gadus (< 10 g. v., no 11 līdz 20 g. v. utt.). Pārlūkojot korpusu šādā vecumgrupu sadalījumā, tika secināts, ka šī pieeja nav pamatota. Liela daļa vecumgrupu pilotkorpusā bija ļoti vāji reprezentētas, un kopaina bija pārāk sadrumstalota, lai sniegtu jēlkādu papildu informāciju. Ņemot to vērā, dienasgrāmatu tekstu pilotkorpusā tika noteiktas sešas vecuma grupas, par pamatu ņemot vispārīgus sociālos, izglītības un ekonomiskos aspektus, kas nosaka cilvēka dzīves posmus. Pirmā grupa: bērni līdz 13 gadu vecumam. Otrā grupa: jaunieši no 14 līdz 20 gadu vecumam. Trešā grupa: jaunie pieaugušie no 21 līdz 35 gadu vecumam. Ceturtā grupa: pieaugušie I no 35 līdz 50 gadu vecumam. Piektā grupa: pieaugušie II no 51 līdz 65 gadu vecumam. Sestā grupa: seniori virs 66 gadu vecuma. Protams, arī šāds iedalījums nav viennozīmīgs, jo patstāvīgas dzīves sākums dažādos laikmetos varētu būt atšķirīgs. Veicot tekstu analīzi, iespējams, vecuma grupas nāksies koriģēt.

Pārlūkojot šādu vecumgrupu sadalījumu dzimumu perspektīvā, redzam, ka bērnu un senioru vecumgrupa ir ārkārtīgi vāji pārstāvēta korpusā un būtu apsverams, vai šīs abas grupas nebūtu lietderīgi apvienot attiecīgi ar 3. grupu (jaunieši) un 5. grupu (pieaugušie II). Līdzsvarota situācija dzimumu perspektīvā vērojama 21–35 gadu vecu cilvēku vecumgrupā, savukārt pārējās dominē viens vai otrs dzimums.

Attēlojot vecumgrupu rakstījumu intensitāti uz vienas laika līnijas (7. attēls), iegūstam papildu informāciju par to, kas jāņem vērā, veicot turpmāku analīzi.

8. un 9. attēls parāda, ka bērnu dienasgrāmatas korpusā ir izkaisītas cauri gadsimtam vienmērīgi mazā apjomā. Taču citu vecumgrupu reprezentācija uz laika līnijas iezīmē noteiktus periodus, kuros izpēte varētu būt sekmīgāka, jo pieejams vairāk materiāla. Piemēram, jauniešu grupas (14–20 g. v.) dienasgrāmatu ieraksti lielā intensitātē tapuši II pasaules kara laikā, bet visvairāk no 1955. līdz 1965. gadam.



8. un 9. attēls. Korpusa sadalījums laikā pēc vecuma grupām.

Jauno pieaugušo (21–35 g. v.) vecumgrupas intensīvākais rakstīšanas laiks korpusā ir no 50. gadu vidus līdz 60. gadu beigām, ar atsevišķiem intensitātes punktiem ap Otrā pasaules kara beigām un 80. gadu vidū. Turklāt, atsaucoties uz 6. attēlu, šī ir arī grupa, kur abi dzimumi ir vienlīdz labi pārstāvēti. Pieaugušo grupa I (36–50 g. v.), kur dominē vīriešu naratīvās balsis, vislabāk reprezentēta ap Otrā pasaules kara laiku līdz pat 50. gadu sākumam. Otrs intensīvākais rakstīšanas posms šai vecuma grupai ir 70. gadi un arī – ap tūkstošgades miju. Pieaugušo grupai II (51–65 g. v.) ir izteikta rakstīšanas intensitāte 30. gados, kā arī ap 1990. gadu, savukārt pārējās desmitgadēs tā ir nemanāma. Savukārt senioru grupa ir maz reprezentēta līdz pat 60. gadu beigām. Pārējo vecumgrupu kontekstā senioru rakstītais veido būtisku pilotkorpusa daļu no 1970. līdz aptuveni 1990. gadam, kas varētu liecināt, ka šīs grupas apvienošana ar pieaugušo grupu II varētu nebūt pamatota.

Atgriežoties pie 3. attēla, kur labi redzams, ka 2. pasaules kara beigas ir visintensīvāk dokumentētais laikmets dienasgrāmatu pilotkorpusā, šajos – 8. un 9. attēlā – redzam, ka autori,

kas rakstījuši ap 1945. gadu, pārstāv trīs vecuma grupas – jauniešus, jaunos pieaugušos un pieaugušo I grupu. Un, salīdzinot šos ar 6. attēlu, redzam, ka šajā laikā dominē vīriešu dzīves pieraksti. Respektīvi, padziļināti analizējot šo laika posmu, būtu jāņem vērā, ka II pasaules kara laiku reprezentē salīdzinoši jaunu vīriešu naratīvās balsis.

Lai gan aina joprojām izskatās neviendabīga un sadrumstalota, jāņem vērā, ka metadati tiek analizēti trijās dimensijās – dzimums, vecumposms un laikmets –, kas katra atsevišķi saskalda sīkāk aplūkojamo korpusa materiālu. Tomēr šāda analīze labi parāda katras vecuma grupas pārstāvību korpusā dažādos laika posmos. Tas ir noderīgi, lai izvēlētos nākamsoļus gan tālās lasījuma, gan kvalitatīvās analīzes veikšanai. Papildus būtu vērts arī sīkāk izpētīt, kā veidojas šie intensitātes periodi, t. i., cik lielu īpatsvaru grupas kopējā pārstāvībā sniedz atsevišķu autoru rakstījumi. Jāņem vērā, ka intensīvi rakstošam autoram varētu būt ievērojama ietekme uz kādas vecuma grupas pārstāvniecību attiecīgajā periodā.

Secinājumi

Metadatu analīzē balstīti mērījumi, kuri veikti, pētot dienasgrāmatu tekstu pilotkorpusu, ir izmantojami secinājumiem par korpusu vai kolekciju, tie var iezīmēt eventuālas tendences vispārējos dienasgrāmatu rakstīšanas paradumos, tomēr ar šiem datiem vien nepietiek, lai izdarītu pamatotus secinājumus par dienasgrāmatu rakstīšanas tradīciju un tendencēm Latvijā.

Vispārēja dienasgrāmatu tekstu korpusa veidošana kopumā ir sarežģīts projekts. Ideālā gadījumā varētu vēlēties redzēt vienlīdz labi pārstāvētus abus dzimumus un vienlīdz labi pārstāvētus abus dzimumus visās vecuma grupās. Mēs varētu arī vēlēties vienlīdz labi pārstāvētas visas vecuma grupas kopumā dažādos laika posmos. Tomēr jāatminas dienasgrāmatu tekstu korpusa veidošanas īpatnības. Proti, dienasgrāmatu apjoms, forma, veids un arī autoru dzimumpiederība vai vecums nav prognozējams un ietekmējams. Veidojot kolekciju un attiecīgi arī korpusu, ir iespēja mērķtiecīgāk meklēt dienasgrāmatas pēc noteiktiem kritērijiem, taču lielākoties izvēle nav liela un nākas rēķināties ar to, kas ir pieejams arhīvos, bibliotēkās vai personīgajās kolekcijās. Šis raksts pievēršas vien autoru dzimumam un vecumam, nediskutējot par izglītības līmeni, dzīvesvietu vai sociālo stāvokli. Bieži vien šāda informācija par dienasgrāmatu autoriem nav zināma, taču iespējamo kritēriju skaits, pēc kādiem sadalīt dienasgrāmatas apakškorpusos un kādā perspektīvā pētīt, var būt krietni liels.

Raugoties no kultūrvēsturiskās perspektīvas, būtu nepieciešams saglabāt un datorizēt visas pieejamās dienasgrāmatas. Savukārt turpmākajā dienasgrāmatu korpusa attīstībā drīzāk būtu jāturpina fokusēties uz eventuāli apjomīgākajiem apakškorpusiem un arī tiem apakškorpusiem, kuri pētniekiem ir nepieciešami konkrētu pētniecisko tēmu izziņāšanai.

Kaut arī dienasgrāmatu tekstu pilotkorpusa apjoms šobrīd nav mazs, tomēr, sastatot tekstus attiecībā pret laiku (gadu, desmitgadi), kurā fiksēts dienasgrāmatas ieraksts, laikmeta

pārklājums ir samērā plāns. Ir gadi, kurus pārstāv līdz pat desmit autoriem, bet ir periodi, kad rakstītāju skaits ir visai niecīgs. Autoru skaita palielināšana varētu būt vēl viens dienasgrāmatu teksta korpusa nākotnes uzdevums. Tiek plānots iekļaut pilnīgi visas digitalizētās, ne tikai stāstošās dienasgrāmatas, metadatos papildus norādot dienas ieraksta garumu, lai turpmākajos izmēģinājumos ar teksta analīzi ierakstus ar noteiktu apjoma ierobežojumu varētu definēt pēc nepieciešamības.

- Dauvarts, Dāvis (1932–1944). Dāvja Dauvarta dienasgrāmata. LFK AK 145. Pieejams: <http://garamantas.lv/lv/collection/1362666/Davja-Dauvarta-dienasgramata> [skatīts 15.10.2021.].
- Dennis-Henderson, Ashley, Roughan, Matthew, Mitchell, Lewis, Tuke, Jonathan (2020). Life Still Goes on: Analysing Australian WW1 Diaries through Distant Reading. *Proceedings of the 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. International Committee on Computational Linguistics, pp. 90–104.
- Hockey, Susan (2004). The history of humanities computing. Schreibman, Susan, Siemens, Raya Unsworth, John (eds.). *A Companion to Digital Humanities*. Oxford: Blackwell, pp. 1–19.
- Jackson, Anna (2010). *Diary Poetics. Form and Style in Writers' Diaries, 1915–1962*. New York: Routledge.
- Jurča, Ērika (1944). Ērikas Jurčas dienasgrāmata. LFK AK 15. Pieejams: <http://garamantas.lv/lv/collection/1134078/Erikas-Jurcas-dienasgramata> [skatīts 15.10.2021.].
- Bandeli, Kumar Kiran, Hussain, Muhammed Nihal, Agarwal, Nitin. (2020). A Framework towards Computational Narrative Analysis on Blogs. *CEUR Workshop Proceedings*, Vol. 2593, pp. 63–69.
- Koncar, Philipp, Fuchs, Alexandra, Hobisch, Elisabeth, Geiger, Bernhard, Scholger, Martina, Helic, Denis (2020). Text sentiment in the Age of Enlightenment: an analysis of spectator periodicals. *Applied Network Science*, No. 5(1).
- Lejeune, Philippe (2009). *On Diary*. Honolulu: University of Hawaii Press.
- Lipša, Ineta (2021). Documenting the queer self: Kaspars Aleksandrs Irbe (1906–1996) in between unofficial sexual knowledge and medical-legal regulation in Soviet Latvia. *Cahiers du monde russe*, No. 62/2-3, pp. 415–442.
- Moretti, Franco (2013). *Distant Reading*. London: Verso.
- Nyhan, Julianne, Flinn, Andrew (2016). *Computation and the Humanities: towards an Oral History of Digital Humanities*. Springer Open.
- Rak, Julie (2019). Big Data and Self-Tracking: Research Trajectories. Barnwell, Ashley, Douglas, Kate (eds.). *Research Methodologies for Auto/biography Studies*. London: Routledge, pp. 116–121.
- Reinsone, Sanita (2018). Participatory practices and tradition archives. Harvilahti, Lauri, Kjus Audun, Cliona, O'Carroll, Österlund-Pörtzsch, Sussane, Skott, Fredrik, Treija, Rita (eds.). *Visions and Traditions: Knowledge Production and Tradition Archives*. Helsinki: SKS Academia Scientiarum Fennica, pp. 279–296.
- Reinsone, Sanita (2020). Searching for deeper meanings in cultural heritage crowdsourcing. Hetland Per, Pierroux, Palmyre, Esborg, Line (eds.). *A History of Participation in Museums and Archives. Traversing Citizen Science and Citizen Humanities*. London: Routledge, pp. 186–207.
- Sula, Chris Alen, Hill, Heather (2019). The early history of digital humanities: An analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004). *Digital Scholarship in the Humanities*, No. 34, pp. 190–206.
- Silis, Kārlis Alberts (1943–1946). Kārļa Alberta Sila dienasgrāmata. LFK AK 174. Pieejams: <http://garamantas.lv/lv/collection/1458840/Karla-Alberta-Sila-dienasgramata> [skatīts 16.10.2021.].
- Thain, Marion (2016). Perspective: Digitizing the Diary – Experiments in Queer Encoding (A Retrospective and a Prospective). *Journal of Victorian Culture*, No. 21 (2), pp. 226–241.
- Tóth, Gábor Mihály (2013). The computer-assisted analysis of a medieval commonplace book and diary (MS Zibaldone quaresimale by giovanni rucellai). *Literary and Linguistic Computing*, No. 28(3), pp. 432–443.
- Upmalis, Laimonis (1944–1946). Laimoņa Upmaļa dienasgrāmata. LFK AK 148. Pieejams: <http://garamantas.lv/lv/collection/Laimona-Upmala-dienasgramata> [skatīts 16.10.2021.].

Metadata-based Analysis of the Diaries' Corpus

Sanita Reinsone, Haralds Matulis,
Ilze Ļaksa-Timinska

Keywords: digital humanities, distant reading, text analysis,
diary corpus, autobiographical texts

Personal diaries are a unique narrative document of an era, revealing the times and events through a deeply personal perspective. Each diary is undoubtedly unique, having been written in a particular place and time, encapsulating many of the authors' present-day experiences. However, together they create a special dimension of personal writing, where the individual style is shaped by the author's taste and writing skills, as well as by the influences of diaristic traditions, fiction and other cultural phenomena. Recognizing the uniqueness and particularity of each diary, in this paper we will focus on personal diaries as a whole and conduct a metadata-driven analysis of a pilot corpus of diaries with the aim of revealing the particularities of the corpus in the juxtaposition of different criteria.

This paper is based on recent experience of building a corpus of personal diary texts from the very beginning, i.e. starting with the collection, digitisation and further processing of the material to build the corpus. In this article, we will analyse the creation of the corpus of diary texts, the methodological challenges and the difficulties encountered in such an initiative.

The metadata-based measurements that we have taken in the pilot corpus of diary texts can be used to draw conclusions about a corpus or a collection, they can highlight possible trends in general diary-writing habits, but these data alone are not sufficient to draw valid conclusions about the tradition and trends of diary-writing in Latvia.

Liepājas Universitātes folkloras krājuma veidošanās, apstrāde un digitālā pieejamība

Ieva Ančevska

Raksts sagatavots projektā "Humanitāro zinātņu digitālie resursi: integrācija un attīstība"
(VPP līguma Nr. VPP-IZM-DH-2020/1-0001).

Raksta mērķis ir sniegt pārskatu par Liepājas Universitātes folkloras krājuma tapšanu dažādos laika periodos kopš augstskolas dibināšanas 1954. gadā. Augstskola kopš tā laika ir mainījusi gan nosaukumus, gan studiju virzienus un mērķus, bet folkloras materiāli ir vākti gandrīz katrā tās attīstības posmā līdz pat mūsdienām. Folkloras materiāli vākti gan studentu folkloras prakšu laikā, gan atsevišķos studiju priekšmetos, gan arī pētnieku organizētās ekspedīcijās jeb lauka pētījumos.

Rakstā apkopota informācija par dažādajiem folkloras materiālu vākšanas, uzglabāšanas un izmantošanas veidiem līdz pat mūsdienām, kā arī par cilvēkiem, kuri vadījuši šīs norises.

Atskats Liepājas Universitātes vēsturē

Liepājas Universitātes aizsākumi meklējami jau 1945. gadā, kad darbu sāka četrgadīga pedagoģiskā skola, kas 1954. gadā pārtapa skolotāju institūtā. Augstskolas sākotnējais nosaukums bija Liepājas Pedagoģiskais institūts (1954–1961), pēc tam nosaukums mainījās, tam pakāpeniski pievienojot vēl papildu apzīmējumus – Liepājas Valsts pedagoģiskais institūts (1961–1966) un Viļa Lāča Liepājas Valsts pedagoģiskais institūts (1966–1990). Deviņdesmito gadu sākumā (1990–1993) uz trim gadiem tika atjaunots pirmais nosaukums Liepājas Pedagoģiskais institūts, pēc tam uz pieciem gadiem institūts tika pārsaukts par augstskolu – Liepājas Pedagoģiskā augstskola (1993–1998), tad uz desmit gadiem tā kļuva par akadēmiju – Liepājas Pedagoģijas akadēmija (1998–2008), bet kopš 2008. gada tā ir pazīstama kā Liepājas Universitāte (LiepU vēsture).

Līdz deviņdesmito gadu sākumam augstskola sagatavoja galvenokārt sākumskolas skolotājus, bērnudārzu pedagogus, matemātikas skolotājus un latviešu valodas un literatūras skolotājus visai Latvijai. 1990. gadā Liepājas Pedagoģiskais institūts uzsāka pāreju no viena profila augstskolas uz reģionālu vairākprofilu augstskolu, un paralēli skolotāju izglītošanas studiju programmām tika izveidots plašs nepedagoģisko studiju programmu spektrs.

Liepājas Universitātes folkloras materiālu vākšana

Jau kopš 19. gadsimta beigām topošie skolotāji dažādās mācību iestādēs tika rosināti vākt folkloras materiālus. Kā pirmais jāpiemin Baltijas skolotāju seminārs, kā arī Rīgas un Jelgavas skolotāju institūti, kuru studenti savākuši ievērojamu daudzumu folkloras vienību, kas

glabājas Latviešu folkloras krātuvē. Arī Liepājas Pedagoģiskā institūta studenti prakšu laikā savākuši krietni daudz folkloras materiālu. Latviešu valodas skolotāju un filoloģijas programmu, kā arī dažu līdzīgu programmu studentiem augstskolās viena no obligātajām bija folkloras prakse. Visbiežāk tā bija jāveic patstāvīgi, pašam izvēloties vietu un teicējus, pašam organizējot savu darbu un kārtojot, atšifrējot savāktos materiālus.

Folkloras prakse lika studentiem iedziļināties folkloras kursā apgūtajā un izmantot zināšanas praktiski, tā rosināja pašiem iepazīt folkloras rašanās un pārņemšanas vidi un apstākļus. Materiālu vākšana ļāva piedzīvot dzīvo tradīciju un apzināties, kuri žanri joprojām tiek nodoti tālāk un kuri vēl veidojas un attīstās.

Ilggadējā Latviešu folkloras krātuves fondu glabātāja un studentu folkloras prakšu vadītāja Māra Viksna raksta: "Folkloras prakses galvenais mērķis ir dot studentam iespēju iepazīt dzīvo folkloru, saprast tās izplatību, funkcionēšanu, apgūt saskarsmes mākslu, iegūt materiālus, kas papildinātu esošos krājumus, dotu jaunas atziņas pētījumos." (Viksna 2002: 310)

Liepājā folkloras kurss un folkloras prakse bija obligāta studiju programmas daļa vairāku specialitāšu studentiem: pirmsskolas pedagoģijas, sākumskolas pedagoģijas, speciālās pedagoģijas studentiem, arī vēstures skolotāju programmas un kultūras vadības studentiem. Kādu laiku pastāvēja latviešu valodas, literatūras un svešvalodas skolotāju programma. Arī šīs programmas studentiem bija jāapgūst folkloras kurss un jāveic folkloras prakse.

Folkloras kursa docētāji augstskolas darbības laikā vairākkārt mainījušies. Par viņiem atrodamas visai skopas ziņas. Septiņdesmito gadu beigās folkloras kursu pasniedza un folkloras praksi studentiem vadīja docētāja Alma Šmite (1935–?), pēc tam kādus gadus folkloras lekciju kursu vadījis pasniedzējs Rihards Kārklis (1923–1991). Pēc viņa astoņdesmito gadu sākumā (1980–1983) folkloras kursu vadīja Latvijas Universitātes profesore *Dr. philol.* Ilze Stikāne (1957). Nākamā folkloras kursa pasniedzēja bija Vineta Trumsiņa (1957–2010). Viņa uzņēmās vadīt arī folkloras praksi. 1994. gadā paralēli Vinetai Trumsiņai folkloras kursu un prakses sāka vadīt arī Humanitāro un mākslas zinātņu fakultātes docētāja Gita Girņus (1970).

Liepājas studentu vāktie folkloras materiāli

Pirmie folkloras materiāli no Liepājas studentiem saņemti jau pirmajā institūta mācību gadā. Tos 1955. gadā iesniedzis students Emils Arājs no Alsungas, šai kolekcijai piešķirts numurs 1921 (Viksna 2002: 309–310).

Nākamais folkloras vākums, kas iesūtīts no Liepājas institūta, ir Benitas Laumanes tautasdziesmu vākums. Benita Laumane (1937) ir *Dr. habil. philol.*, valodniece, dialektoloģe, profesore, kura strādā Liepājas Universitātē kopš 1994. gada un aktīvi piedalās folkloras un dialektoloģijas ekspedīcijās. Vācot dialektoloģijas materiālus, viņa savākusi arī daudz folkloras materiālu.

Benita Laumane atceras, ka piecdesmitajos un sešdesmitajos gados Liepājas institūta folkloras ekspedīcijas bijušas apvienotas ar Zinātņu akadēmijas organizētajām ekspedīcijām. Kādu laiku apvienotas bijušas arī folkloras un dialektoloģijas prakses. Benitai Laumanei saglabājušās atmiņas par 1959. gada ekspedīciju, kas viņai kā studentei radījusi dziļāku interesi un devusi impulsu tālākiem lauka pētījumiem (Laumane 2021). Nākamajā gadā (1960) Benita Laumane ekspedīcijā uz Bārtu jau devusies patstāvīgi un līdzās dialektoloģijas materiāliem savākusi arī tautasdziesmas. Tās viņa iesūtījusi folkloras krātuvei, un šim vākumam piešķirts kolekcijas numurs LFK [1947]. Vākumā ir 230 Bārtas pagastā pierakstītas tautasdziesmas.

Folkloras pētniecei Ritai Drīzulei adresētā vēstulē, nododot savu vākumu, Benita Laumane norāda vairāku Bārtas teicēju vārdus, kuriem ir bagāts tautasdziesmu pūrs un kurus būtu ieteicams apmeklēt, piemēram, Bārtas etnogrāfiskā ansambļa vadītājs Jēkabs Ķiburs, Margrieta Pūšēna un Kūmas māte, “kas ir jau deviņtā gadu desmitā. Agrāk bijusi bērnu saņēmēja un audēja” (LFK 1947, 2). Kopumā par Bārtā saglabāto folkloras mantojumu viņa raksta: “Bārtā var savākt bagātīgu tautas dziesmu pūru.” (LFK 1947, 2)

Par turpmākajām Liepājas studentu folkloras praksēm un ekspedīcijām ir atrodamas tikai fragmentāras ziņas līdz pat 1975. gadam, kad studentu folkloras prakses sāka vadīt Māra Viksna (1949), *Dr. philol. b. c.*, folkloras pētniece, LU LFMI Latviešu folkloras krātuves fondu glabātāja. Viņa pētījusi folkloras vākšanu un folkloristikas vēsturi, daudz rakstījusi par vācējiem un teicējiem. Māra Viksna aktīvi iesaistījusies arī studentu izglītošanā, bijusi folkloras prakšu vadītāja vairāku augstskolu studentiem: LU Filoloģijas fakultātē, Latvijas Kultūras akadēmijā, Jāzepa Vītola Latvijas Mūzikas akadēmijā un kādreizējā Viļa Lāča Liepājas Valsts pedagoģiskajā institūtā. Liepājas studentiem Māra Viksna vadījusi folkloras prakses no 1974. līdz 1989. gadam (Literatura.lv; Viksna 2021).

Māra Viksna atceras, ka no 1974. gada viņa reizi gadā devusies uz Liepāju, kur sniegusi pārskatu par padarīto un devusi ievirzi nākamajiem folkloras materiālu vācējiem. Tajā laikā katrs students ir strādājis savās mājās vai pie radiem un pamazām sākusi veidoties Liepājas kolekcija LFK [2011], kurā apkopotas 24 075 folkloras vienības¹ (Viksna 2002: 311).

Māra Viksna norāda, ka tolaik Liepājā “par praksi rūpējās un gādāja profesore Antoņina Reķēna. Laikam viņa saskatīja dialektoloģijas un folkloristikas saskari.” (Viksna 2021)

Arī citu laikabiedru liecības vēsta, ka līdz pat deviņdesmito gadu sākumam folkloras materiālu vākšanu Liepājā galvenokārt organizēja nevis folkloras pasniedzēji, bet gan valodniece *Dr. habil. philol.* Antoņina Reķēna (1928–1998). Viņa strādāja Liepājas Pedagoģijas akadēmijā no 1953. līdz 1998. gadam. Reķēna bija sevišķi ieinteresēta dialektoloģijas materiālu vākšanā, kas veidoja būtisku pamatu viņas pētījumiem dialektoloģijā un areālajā lingvistikā (Medveckis 2007: 123–124).

1 Sarakstē Māra Viksna norāda, ka šajā kolekcijā ir 25 084 folkloras vienības.

Māras Vīksnas vadīto folkloras prakšu materiāli nodoti Latviešu folkloras krātuves glabāšanā, bet vēlākajos gados vāktie materiāli glabājas Liepājas Universitātes folkloras arhīvā.

Latviešu valodas un literatūras pasniedzēja, vēlāk Pedagoģijas fakultātes dekāne *Mg. philol.* Vineta Trumsiņa (1957–2010) daudz nodarbojās ar augstskolas pedagoģijas jautājumiem un no 1990. gada pārņēma studentu folkloras prakšu vadīšanu (Medveckis 2007: 151). Kopš tā laika studentu savāktie materiāli palika un glabājās Liepājā.

Mg. philol. Gita Girņus sāka strādāt Liepājas Universitātē 1993. gadā un no 1994./95. studiju gada sāka vadīt folkloras kursu pirmsskolas pedagoģijas studentiem. Vēlākos gados viņa vadījusi praksi arī sākumskolas pedagoģijas, speciālās pedagoģijas, vēstures skolotāju programmas un kultūras vadības studentiem. Kādu laiku latviešu folkloras kursu un folkloras praksi latviešu valodas un literatūras skolotāju specialitātes studentiem vēl turpinājusi vadīt Vineta Trumsiņa, bet no 1997. gada folkloras prakses dažādu specialitāšu studentiem jau notikušas tikai Gitas Girņus vadībā.

Gita Girņus sastādījusi metodisku līdzekli Liepājas studentiem “Norādījumi folkloras praksei”, kur uzsvērts, ka “folkloras prakses mērķis ir iesaistīt studentus folkloras materiālu vākšanas darbā, kas nodrošinātu vēl nefiksētu tradicionālās kultūras vienību saglabāšanu, kā arī sniegtu liecības par folkloras dzīvotspēju, noturīgumu un veidošanos mūsdienās” (Girņus 2001: 3). Šajā materiālā studentiem dotas galvenās norādes, kā atrast un izvēlēties teicējus, kādas ziņas sniedzamas par teicējiem. Tāpat doti ieteikumi, kā veidot sarunu, kā atpazīt un piefiksēt stāstījumā iekļautās folkloras vienības. Turpat pieejami arī praktiski norādījumi savāktā materiāla sakārtošanai un noformēšanai.

Ap 2001. gadu Liepājas Pedagoģijas akadēmijā pie Latviešu valodas un literatūras katedras tika izveidots Latviešu folkloras un dialektoloģijas kabinets, par tā vadītāju kļuva Gita Girņus. Šī kabineta paspārnē nodibinājās un aktīvi darbojās studentu folkloras draugu kopa “Baļķi”, kas apvienoja radošus latviešu tautas tradīciju un folkloras interesentus, kuri muzicēja, dziedāja tautasdziesmas, izzināja, un iedzīvināja jau savāktu sava novada folkloras mantojumu. Studentu folkloras kopa “Baļķi” Ievas Tālbergas vadībā darbojās no 2000. līdz 2010. gadam (Girņus 2021).

Latviešu folkloras un dialektoloģijas kabinets pēc pāris gadiem pārtapa par Kurzemes folkloras un valodas centru un tika atklāts 2003. gada 16. decembrī. Šī centra darbības galvenie mērķi un uzdevumi bija latviešu folkloras un latviešu valodas izloksņu (īpaši Kurzemes reģiona) popularizēšana un kultūras mantojuma apzināšana. Tajā turpinājās gan latviešu dialektoloģijas (Benita Laumane, Liene Markus-Narvila), gan latviešu folkloras (Gita Girņus) pētniecība. Centra atklāšanas dienā sniedzot interviju laikrakstam “Kurzemes Vārds”, Gita Girņus akcentēja šī centra nozīmību: “Ir sperts pirmais nopietnais solis, lai vienkopus būtu pieejami studentu un mācībspēku nopietnā darba materiāli, jo folklorā ir krātuve, no kuras spēku ņemt.” (Meistere 2003)

Sadarbībā ar Kurzemes folkloras un valodas centru folkloras draugu kopa “Baļķi” vairākās programmās iedzīvināja latviešu svētku tradīcijas, piedalījās pasākumos Liepājā, dažādās

Kurzemes vietās un citviet Latvijā. “Baļķi” viesojās arī Lietuvā un iepazīstināja ar tautasdziesmu dzīvo skanējumu Klaipēdas Universitātes studentus, kuri apguva latviešu valodu un studēja latviešu folkloru (Girņus 2021). “Baļķu” dziedātās tautasdziesmas ir ierakstītas un izdotas kompaktdiskā.

2007. gada septembrī uz Kurzemes folkloras un valodas centra bāzes tika nodibināts Liepājas Universitātes (LiepU) Kurzemes Humanitārais institūts (KHI). Institūta (KHI) darbības galvenie mērķi ir latviešu kultūras, īpaši Kurzemes etnoģeogrāfiskā novada, folkloras, valodas, vēstures, mākslas un literāro procesu mantojuma apguve un zinātniska izpēte, kā arī humanitāro zinātņu studiju veicināšana un pētījumu popularizēšana (LiepU KHI).

Kurzemes folkloras un valodas centra pētnieki Gītas Girņus vadībā katru gadu devās folkloras un valodas materiālu vākšanas ekspedīcijās. Vēlāk šis darbs tika turpināts Kurzemes Humanitārā institūta paspārnē. Ekspedīcijās kopā ar pētniekiem piedalījās arī studenti. Tās notika šādos Liepājas apkaimes pagastos:

Rucavā (2004, 2005, 2006, 2007), Nīcā (2008, 2009, 2011), Virgā (2012), Bārtā (2013, 2014), Priekulē (2018).

Ekspedīciju laikā savāktie folkloras, dialektoloģijas un toponīmas materiāli ir izmantoti pētījumiem, kas apkopoti kolektīvajās monogrāfijās “Rucavā, tur Paurupē...” (2007) un “Bārtas dzīpari” (2017). Grāmatas “Rucavā, tur Paurupē...” radošais kolektīvs un redakcijas kolēģija – *Dr. habil. philol.* Benita Laumane, *Dr. philol.* Ieva Ozola, *Mg. philol.* Gita Girņus par zinātnisko izpēti, materiālu dokumentēšanu un publicēšanu saņēmušas Lielo folkloras gada balvu.

Daļa no Kurzemes reģionā pierakstītajiem stāstījumiem un studentu folkloras prakšu materiāliem apkopoti un izdoti tematiskā krājumā “Bija tāds gadījums” (2012).

Liepājas Universitātes folkloras arhīvs

Liepājas Universitātes uzglabātajā folkloras materiālu arhīvā pieejami gan pētnieku organizēto etnogrāfisko ekspedīciju materiāli, gan arī 310 studentu iesniegti folkloras prakšu vākumi, kas vākti no 1987. līdz 2020. gadam.

2020. gadā Liepājas Universitātes Kurzemes Humanitārā institūta (KHI) pētnieces sadarbībā ar LU LFMI Latviešu folkloras krātuvi izstrādāja projektu “Tradicionālā kultūra digitālajā vidē” un saņēma VKKF atbalstu folkloras materiālu digitalizēšanai un sagatavošanai publicēšanai digitālajā vietnē *garamantas.lv*.

Liepājas studentu folkloras prakšu materiāli gadiem ilgi glabājās viņu veidotajās kladēs, kur bija piešķirts viens numurs visam studenta savāktajam materiālu kopumam. Lai materiāli būtu plašāk pieejami un pārskatāmi, KHI pētnieces reģistrēja katru atsevišķo folkloras vienību, saskaņojot to reģistrācijas numurus ar Latviešu folkloras krātuves sistēmu. Materiālu skenēšanā tika iesaistītas studentes, kas tos ieskenēja un sagatavoja publicēšanai tiešsaistē. Pēc tam tika uzsākta materiālu digitalizēšana un sagatavošana publicēšanai digitālajā vietnē *garamantas.lv*.

Materiālu digitalizēšana turpinās ar VPP projekta Nr. VPP-IZM-DH-2020/1-0001 “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” atbalstu. Projektā Liepājas folkloras materiāli tiek apstrādāti, digitalizēti un publicēti tiešsaistē digitālajā arhīvā *garamantas.lv*. Projekta īstenošanas gaitā pētnieces un studentes darbu turpina, katrai folkloras vienībai veidojot metadatus – ziņas par teicējiem, vācējiem, folkloras tekstu savākšanas vietām, pierakstīšanas laiku un žanrisko piederību.

Liepājas Universitātes studentu materiāli ir sakārtoti 28 kolekcijās, kur reģistrētas 25 290 folkloras vienības, kas 2021. gadā jau ievietotas digitālajā arhīvā *garamantas.lv*, taču pilnībā to apstrāde un metadatu veidošana tiks pabeigta līdz 2022. gada septembrim.

Studentu savāktais folkloras materiālu kopums ir nozīmīga Kurzemes kultūras mantojuma dokumentācija, jo vairums materiālu savākti un pierakstīti tieši Kurzēmē: Liepājas apkārtnē (140 vākumu), Kuldīgas novadā (40), Saldus (25), Ventspils (23), Talsu (20), Tukuma (17), Dobeles novadā (17). Vairāki folkloras materiālu vākumi nāk arī no Bauskas, Jelgavas, Cēsu, Jēkabpils, Preiļu un Rīgas apkārtnes.

Lai digitalizētie folkloras materiāli būtu sabiedrībai vieglāk pieejami un plašāk izmantojami, 2021. gada maijā un jūnijā tika organizēta akcija “Atšifrēšanas talka”, kurā folkloristi aicināja ikvienu interesentu izzināt dzimtajā pusē savāktos folkloras un tradīciju materiālus, palīdzot atšifrēt un pārrakstīt digitalizētos studentu pierakstus. Intereses rosināšanai tika norādīts, ka starp šiem materiāliem varbūt atrodams arī pašu ģimenes, radu vai draugu lokā pierakstītais. Īpaši uzrunāti tika kurzemnieki, jo liela daļa materiālu vākti Kurzēmē, taču ir krietni daudz materiālu arī no citiem novadiem. Kopš akcijas sākuma interesenti interneta vietnē *garamantas.lv* var piedalīties virtuālajā talkā – tradicionālās kultūras materiālu pārrakstīšanā. Katram tur iespējams atrast un izvēlēties savām interesēm atbilstošus materiālus.

Žanru daudzveidība

Liepājas studentu vākumos pārstāvēti visdažādākie folkloras žanri. Tur atrodami folkloras materiāli, kas viegli ierindojami klasisko žanru kategorijās: pasakas, teikas, nostāsti, tautasdziesmas, mīklas, ticējumi, parunas, sakāmvārdi, anekdotes, paražas un tautas ārstniecības

padomi. Studentu materiālos atrodams bagātīgs praktisko padomu klāsts un daudz ēdienu recepšu, savākts liels skaits dažādu laiku dziesmu un ziņģu. Pierakstītajos materiālos fiksēti arī dzīvesstāsti, personīgās pieredzes naratīvi, sapņu tulkojumi, rotaļas, spēles un mēles mežģi.

Atsevišķu samērā plašu grupu veido kultūrvēstures un etnogrāfiskie materiāli – starp tiem ir plaši pagastu vēstures un vietu, arī politisko notikumu un vietējo sabiedrisko norišu apraksti. Daudzviet ir fiksēti teicēju atmiņu stāsti par īpašiem dzīves notikumiem, naratīviem pievienotas fotogrāfijas, kas raksturo kādu vietu vai laiku, arī dokumentu kopijas: gan skolu liecības un apliecības, gan saimniecības dokumenti.

Līdzās atmiņu stāstiem un nostāstiem studenti aktīvi vākuši teicēju personīgās pieredzes stāstus un memorātus. Īpaša uzmanība pievērsta stāstījumiem, kas saistīti ar konkrētām un pašiem vācējiem pazīstamām vietām, piemēram, memorāts par Liepājas pilsētas bankas ēku:

“Mana mammīte visu savu darba mūžu nostrādāja bankā. Un viņiem tūlīt pēc kara, nu nepateikšu vairs, kas tās bija par naudas operācijām, bet bija jāstrādā arī naktīs. Tā banka atradās Teātra ielā, un kādreiz tā bija Liepājas pilsētas banka. Un tad viņa stāstīja, ka bija starp pulksten divpadsmitiem un trijiem, viņa dzirdēja soļus. Un tādus smagus, tā kā zābaku. Un tad viņa prasīja apsargiem, kas tur toreiz bija. Un viņi bija teikuši, ka tas ir cauri vien. Un izrādījās, ka tur kara laikā ir bijis hospitālis uz kaut kādu laiku ierīkots. Vai nu tur dikti daudz bija miršanas gadījumi, vai traģiski gadījumi, bet to viņa stāstīja, ko viņa pati pieredzēja.” (LFK LiepU 27, 202)

Par pārdabiskas dabas notikumiem vēsta diezgan daudz Liepājas krājuma materiālu, piemēram, pieredzes stāsts par lādēšanas ietekmi, neparastas pieredzes stāsts par Melno grāmatu vai stāsti par svētvietās piedzīvoto:

“Man stāstīja, ka viens vīrietis no Talsiem bija Pokaiņu mežā, izstaigāja visu un piemiņai paņēma akmeni. Aizveda līdz uz mājām. Tad naktīs viņš sāka murgot, vispār dzīvē neveicās. Ar vienu cilvēku, kas no tādām lietām kaut ko saprot, bija aizrunājušies līdz tam akmenim, un viņš teicis, ka tādas lietas nevar un nedrīkst nest mājās. Vīrietis aizveda atpakaļ, nolika, un dzīvē viss nokārtojās.” (LFK LiepU 15, 14255)

Šie vēstījumi ierindojami pie ticējumteikām, kas noteiktā apvidū veidojušas pamatu uzskatam, ticējumam, ka no svētvietas neko nedrīkst ņemt līdzī un svētajiem kokiem nedrīkst lauzt zarus. Vairāki iespaidīgi piemēri vēsta par Užavas svētliepas īpašo spēku: “Kādu reizi tuvējo māju puisis, nevarēdams tikt pāri grāvim, nocirtis liepai trīs zarus. Otrā dienā viņu piemeklējusi nelaime. Cērtot malku, nocirtis rokai trīs pirkstus.” (LFK LiepU 17, 15492)

Otrs ticējumteikas piemērs vēl vairāk pastiprina radīto iespaidu un nodod pamācošo vēsti, kā tas raksturīgi ticējumiem: “Senos laikos kāda sieviņa salasījusi liepziedus tējai. Tēju dzerot, palikusi akla. Jo nedrīkstot aiztikt Svētliepu. Ja to darot, tad piemeklējot ļauna slimība vai kāda cita nelaime.” (LFK LiepU 17, 15493)

Sarp studentu savāktajiem personīgās pieredzes stāstiem atrodami arī smieklīgi, komiski pieredzes stāsti jeb anekdotiski notikumi no paša teicēja vai tuvinieku dzīves:

“Un tad atceros, kad Alberts bij mazs, nu, mans brālis, viņam laikam trūka kaļķis. Tad jau nekādas kaļķa tabletes, nekas tāds nebija, un laukos jau neviens tos bērņus tā neuzpasēja. Un tad viņš sēdēja pie plīts un tur sienu rubināja, un to ēda. Jā, un Milda, mana māsa, atkal skrēja pie mammas un teica: “Mamm, nāc ātrāk, Alberts mūsu māju apēdis!” (LFK LiepU 19, 17956)

Saviem vākumiem studenti pievienojuši rokdarbu rakstus, kas gan rūpīgi pārzīmēti un izkrāsoti pa rūtiņām, gan fiksēti fotogrāfijās. Iesniegti pat pašu rokdarbu paraugi vai fragmenti. Sarp folkloras materiāliem studentu vākumos atrodami arī valodniecības materiāli: vecvārdi un to skaidrojumi, kā arī bērnu valodas piemēri.

No jaunākiem folkloras žanriem, kuri sastopami Liepājas Universitātes krājumā, daudz savākti atmiņu albumu panti un dažādi apsveikumu pantiņi.

Atsevišķu grupu veido skolēnu folklorā, kur fiksēti gan skolotāju, skolēnu un studentu izteicieni, gan neparastu situāciju apraksti un pat piezīmes dienasgrāmātā.

Nosacīti šai grupai pieskaitāmas pārveidotas un situācijai piemērotas tautasdziesmas, piemēram, vidusskolēnu veidotas apdziedāšanās dziesmas:

“Divpadsmitie danci veda
Ap to lielo siena kaudzi,
Citiem mati kā kodaļas,
citiem ausis kā lemeši.”
(LFK LiepU 17, 14)

Studentu vākumos atrodami sabiedrības aktualitātēm atbilstoši jaunu folkloras žanru meklējumi un jaunievedumi, kas, kā redzams, savulaik tikuši prakšu vadītāju apstiprināti, piemēram, mobilo tālrunu SMS ziņas. Divu studentu materiālos savāktas visdažādāko veidu īsziņas. Tās bija populāras pirms interneta izplatīšanās un tika savāktas 2004. gadā. Studenti īsziņas sistematizējuši pēc tematikas: dzimšanas un vārda dienu īsziņas, Lieldienu, Jāņu laika, Ziemassvētku un Jaunā gada īsziņas, dienesta un romantiskās īsziņas. Šobrīd arī tās jau ir daļa no vēstures, jo tolaik aktīvo dalīšanos ar mobilo tālrunu īsziņām mūsdienās ir aizstājušas ziņas sociālo tīklu vietnēs.

Neparasts materiāls, kas tā īsti neiekļaujas nevienā no žanriem, ir teicējas Ilzes Klagišas (dz. Rubenes) nodotais puķu pulkstenis, ko izmantojuši mīloši ļaudis, lai citiem nemanāmi norunātu satikšanos: “Ja grib satikšanos caur puķēm norunāt, tad pulksten vieniem pasniedz puķīti zaļu, diviem – baltu rožu, trijiem – magoņu, četriem – zilu rudzupuķu [...] Pušķītis, apsīts ar baltu lentu, nosaka dienas laiku, bet pušķītis, apsīts ar zilu lentu, nosaka vakara laika stunda.” (LFK LiepU 2, 257)

Kārtojot daudzveidīgo un neparasto materiālu klāstu, reizēm rodas pārdomas un šaubas par dažu materiālu iedalību starp folkloras materiāliem. Meklējot atbildes par kādu neparastu materiālu piederību, redzeslokā nokļuva pirmās Latviešu folkloras krātuves pārzīnes Annas Bērzkaines sniegtā atbilde par vācamo folkloras materiālu un žanru dažādību. Tā bijusi šāda: “Sūtat vien visu, ko Jūs savā apkaimē varat savākt no dzirdēšanas, lai tas arī nebūtu visai pieklājīgs, jo zinātnei no svara pazīt arī tumšo pusi un nevis gaišo vien.” (Vikсна 2014: 126)

Tā nu arī daži studentu vāktie materiāli būtu ierindojami šajā daļā – nevis nepieklājības dēļ, bet drīzāk neiederības dēļ. Tie ir materiāli, ko grūti ievietot kādā noteiktā žanra kategorijā vai vispār pieskaitīt folkloras materiāliem. Piemēram, tie ir dīvaini pārfrāzēti frazeoloģismi un parunas, bērnu pašsacerētas pasakas un pašizdomātas spēles, gan arī vecāku sacerētas pasakas bērniem. Vairākos vākumos pierakstīti arī teicēju pašu sacerēti dzejoļi. Šāda veida materiāli tomēr nav pieskaitāmi folklorai, jo ir viena konkrēta un pierakstītājam zināma autora radīti.

Savdabīgi un folkloras materiāliem ne īsti piederīgi ir arī mūsdienu kāzu norišu aprakstiem pievienotie materiāli, piemēram, padomi jaunajai sievai un vīram, laulību protokols, kāzu sviņībām un konkrētiem cilvēkiem īpaši sacerēti dzejoļi, izmainīti pazīstamu dzejoļu un dziesmu teksti utt. (LFK LiepU 15, 40).

Nozīmīgu daļu Liepājas studentu materiālu veido ēdienu receptes. Tās tiek ierindotas pie sadzīves paražām apakškategorijā “Nodarbošanās, kas saistīta ar apģērbu, uzturu, higiēnu” (m4 – apģērbs, uzturs, higiēna, ēdieni, dzērieni, tabaka, krāsošana, apģērba pagatavošana, vērpsana, adīšana, aušana, veselība, istabas iekārta, gultas vietas, skaistuma kopšana) (LFK žanru rādītājs).

Kārtojot savāktos folkloras materiālus un apzinoties, cik daudz laika pētniekiem jāpavada, atlasot un analizējot dažādos materiālus, radās jautājums, kā būtu iespējams atvieglot darbietilpīgo meklēšanas procesu, precizējot materiālu žanrisko piederību un specifiku. Ja šobrīd kāds materiālu kopumā vēlas atrast noteikta veida receptes vai konkrētu produktu lietošanas veidus vai arī pētīt tautas uztura tradīcijas, tad nākas doties cauri visiem sadzīves paražu aprakstiem, kas ierindoti šajā plašajā kategorijā.

Kārtojot sadzīves paražu materiālus un apzinoties, ka sabiedrībā pēdējā laikā vērojama samērā liela interese par ēdiena gatavošanu un kulinārajām tradīcijām, ko apliecina daudzveidīgie ēdieniem veltītie raidījumi un periodiskie recepšu izdevumi, aktuāls kļuva jautājums par attiecīgo folkloras materiālu atrašanas iespējām. Tāpat sabiedrībā vērojama interese par neparastām un tradicionālām receptēm, par dabas produktu izmantošanu un veģetāro ēdienu gatavošanu. Piemēram, sabiedrībā pieaugošo interesi par veģetāriešu un vegānu produktiem apliecina sabiedriskās domas un tirgus pētījumi: “Pieprasījums pēc specializētiem veģetāriešu un vegānu pārtikas produktiem, kas atrodami īpašā stendā, katru mēnesi pieaug, sasniedzot pat 254 % pieauguma pēdējā gada laikā.” (Popova, Rimša 2019: 47)

Liepājas folkloras materiālos šīs jomas interesentiem būtu pieejams bagātīgs pētījumu un iedvesmas avots, jo studenti fiksējuši teicēju krājumus, kur apkopots liels skaits dažādu ēdienu un

dzērienu receptšu. Starp tām pieejams bagātīgs veģetāro ēdienu receptšu krājums, kur atrodamas dažādas mazāk zināmas receptes, piemēram, griķu kotletes, kartupeļu pudiņš, ābolu siers un citas.

Savukārt rokdarbnieces un amatnieki, kuri interesējas par senajiem tekstiliju krāsošanas paņēmieniem, par rokdarbu rakstiem un apģērbu darināšanu tautas tradīcijā, varētu atrast vērtīgus praktiskus padomus un radošus impulsus, iepazīstot studentu savāktos materiālus par šīm tēmām.

Vairāku Liepājas studentu vāktajos materiālos fiksēti dažādi apģērbu krāsošanas ieteikumi, kas būtu vērtīgi attiecīgās jomas pētniekiem. Piemēram, studente Linda Zande intervējusi savu vecmāmiņu Zeltu Zaši-Zasi, kura spilgti raksturojusi ne vien krāsošanas paņēmienus, bet arī dabas krāsvielu atklāšanas prieku:

“Es jau pati ar’ dziju krāsoju. Rudenī, kad bija uzsnidzis tas sniegs, gāju gar ābeli, un tai tās lapas bij sakritušas, un tik skaisti dzeltens sniegs nokrāsojies! Un es mudīgi salasīju...redz, šitais ir ābeļu lapās krāsots (skat. 2. pielikumu). Bet ar ko es viņu kodināju, es nezinu. Man liekas, ar pelnu sārmu. Man bij pelni, kā jau pie plīts.” (LiepU19-2005-13-0005)

Pēc žanru iedalījuma sanāk, ka rokdarbu raksti un apģērba darināšanas apraksti atrodami tajā pašā sadzīves paražu apakšskategorijā, kur ēdienu receptes. Turpat meklējami arī praktiskie padomi apģērba kopšanā, dziju krāsošanā, traipu tīrīšanā, kā arī ķermeņa un skaistuma kopšanā.

Nodrošinot folkloras materiālu digitālu pieejamību, iespējams, būtu vērtīgi interesentu un pētnieku meklējumus vienkāršot, minētajā sadzīves paražu kategorijā (m4) izdalot precizējošas apakšskategorijas:

- ēdienu, dzērienu gatavošana un receptes,
- rokdarbi, apģērbu darināšana, krāsošana un kopšana,
- veselība, higiēna un skaistumkopšana,
- praktiski padomi mājāsaimniecībā.

Tā kā materiālu apstrāde vēl turpinās, joprojām tiek meklēti veiksmīgākie risinājumi materiālu sakārtošanai un sistematizēšanai.

Folkloras materiālu vākšana notikusi ne vien folkloras prakšu laikā, bet arī Liepājas Universitātes studiju priekšmetos. Baltu filoloģijas un kultūras studentiem bijuši specializēti studiju kursi “Folkloras praktikums” un “Folkloras par Liepāju”, kur studentiem doti uzdevumi vākt folkloras materiālus. Šādu uzdevumu veidotāja un studentu ievirzītāja jau vairāk nekā divdesmit gadu garumā bijusi docētāja Gita Girņus. Daudzveidīgo studentu savāktos materiālu tapšanā liela loma ir viņas personīgajai aizrautībai un spējai modināt jauniešos interesi par kultūrvēstures mantojumu un lauka pētījumiem.

Materiāli, ko studenti savākuši folkloras praktikuma laikā, ir apkopoti vairākās Liepājas Universitātes krājuma kolekcijās un atrodami pēc pierakstīšanas gada, bet Liepājas folkloras kursa vākumi pieejami atsevišķā – 27. kolekcijā.

Liepājas folkloras materiālu klāsta pamatā ir atmiņu un personīgās pieredzes stāstījumi par pilsētas veidošanos un attīstību, par atsevišķu ielu un ēku tapšanu, arī par dažādu vēstures notikumu izpausmēm, par kultūras un sabiedriskās dzīves notikumiem, par personībām, kuras dzīvojušas un darbojušas Liepājā.

Liecības par Liepājas Pedagoģiskā institūta tapšanas laiku sniedz Jura Braša atmiņu stāsts (pierakstījusi studente Ieva Ozola). Viņš studējis institūtā tūlīt pēc tā dibināšanas (1954–1958) un atceras, ka studenti piedalījušies arī celtniecības darbos: “Kur tagad institūta māja lielā, tur bija paredzēta Kurzemes apgabala māja. Kādreiz, piecdesmitajos gados, Hruščovs taisīja apgabalus, tad tur dabūja to māju. Viņa nebija pabeigta, tur jau nekas nebija, tikai jumts bija uzlikts. Tad mēs tur strādājām pie mājas būvēšanas. Ko lika, to darīja. Betonu pa trepēm vilka augšā ar ķeru. (..) Tur apgabala pārvalde bija paredzēta iekšā. Un tad tos apgabalus neuztaisīja un atdeva institūtam to māju. (..) Kādā piecdesmit sestajā gadā institūtu pārcēla uz to māju. Līdz tam bija Baseina ielā.” (LiepU 27, 44)

Plašāku atmiņu stāstu par Liepājas Peldu iestādes darbību studente Inga Miglone pierakstījusi no Saulīšu ģimenes. Izoldes Saulītes (dz. Kunstmanes) tēvs bijis ārsts, kurš ilgus gadus nostrādājis Nikolaja Peldu iestādē. Fragments no viņas stāsta: “Tēvs taču strādāja, vis gatavoja tās dūņas, viņi viņas sildīja. Var iet apskatīties peldu iestādi, tur ir arī tagad tāds koka šķūnis, tajā ievada dūņas no Liepājas ezera, tur ir vieta, kur tās dūņas rakt, atvest. Tad viņas sildīja un pēc tam padeva pa konveijera lenti uz galveno ēku. (..) Tur bij tādas flīzētas vannas ar metāla kāpnītēm, kur varēja iegult, viņas bij diezgan dziļas. (..) Visa tā sanatorija darbojās līdz deviņdesmit kaut kādiem tur gadiem (1992/1993).” (LiepU 27, 94)

Pierakstot liepājnieku atmiņu stāstus, studenti ir fiksējuši svarīgu aculiecinieku pieredzi, kas vēsta par attiecīgā laika vēsturiskajām un sabiedriskajām norisēm, kā arī pauž vietējo iedzīvotāju attieksmi pret attiecīgajiem notikumiem. To raksturo vēl kāds piemērs no stāsta par Liepājas Peldu iestādi: “Tad redz par kūrortu, Peldu iestādi. Tā vēl bij pēc kara un pirms kara bij. Viņa gandrīz konkurēja ar Ķemeriem. Nu, Ķemeros pēc tam, kad uzbūvēja to jauno māju, tad to Liepāju noballēja.” (LiepU 27, 92)

Vērtīgas liecības teicēju sniegtajos naratīvos var atrast arī kultūras nozares pētnieki, jo materiālos atrodami stāsti par Liepājas kino un teātri, par aktieriem, dziedātājiem un citiem kultūras nozares pārstāvjiem. Šo naratīvu grupu raksturo atmiņu stāsts par Liepājas operu: “Liepājā bij opera, līdz 1949. gadam bij pilnīgi regulāras izrādes, ir operetes, ir operas izrādes bij. Arī kad te bij vācieši ienākuši, tad ar tāpat darbojās opera un teātris, tas bij kara laikā no 1942. līdz 1944. gadam ieskaitot. Tie bij štata aktieri un solisti, pazīstami visā pilsētā, nu un tad to nolikvidēja..” (LiepU 27, 91)

Liepājas Universitātes folkloras prakšu materiālus docētāji izmanto folkloras kursa uzdevumiem un folkloras praktikuma nodarbībās, sniedzot piemērus, kā un kādi materiāli agrāk tikuši vākti, kādi bijuši teicēji un kādas pierakstītāju intereses dažādos periodos. Tādējādi studentiem ir iespēja gūt dziļāku priekšstatu par lauka pētījumu būtību un labāk sagatavoties patstāvīgi veicamajam uzdevumam. Līdz šim studiju procesā lauka pētījumu raksturošanai

tikuši izmantoti atsevišķi Liepājas Universitātes arhīvā atrodamie studentu vākumi, kas bija pieejami kladēs un mapēs rokrakstos vai izdrukātā veidā. Kopš materiālu digitalizēšanas docētājiem un studentiem iespējams piekļūt visiem Liepājas studentu savāktajiem materiāliem un salīdzināt tur vērojamās attiecīgajam laika periodam raksturīgās iezīmes un aktuālos žanrus. Tāpat ir iespējama arī plašāka salīdzinošā izpēte un visa materiālu korpusa analīze.

Folkloras kursa laikā studentiem tiek dots ieskats digitāli pieejamo folkloras materiālu krātuvēs, raksturojot katru vietni un tur atrodamos materiālus. Tas būtiski atvieglo studiju procesu un tika īpaši novērtēts laikā, kad nebija iespējams apmeklēt bibliotēkas un pārļapot drukātos folkloras avotus.

Lai popularizētu folkloras un kultūras mantojuma pētniecību, studenti tiek iesaistīti gan folkloras vākšanā, gan ekspedīciju darbā kopā ar pētniekiem, gan materiālu apstrādē un atšifrēšanā. Tādējādi studentiem tiek dota iespēja praktiski iepazīt pētniecības darbu.

Liepājas folkloras materiālu digitalizēšanas projektu darbos tika iesaistītas divas Baltu filoloģijas programmas studentes, kuras jau folkloras kursa apgūšanas laikā izrādīja padziļinātu interesi par šo jomu un parādīja spēju patstāvīgi un sistemātiski strādāt. Tās ir pētniecības darbam nepieciešamas kvalitātes, ko studentēm šajā projektā ir iespēja gan apliecināt, gan tālāk pilnveidot.

Secinājumi

Liepājas Universitātes folkloras krājums veidojies dažādos laika periodos kopš augstskolas dibināšanas 1954. gadā. Augstskola kopš tā laika ir mainījusi gan nosaukumus, gan studiju virzienus un mērķus, bet folkloras materiāli ir vākti gandrīz katrā tās attīstības posmā līdz pat mūsdienām. Folkloras materiāli ir vākti gan studentu folkloras prakšu laikā, gan atsevišķos studiju priekšmetos, gan arī pētnieku organizētās ekspedīcijās jeb lauka pētījumos. Liepājas Universitātes studentu un pētnieku savāktie folkloras materiāli (kladēs un mapēs) glabājas Latviešu folkloras krātuvē un Kurzemes Humanitārā institūta folkloras arhīvā.

Laika gaitā Liepājas Universitātes krājumā savāktie folkloras materiāli ir izmantoti gan studiju procesā atbilstošos mācību priekšmetos, gan arī zinātniskajos pētījumos, kas apkopoti tematiskos rakstu krājumos.

Sistematizējot un digitalizējot folkloras materiālus, projektā bija iespēja gūt pārskatu par visu savākto krājumu, fiksējot žanrisko daudzveidību un pamanot atsevišķu studentu intereses tradīciju jomā, kā arī docētāju netiešo ietekmi un sabiedrībai aktuālu tēmu izpaušmes. Piemēram, pēdējo gadu norises sabiedrībā, iespējams, ir veicinājušas lielāku studentu interesi par veselības uzturēšanas līdzekļiem un metodēm, jo pēdējo pāris gadu vākumos vērojams krietni lielāks tautas ārstniecības materiālu daudzums nekā iepriekšējos gados.

Rakstā atainota un ar piemēriem raksturota folkloras materiālu žanriskā daudzveidība un plašais tēmu loks, kāds pieejams Liepājas Universitātes folkloras materiālu krājumā. Iespējams, ka tas rosinās dažādu nozaru pētniekus izmantot šos materiālus saviem darbiem un dos impulsus jaunu pētījumu tapšanai.

Projekta “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” īstenošanas gaitā digitalizētos materiālus turpmāk būs iespēja izmantot gan folkloras un tradīciju interesentiem, gan pētniekiem. Nozīmīgs ir apjomīgais darbs pie metadatu veidošanas un tehniskās apstrādes, kas nodrošina veiksmīgus materiālu atlases un tālākas sistematizēšanas procesus.

Projekta rezultātā ir veikta Liepājas Universitātes folkloras arhīva materiālu sistematizēšana, apstrāde un digitalizēšana, tādējādi sagatavojot un pievienojot šos materiālus Latviešu folkloras krātuves digitālajam arhīvam *garamantas.lv*. Tādējādi šie iepriekš maz zināmie un nesistematizētie folkloras materiāli kļūst pieejami gan pētniekiem, gan folkloras interesentiem Latvijā un pasaulē, kā arī veidojas pilnīgāka dokumentēto folkloras materiālu kopaina. Pēc projekta īstenošanas šie nemateriālā kultūras mantojuma dokumentējumi būs brīvi pieejami plašai sabiedrībai.

Giņģus, Gita (2001). *Norādījumi folkloras praksei. Metodisks līdzeklis studentiem*. Liepāja: Liepājas Pedagoģijas akadēmija.

Medveckis, Arturs (2007). *Akadēmiskā personāla biogrāfiskais rādītājs*. Liepāja: Liepājas pedagoģijas akadēmija.

Meistere, Daina (2003). Vieta, kur glabāt vārdu, ir atklāta. *Kurzemes Vārds*, 18.12.

Popova, Rimša (2019). Uzturs vegāniem – cik pieejams Latvijā? *Inquisitio*, RSU, Nr. 5, 45.–48.lpp. Pieejams: https://issuu.com/rsu49/docs/18-070-inquisitio-5_makets-low [skatīts 12.10.2021.].

Viksna, Māra (2002). Liepājas pedagoģiskā institūta (augstskolas) ieguldījums latviešu folkloras vākšanā. *Aktuālas problēmas literatūras zinātnē*, Nr. 7, 308.–318. lpp.

Viksna, Māra (2014). Latviešu folkloras krātuves izveide un darbība. *Latviešu folkloristikā starptaru periodā*. Rīga: Zinātne, 101.–151. lpp.

Elektroniskie resursi

Garamantas. Latviešu folkloras krātuves digitālais arhīvs. Pieejams: <http://garamantas.lv/> [skatīts 20.10.2021.].

LFK LiepU. Latviešu folkloras krātuve, Liepājas Universitātes krājums. Pieejams: http://garamantas.lv/lv/collection/index?CollectionSearch%5Brepository_id%5D=1477981 [skatīts 12.10.2021.].

LFK 1947. Benitas Laumanes folkloras vākums (LFK 1947, Benita Laumane, Pieejams: <http://garamantas.lv/lv/collection/887922/Benitas-Laumanes-folkloras-vakums> [skatīts 12.10.2021.].

Liepājas Universitātes Folkloras un valodas centrs. Pieejams: <https://kfvc.liepu.lv/index.php?lapa=sakums> [skatīts 10.10.2021.].

LiepU KHI. Liepājas Universitātes Kurzemes Humanitārais institūts. Pieejams: <https://khi.liepu.lv/parkhi> [skatīts 20.10.2021.].

LiepU vēsture. Liepājas Universitātes vēsture. Pieejams: <https://www.liepu.lv/lv/173/vesture> [skatīts 20.09.2021.].

Literatura.lv. LU Literatūras, folkloras un mākslas institūta literatūras nozares virtuālā resursa *literatura.lv* personu rādītājs. Pieejams: <https://literatura.lv/lv/person/Mara-Viksna/876090> [skatīts 05.10.2021.].

LU LFMI. Latvijas Universitātes Literatūras, folkloras un mākslas institūts. Pieejams: <http://lulfmi.lv/jaunumi> [skatīts 10.10.2021.].

Intervijas un sarakstes (intervējusi Ieva Ančevska)

Giņģus (2021) – Gita Giņģus, sarakste 12.10.2021.

Laumane (2021) – Benita Laumane, intervija 11.10.2021.

Viksna (2021) – Māra Viksna, sarakste 05.10.2021.

Formation, Processing and Digital Accessibility of the Liepaja University Folklore Collection

Ieva Ančevska

Keywords: folklore practices, variety of genres, sorting, digitizing and using folklore materials

This article summarizes the information about the creation of the folklore collection at the Liepaja University during different time periods since the foundation of the University in 1954. Over time, the University has changed both its name and its pedagogical directions and goals, but folklore materials have been collected at almost every stage of its development, up to the present day. The folklore materials have been collected during students' folklore practices, within the framework of specific course subjects, as well as during expeditions or field research organized by researchers.

Eventually, the folklore materials gathered in the collection of the Liepaja University have been used both in the study process of relevant subjects as well as in scientific research, which is summarized in thematic collections of articles.

This article sheds light on the various ways of collecting and storing folklore materials up to the present day. It also provides an overview of the work done over the last year – on the processing and digitizing of the collection in order to add these materials to the digital archive of the Latvian Folklore Repository *Garamantas.lv*. The folklore materials are characterised by genre diversity and thematic ranges, as well as the problems that have been encountered during the processing and digitizing of the collection.

The availability of digitized materials today is an important facilitator of study and research processes. The article investigates how the integration of digitized materials in studies and research is carried out, as well as outlines the possibilities of using the folklore collection of the Liepaja University in the future.

Liriskās ģeogrāfijas: literārās telpas kartēšana latviešu rakstnieču romānos un īsprozā

Zita Kārkla, Eva Eglāja-Kristsone

Publikācija tapusi divos projektos: “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001), “Ķermeņa ģeogrāfijas: latviešu sieviešu rakstniecības vēsture” (Nr. 1.1.1.2./VIAA/3/19/430).

Ievads

20. gs. 20. gados Latvijas literatūras procesa apskatos un kritikās ienāk apzīmējumi “literārās ekskursijas”, “ģeogrāfiskā beletristika” un apzīmējums, ko arī izmantojam šī pētījuma virsrakstā – “liriskā ģeogrāfija”. Visus apzīmējumus lietojis Andrejs Upīts savos pārskatos par jaunāko latviešu literatūru žurnālā “Domas”, visbiežāk tieši pēdējo apzīmējumu, pamatojot, ka liriskā ģeogrāfija latviešu literatūrā “aug līdz ar rakstnieku ārzemju ceļojumiem ar Kultūras fonda biļetēm un bez tām” (Upīts 1927). Upīts šeit atsaucas uz 1920. gadā tika nodibināto Kultūras fondu, kurš atbalstīja gan dažādas kultūras iestādes, gan individuālus māksliniekus, tajā skaitā to braucienus uz ārvalstīm. Zīmīgi, ka, raksturojot Emīlijas Prūsas romānu “Itāliete”, Upīts atzinīgi izsakās nevis par darba literārajām kvalitātēm, bet gan par to, ka Prūsa “pārāk nejūsmo, izsargās no lieka lirisma, tēlo pietiekoši sīki un lietišķi, katrā vietā un atsevišķā priekšmetā uztver to raksturīgāko un nekur daudz neatkāpjās no zinātniskā ģeogrāfijā un ceļa vadoņos lasītā” (Upīts 1927).

Literāro ģeogrāfiju kartēšana nav jauns paņēmieni: literatūras atlanti ir bijuši aprītē kopš deviņpadsmitā gadsimta beigām, pēc tam sekoja eksperimenti literārajā kartogrāfijā (Reuschel, Hurni 2011: 293). Jautājumi, kas radās šajos agrīnajos literatūras kartēšanas gados, ir palikuši nemainīgi. Pirmkārt, kā kartēt naratīvus un to sarežģīto telpisko struktūru? Un otrkārt, kāpēc kartēt literatūru, ko mēs ar to iegūstam? Anna Katrīna Reišela (*Anne-Kathrin Reuschel*) un Lorencs Hurni (*Lorenz Hurni*), piemēram, ir ierosinājuši divas pamata metodoloģijas pieejas literārajai kartēšanai: kartēs var “attēlot atsevišķus tekstus, lai iegūtu dziļāku, analītisku ieskatu naratīva telpiskajā struktūrā”, vai arī var koncentrēties uz pazīmju grupu, lai izceltu autora, žanra, motīva vai laikmeta literāro ģeogrāfiju. (Reuschel un Hurni 2011: 293). Savā rakstā izmantojam otro pieeju, akcentējot prozas žanru caur sieviešu rakstnieču tekstiem konkrētā laika posmā no 1872. līdz 1940. gadam.

Mūsu interesi par telpu un vietām latviešu rakstnieču tekstos veicinājuši gan sieviešu mobilitātes palielināšanās 19. gadsimta pēdējās desmitgadēs, gan tendences literatūrpētniecībā, kas saistītas ar digitālo humanitāro zinātņu attīstību. Digitālo humanitāro zinātņu rīki un metodes literatūras kartēšanai un vietu analīzei izmantoti arī latviešu literatūrzinātnē. Pauls Daija un Benedikts Kalnačs, digitālās analīzes pamatā izmantojot latviešu romānu korpusu (1900–1914), norādījuši uz ģeogrāfiskās telpas paplašināšanās attēlojumu romānos, saistot literārās reprezentācijas ar vēsturiskajiem un sociālajiem notikumiem,

piemēram, ar nācijas veidošanās procesu deviņpadsmitā gadsimta un divdesmitā gadsimta sākuma Latvijā (Daija, Kalnačs 2019; Daija, Kalnačs 2020). Pētnieki digitālajai analīzei izmantojuši latviešu romānu korpusu (46 teksti), kas – ar diviem izņēmumiem (Tīrzmalieta un Ivande Kaija, kuru darbi netiek izmantoti tuvlasījumam) – pamatā sastāv no vīriešu rakstītajiem romāniem, tādējādi sievietes šajā analīzē “tiek izlaistas”. Savukārt Inguna Daukste-Silasproģe, pievēršot uzmanību atsevišķu autoru atsevišķiem tekstiem, kartējusi trimdas rakstnieku Andreja Johansonu un Indras Gubiņas biogrāfijas un darbus (Daukste-Silasproģe 2020; Daukste-Silasproģe 2022). Neviens no šiem pētījumiem nav pievērsies latviešu rakstnieču sieviešu biogrāfiju un literāro darbu ģeogrāfiskajiem aspektiem pirms 2. pasaules kara.

Raksta mērķis ir izziņāt digitālās kartēšanas iespējas sieviešu rakstniecības vēstures pētīšanā. Analizējot latviešu rakstnieču prozas tekstu korpusu (1872–1940), mūs īpaši interesēja sieviešu literārās telpas ģeogrāfija un tas, ko jaunu, pievēršot uzmanību vietai, iespējams uzzināt par sieviešu rakstniecību un sociālo vēsturi, kas savā starpā ir cieši saistītas. Raksta pirmajā daļā aplūkots datu vizualizācijas un digitālo kartēšanas metožu un rīku lietojums literatūrpētniecībā, kā arī ģeokritikas un *geo-parler femme* jēdziens. Raksta otrajā daļā, izmantojot literārās digitālās kartēšanas rezultātus, kas iegūti, lietojot ģeogrāfiskās anotācijas un vizualizācijas sistēmu – *Google Maps*, Ģeogrāfiskās informācijas sistēmu (GIS) un statistikas metodes, pievēršamies sieviešu romānu un īsprozas ģeogrāfijas analīzei, uzdodot jautājumus par sieviešu literārās telpas teritoriju un attiecībām starp lokālā un globālā reprezentācijām. Savukārt raksta trešajā daļā, izmantojot *geo-parler femme* pieeju un izveidotās kartes, lai vizualizētu autorens biogrāfiju un ģeogrāfiskās atsauces literatūrā, salīdzinām Martas Grimmas autobiogrāfiskajos darbos (ceļojumu aprakstā, atmiņās) un prozas darbos (stāstos, romānā) minēto vietu sabal-sošanos, izdarot secinājumus par reālās un izdomātās telpas krustpunktiem.

Digitālā kartēšana, ģeokritika un *geo-parler femme*

Lai arī pastiprināta interese par literatūru un telpu, literatūras teritoriālajiem un topogrāfiskajiem aspektiem iekļaujas plašākā “telpiskā pagriezienā” humanitārajās zinātnēs (20. gs. 70., 80. gados), īpaši ar postmodernisma un postkoloniālās teorijas parādīšanos telpas problemātika nokļuvusi daudzu pētījumu centrā. Kā to skaidrojis viens no vadošajiem kartēšanas vēsturniekiem un teorētiķiem Roberts Tallijs (*Robert T. Tally*), “pieeja naratīvam kā telpiski simboliskam aktam ļauj orientēties literatūrā un pasaulē interesantā un jaunā veidā, uzdodot dažādus jautājumus, pētot dažādas teritorijas un atklājot dažādus aspektus. Rakstnieki kartē savu pasauli, savukārt lasītāji vai kritiķi var izmantot šīs stāstījuma kartes, lai orientētos un izprastu lietas mainīgajā pasaulē” (Tally 2011). Tātad ieguvums no jebkuras literārās kartes ir tas, ka tā sniedz jaunas perspektīvas pašizziņā un pasaules izziņāšanā.

Ja literārās ģeogrāfijas pētījumi 20. gadsimta sākumā lielākoties ir ilustratīvi, proti, kartes pamatā atbalsta tekstu, tad līdz ar Franko Moreti (*Franco Moretti*) “Eiropas romāna atlantu”

(1998) iezīmējas jauna laikmeta sākums literārajā ģeogrāfijā, kur kartes kļūst par interpretācijas instrumentiem, ļaujot ieraudzīt to, kas līdz šim nav bijis acīmredzams (Piatti et al. 2009). Pēc Moreti domām (“Eiropas romāna atlants” (1998); “Grafiki, kartes, koki” (2005)), skaidrojot saikni starp ģeogrāfiju un literatūru, literatūra ir jākartē, jo karte ļauj mums ieraudzīt nozīmīgas attiecības, kam līdz šim neesam pievērsuši uzmanību, un “ar nelielu veiksmi šīs kartes būs vairāk nekā to atsevišķo daļu summa: tām būs “jaunās” īpašības, kas nebija pamānāmas iepriekš” (Moretti 2005: 53). Reducējot tekstu līdz dažiem elementiem un abstrahējot šos elementus no stāstījuma plūsmas, tiek izveidots jauns mākslīgs objekts – kartes, un tādējādi teksts tiek sagatavots tālākai analīzei.

Tieši pēdējās desmitgadēs, attīstoties digitālajām humanitārajām zinātnēm, interese par telpu ir strauji pieaugusi (Tally 2017: 2). Digitālo ģeotelpisko tehnoloģiju, tai skaitā digitālo kartēšanas rīku izveide un attīstība, ir bagātinājusi un pārveidojusi arī literāro karšu veidošanas prakses. Pavēršiens uz digitālajām metodēm un rīkiem literārajā kartēšanā piedāvā gan lielāku mērogu, gan precizitāti, radot iespējas pētīt jauna veida kritiskos objektus. Parādoties digitālajai literārajai kartogrāfijai, centieni vizualizēt un analizēt attiecības starp literatūru un ģeogrāfiju ir ievērojami palielinājušies. Piemēram, Literatūras atlanta projekts ir iespēja pārrakstīt literatūras vēstures, apvienot milzīgu daudzumu atsevišķu pētījumu visaptverošā projektā, tas ir interaktīvs tiešsaistes atlants, kas piedāvā dažādas kartes, kurās atrodami Velsā izdotie romāni angļu valodā (sk: <http://www.literaryatlas.wales/en/>). Gan pētniekiem, gan citiem lietotājiem (skolēniem, studentiem, skolotājiem, lasītāju kopienai kopumā) literatūras atlants piedāvā ne vairs hronoloģiski, bet telpiski sakārtotu Eiropas literārā mantojuma vēsturi un procesu, kas neapstājas pie valstu vai valodu robežām (Piatti et al. 2009). “Ģeohumanitātes” ir veicinājušas vietas nozīmes pārkonceptualizāciju attiecībās ar naratīvu, reprezentācijas praksi un digitālajām tehnoloģijām (Presner and Shepard 2016: 202). Iespēja ģeoreferencēt autores radīto pasauli digitālās kartes saskarnē ir viens no veidiem, kā samazināt telpu starp literāro ģeogrāfiju un literāro kartogrāfiju, tas ir, autores radīto pasauli un lasītāja apgūto pasauli. Lai gan kartēšana pilnībā nenovērš plaisu starp tām, jo nenodod katrai vietai autores iztēloto elementu, kas piedod dzīvību tekstā, tā vismaz attēlo autores iztēles vērienu, vizualizējot viņas ģeogrāfiskās pasaules centrus, kopas un perifērijas.

Literāro telpu kartēšana kļūst īpaši nozīmīga ģeokritikas pieejā. Tallijs, pētot telpas un vietas nozīmi daļēji literatūrā, telpu un literatūras mijiedarbību, tostarp interpretējot gan reālas, gan iedomātas telpas ar literārajām kartogrāfijām, ir izstrādājis pieeju, ko sauc par “ģeokritiku”. Tallijs norāda, ka ģeokritika ir ne vien veids, kā analizēt literāros tekstus, bet arī kā iespēja sociālajai kritikai: “ģeokritika, iespējams, var atklāt slēptās varas attiecības telpās, kuras citas kritiskās teorijas, kas ir mazāk vērstas uz telpas/telpiskuma izpēti, varētu nepamanīt” (Tally 2013: 114). Pievēršot uzmanību tam, kā naratīvi ne vien reprezentē, bet arī veido sociālās telpas, Eimija Velsa (*Amy Wells*), savienojot ģeokritiku ar feminisma kritiku paustajām *écriture féminine* un *parler femme* idejām, kas prasa pēc sieviešu valodas, lai izteiktu sieviešu pieredzi, ierosina konceptu *geo-parler femme*, feministisku ģeokritiku kā metodoloģiju, ar kuru identificēt un analizēt attiecības starp dažādām sociāli kulturālajām sieviešu identitātēm un vietu. Pamatojoties uz feministiskās naratoloģijas uzskatu, ka dzimums, dzimte un seksualitāte ir nozīmīgi ne tikai teksta interpretācijā un lasītāju uztverē, bet arī teksta poētikā un līdz ar to

teksta formās, struktūrās, reprezentācijas praksēs un komunikatīvajos kontekstos, Velsa savos pētījumos secina, ka “sievietes rakstnieces bieži izmanto ģeogrāfiju kā slepenu vai kodētu valodu, paužot emocionālas vai simboliskas nozīmes, kuras var saprast vien tie lasītāji, kuriem ir kopīgs konteksts ar autorēm”. Šie ģeogrāfiskie kodi jeb stratēģijas dod iespēju rakstniecēm izteikt jūtas un pieredzes, kuras viņas nespētu citādi izteikt. Turklāt bieži rakstnieces kodē šo ģeogrāfiju caur personīgu mijiedarbību ar dažādām telpām (Wells 2014; Wells 2017). Tā vietā, lai risinātu jautājumus, kā dzimtes identitāte ietekmē rakstnieka/rakstnieces darbības vietas izvēli tekstos, vai vīriešu un sieviešu tekstos parādās vienas un tās pašas vietas un vai vīrieši un sievietes šīs ģeogrāfiskās vietas apdzīvo atšķirīgi, *geo-parler femme* pieeja ļauj sašaurināt fokusu, koncentrējoties uz ģeogrāfiskajām stratēģijām, kas sieviešu tekstos tiek izmantotas kā stāstījuma stratēģijas.

Sieviešu prozas korpusa kartēšana

Kartējot rakstnieču literārās ģeogrāfijas, jāņem vērā laikmeta konteksts, kad sieviešu mobilitātes palielināšanās, sākot ar 19. gadsimta otro pusi, nāca kā modernitātes, urbanizācijas un industrializācijas sekas, iekļaujot arī izglītības un pārticības pieaugumu. Kopš deviņpadsmitā gadsimta otrās puses latviešu sievietes iekļaujas arī darbaspēka migrācijā, vai nu dodoties uz fabrikām, rūpnīcām pilsētās, vai caur savu darba devēju dzīvesvietas lokācijas maiņu. Pasaules karš un pārmaiņas, ko piedzīvoja Eiropas sabiedrība, papildus dekolonizācijai, globalizācijai un transporta attīstībai un ātrumam mainīja migrācijas plūsmu apjomu un raksturu Eiropā un virzienā uz Eiropu. Šī mobilitāte darbojās daudzējādā ziņā kā modernitātes zīme, un viens no pirmā viļņa feminisma panākumiem bija tieši sieviešu mobilitātes palielināšanās. Pārcelšanās darba vai ģimenes apstākļu dēļ, emigrēšana politisku vai ekonomisku faktoru iespaidā un ceļojumi aizveda sievietes pa pasaules ceļiem un pilsētām, un to pārvēršana prozā ienesa viņu balsis un viedokļus publiskā aprītē, veicinot sieviešu pārtapšanu no statiskām būtnēm par politikas un kultūras radītājām un sieviešu rakstību no ceļojumu stāstiem par kultūras aģenci (Schriber 1997: 7). Jau agrīnajā latviešu sieviešu rakstniecības vēsturē 19. gadsimta pēdējās desmitgadēs kā piemērus var minēt gan prozisti Mariju Valdemāri-Medinsku, kura savus darbus saraksta Tīgodas muižā Novgorodas guberņā, kur viņa vadīja brāļa saimniecību, gan pirmās ceļojumu aprakstu autore, kuru teksti tapa, mobilitātes ierosināti (Minna Freimane, Piliņu Katrīna, Marija Šmite u. c.).

Lai izpētītu un analizētu rakstnieču literāro darbu ģeogrāfiju, vispirms izveidojām tekstu korpusu un sagatavojām to analīzei, cenšoties tajā iekļaut pēc iespējas vairāk rakstnieču, kas savus darbus publicējušas gan grāmatās, gan periodikā laika posmā no pirmā latviešu valodā publicētā sievietes rakstītā prozas darba 1872. gadā līdz 1940. gadam. Tekstu korpusā iekļautas gan autores, kuras aplūkotojā laika posmā prozu publicējušas grāmatās, gan tās, kuru prozas darbi atrodami vien periodikā. Tekstu korpusā ietver 60 autores un 110 darbus. Abos žanros (romānā un īsprozā) līdz 1940. gadam darbojas: Anna Brigadere, Eiženija Tālmane, Lūcija Zamaiča, Marija Liepiņa, Emīlija Prūsa, Aida Niedra, Hilda Vīka, Austrā Ozoliņa-Krauze,

Tija Banga, Tālriņa, Marta Grimma, Tizmaliete, Ivande Kaija, Lidiņa Polis. Romānu autores: Paula Līcīte, Amanda Kloto (īst. v. Emīlija Jureviča), Emīlija Krišjāne, Emīlija Prūsa, Paula Balode, Elita (īst. v. Elza Miķelsone), Alma Dēķena, Vera Ziemeļniece (īst. v. Vera Pozņaks), Berta Pīpiņa, Aspazija, Herta Brēmane-Moora, Elza Stērste, Elvīra Kociņa, Erna Āboliņa-Martini, Rūta Skujiņa, Mare Bērziņa, Ilze Kalnāre, Elga Bite (īst. v. Vera Vanaga), Milda Pētersone. Autores, kas šajā laikā rakstījušas tikai īsprozu: Marija Valdemāre-Medinska, Hermīne Zālīte, Birznieku Latīņa, Birznieku Sofija, Anna Rūmane-Ķeniņa, Zemgaliešu Biruta, Antija, Alija Baumanē, Angelika Gailīte, Marta Altberga, Mirdza Bendrupe, Lija Bendrupe, Vilma Delle, Ausma Roga, Elvīra Vankina, Zelma Mežsēta, Ilona Leimane (romāns "Vilkaču mantinieci" publicēts 1942. gadā), Alise Stunda, Alma Zaprauska, Elza Ābele, Anete Stūre, Emīlija Ziņģe, Alida Roga, Marta Dakne, Emīra Rozīte, Ženija Pētersone, Felicita. Korpusā nav iekļauti pilnīgi visi šajā laika posmā publicētie sieviešu prozas teksti, taču autores tiekušas pēc iespējami plašākas tekstu daudzveidības (romāni, īsproza, gan grāmatās, gan periodikā publicētie darbi), lai korpusā ar vismaz vienu prozas vienību pārstāvētu visas autores, kas šajā laika posmā publicējušas savus tekstus. Tādējādi izveidotais korpusa kopumā vispusīgi atspoguļo šī perioda latviešu sieviešu prozas situāciju.

Lai atlasītu informāciju un kartētu ģeogrāfisko telpu, tekstus vispirms rūpīgi izlasījām un sagatavojām analīzei. Barbara Pjati (*Barbara Piatti*), viena no vadošajām digitālās kartēšanas teorētiķēm, piedāvā piecas kategorijas: darbības vieta, darbības zona (visas vai vairākas teksta darbības vietas kopā), projicētā vieta (sapņi, atmiņas ilgās), marķieris (vieta minēta tekstā, bet neatbilst nevienai no iepriekšējām kategorijām) un maršruts (pārvietošanās) (Piatti et al. 2009). Saskaņā ar mūsu izveidoto sistēmu izmantojām trīs no tām: darbības vieta, projicētā vieta un marķieris. Atsevišķi izdalīti darbi, kuros ģeogrāfiskās vietas nav minētas vispār, pārsvārā tie ir īsprozas darbi. Darbiem, kuros vietvārdi neparādās, un tiem, kuros darbība notiek iedomātās vietās, tāpat arī pārvietošanās maršrutiem šis raksts tuvāk nepievēršas.

Respektējot Pjati argumentu, ka "tā jau ir daļa no sagatavošanās: literatūras ģeogrāfiskās vietas attēlojošu vizualizāciju izstrādes process ir tulkošanas process. Viena simbolu kopa – teksts – tika pārtulkota uz citu simbolu kopu – kartes simboliem" (Piatti et al. 2009), kartēšanas procesam pievērsām īpašu uzmanību. Izmantojām vienu no vienkāršākajiem rīku komplektiem, datus ievadot *Google Sheets* un kartes veidojot *Google My Maps*. Balstoties uz izveidoto tekstu korpusu, strukturējām datus, kas attiecas uz ģeogrāfiski kartējamām (reālām) vietām, vairākos līmeņos (kur tas bija iespējams, proti, minēts tekstā): no konkrētas valsts līdz pilsētai un adresei, kur tā bijusi minēta. Ja viena un tā pati vieta vienā tekstā minēta atkārtoti, atzīmējam to tikai vienu reizi, lai kartē parādītos nevis pieminējuma biežums kādā darbā, bet gan unikāli minēto vietu blīvums. Dati ir strukturēti: autore, darba nosaukums, publicēšanas gads (pirmpublicējuma) un saistībā ar teksta kartējumu: valsts, pilsēta, adrese, citāts. Izmantojot ievadītos datus, veidotas kartes, kas sniedz pārskatu par sieviešu rakstītās literatūras ģeogrāfiju. Pēc tam katrai atrašanās vietai piešķirta īpaša atzīme un noteikta krāsa. Vietas, kuru apzīmējums tekstā bijis atšķirīgs no mūsdienu nosaukuma, apzīmētas ar nosaukumu, kāds tiek izmantots šodien (un kuru atpazīst ĢIS): piemēram, Pēterpils – Pēterburga – Sv. Pēterburga – izmantota Sanktpēterburga, mūsdienu vietvārds, tāpat arī ar ielu un iestāžu nosaukumiem, citātā paturot teksta variantā minēto.

Papildus kartes automatiskai izveidošanai digitālā kartēšana sniedz arī citas priekšrocības. Slāņu veidošana ļauj vienā kartē atlasīt vairākas datu grupas, piemēram, palūkoties uz īsprozu pirms un pēc Pirmā pasaules kara vai sastatīt konkrētas autores romāna karti ar viņas biogrāfisko karti. Vēl viena iespēja ir ievadīt metadatus un kartes pievienot kā hipersaites, papildinot ar attēliem vai teksti. Nākotnē plānots izmantot arī maršruta rīku, lai dažādās atrašanās vietas tiktu savstarpēji saistītas un nodrošinātu hronoloģisku autora dzīves vai literārā teksta notikumu attēlojumu. Iespēja attēlot autores literāro darbu un/vai dzīves reālo ģeogrāfisko vidi, atspoguļojot viņas bieži apmeklētās vietas, mīļākās vietas un vietas, kur notikuši būtiski notikumi rakstnieces dzīvē.

Tekstu korpusa kopīgajā kartē (skat. 1. attēlu) redzams sieviešu prozas ģeogrāfiskais plašums. Salīdzinot īsprozas un romānu ģeogrāfisko telpu (skat. 2. un 3. attēlu), iezīmējas ļoti līdzīga aina. Visumā sieviešu rakstītajos tekstos iezīmējas diezgan plaša teritorija, iekļaujot gan laukus un pilsētas, gan Krievijas Impēriju un Rietumeiropu. Pēc Pirmā pasaules kara, uzplaukstot Latvijas valstij kā līdzvērtīgai starp citām Eiropas un pasaules valstīm, mobilitāte kļūst neierobežota, ja to ļauj materiālie apstākļi. Arī sieviešu ceļošana, kultūras baudīšana vai studēšana ārzemēs vairs nav kas īpašs. Lielākoties sieviešu mobilitātes iemesli ir studiju iespējas (Pēterpils, Drēzdene), ceļojums (Rietumeiropa: Itālija, Šveice, arī Vācija, tai skaitā arī kāzu ceļojums un došanās ārstēties), kā arī bēgļu gaitas 1. pasaules kara laikā (Krievija). Taču, palūkojoties uz prozas kartējumu, jāsecina, ka darbība lielākoties norisinās Latvijā, visbiežāk Rīgā un Jūrmalā. Iespējams, tas daļēji izskaidrojams ar sižetu, kur centrālie ir varoņu attiecību, visbiežāk mīlestības jūtu, sarežģījumi, un ģeogrāfiskā vieta neieņem būtisku lomu. Tāpat arī plaukstošā Latvijas vidusšķira ar tās izpriecām, modernām kafejnīcām, bāriem, atpūtu jūrmalā un izrīkojumiem ir pateicīgs fons romānu konfliktiem.

Sieviešu prozas ģeogrāfijas pamatu veido Latvijas teritorija (vairāk nekā 200 pieminējumi). Tekstos skaidri nolasāms ir pavērsiens pilsētas virzienā, darbība lielākoties notiek pilsētvidē – dominē Rīga (114) un Jūrmala (16, vēl atsevišķi arī Lielupe, Bulduri, Dzintari (Edinburga), Majori, Asari). Minēta arī Jelgava, Sigulda, Cēsis un Valmiera. Citas konkrētās vietas Latvijas teritorijā prozas tekstos ienāk retāk, ir vispārīgāki novada iezīmējumi, dominē Kurzeme, Vidzeme, Zemgale.

Nemot vērā, ka rakstnieču prozā dominē Rīga, tā arī tekstos tiek visprecīzāk kartēta. Piemēram, minētas konkrētas ielas, starp biežāk minētajām ir Elizabetes iela, Kaļķu iela, Aspazijas bulvāris, Raiņa bulvāris, Bastejkalns, Brīvības bulvāris, Stabu iela, tāpat minēta arī Pārdaugava, Torņakalns un Jugla. Taču visumā dominē Rīgas centrs: teātri, kafejnīcas, konkrēti centra nami, arī tādas vietas, par kurām šodien var uzzināt tikai Latvijas Nacionālās bibliotēkas projektā “Zudusī Latvija” vai atsevišķās publikācijās, kā restorāns “Alhambra” un citi.

Krievija tekstu korpusā parādās 55 reizes, un lielākoties tā minēta vai nu kā abstrakta, plaša teritorija, iekļaujot tajā, iespējams, visu Krievijas Impēriju, vai arī divas konkrētas pilsētas – Sanktpēterburga (Pēterburga, Pēterpils – 13) un Maskava (12), ienāk arī konkrētas vietas, kā muzeji, parki, promenāde. Krievija kā darbības vieta izteiktāk parādās pirms 1. pasaules kara

vai tekstos, kas atspoguļo notikumus pirms kara un tā laikā. Amerikas Savienotās Valstis piesauktas kā vieta, kur kāds dodas, kā eksotika, bet nav saistāma ar galvenajiem tēliem vai pašu autoru pieredzi. Eksotiskas vietas, kā, piemēram, Šanhaja Ķīnā izmantota kā darbības vieta romānā, kas uzskatāms par populāro (lubu) literatūru un ir izņēmuma gadījums. Tomēr šajā publikācijā akcentēta “skatīšanās Rietumu virzienā” – Rietumeiropas valstis tekstos minētas bieži, no tām – Vācija (31), Itālija (54), Francija (34), Šveice (24), Austrija (12). Rietumeiropas pilsētas, kas parādās tekstos, – Prāga (33), Parīze (21), Ospedaleti (4), Drēzdene (6), Florence (5), Roma (9), Cīrihe, Berlīne (12), Vīne (9). Ne mazums prozas darbos autore sīžeta centrā vai fonā ievij dažādas Eiropas valstis un vietas, un viena no acīmredzamākajām iezīmēm ir paralēles starp rakstnieču biogrāfiju un literāro ģeogrāfiju. Autore pamatā izvēlas tēlot vietas, kur pašas ir bijušas.

Īsproza

Rakstā analizēti īsprozas tekstu korpusā iekļautie 80 teksti (stāsti, tēlojumi, skices), kuros minēti vietvārdi. Ņemot vērā, ka romānus latviešu rakstnieces sāka publicēt vēlū, pirmajā sieviešu rakstniecības vēstures posmā tieši īsproza sniedz priekšstatu par literāro ģeogrāfiju. Pirmais latviešu rakstnieces prozas darbs grāmatā izdots 1872. gadā: Marijas Medinskas-Valdemāres lokalizētais stāsts “Ozols, viņa dzīve un gals”, kam seko oriģinālstāsts “Zemnieks un muižnieks” (1877). Stāstu darbība norisinās Kurzemē, un sievietes šo vietu nepamet. Tie, kuri pārvietojas, ir vīrieši, piemēram, dodoties studēt uz Sanktpēterburgu. Salīdzinājumam Hermīnes Zālītes 19. gadsimta 90. gados periodikā publicētajos tekstos studēt dodas jau sievietes (piemēram, uz Drēzdeni studēt mūziku vai uz Parīzi studēt medicīnu), mobilitātei un sieviešu tieksmei pēc izglītības iezīmējot arī pārmaiņas sieviešu situācijā, kas sabalsojas arī ar plašāku autoru sieviešu ienākšanu literārajā telpā: tekstus sāk publicēt Birznieku Latīņa, Anna Brigadere, Tīrzmaliete, nedaudz vēlāk arī Emīlija Prūsa u. c., arvien paplašinoties autoru lokam.

Lai arī stāsta žanrs pēc būtības uzskatīts par “neģeogrāfisku,” jo salīdzinājumā ar romāniem lielā daļā īsprozas tekstu vietvārdi netiek minēti vispār un darbības norises vieta tajos ir nolasāma no konteksta; stāsts arī nespēj sniegt tādu ģeogrāfisko materiālu, kāds atrodams romānos, kur iespējami plaši izvērsti vietu apraksti, kas ļauj, piemēram, izdarīt secinājumus par telpas ietekmi uz tēlu attīstību. Reizē šis īsprozas acīmredzamais “ģeogrāfiskais trūkums” var kļūt par priekšrocību – stāstā iespējams ļoti ierobežots vietu skaits un galvenokārt tie ir vispārīgi vietu pieminējumi. Tieši tādēļ, ka ir nespecifiski un viegli atpazīstami, lielākoties arī stereotipiski, tie spēj funkcionēt tematiski (Hones 2010; Brosseau 2017). Piemēram, Sanktpēterburga, Drēzdene, Parīze visbiežāk parādās kā izglītības vietas, taču, ja tekstā minēta Itālija, tā saistīsies ar atpūtas un mākslas baudīšanas ceļojumiem.

Kad 1902. gadā “Pēterburgas Avīzēs” parādās nelieli liriskas noskaņas tēlojumi, pie vairākiem tēlojumiem atzīmēts arī to sacerēšanas laiks un vieta: 1901. gada novembris un decembris,

Ospedaleti. Tēlojumu autore ir Anna Ķeniņa, kura 1901. gada nogalē dienvidos, Itālijas kūrortā Ospedaleti, uzturējās veselšanās nolūkā. Šīs vietas nosaukums ir mazzināms un citu rakstnieču tekstos neparādās, tādējādi biogrāfiskā sasaiste literatūras kartē iezīmē ļoti konkrētas, bieži vien tikai autoras piedzīvotas teritorijas. Apjomā nelieli teksti, kuros pārdomas un vērojumi mijas ar fantāziju un kuru nosaukumi jau liecina par cikla kopīgo noskaņu – “Dzīvot gribas”, “Sauli noejot”, “Garāmejot”, “Kā dienvidu priedīte mira” u. c. –, raksturīgi ar tajos paustajām pārdomām par dzīves īslaicīgumu un nāves vienmēr klātesošo tuvumu (Rūmane 1902). Spēcīgu noskaņu rada krāšņās, mirdzošās un vitālās Itālijas dabas klātesamība, kas pretstatīta slimīgajiem, fiziski izvārgušajiem un dzīvot alkstošajiem cilvēkiem, tādējādi sniedzot vietas subjektīvo uztveri.

Lielākoties rakstnieču īsprozā nav izstrādātu vides tēlojumu, nereti autorēm pievērsties romantiski impresionistiskiem jūtu tēlojumiem, kvēlas kaislības un izmisums norisinās specifiskā, nereti stereotipiskā vidē, kura kalpo kā darbības fons, kas viegli atpazīstams lasītājiem, taču var ietvert arī dziļāku un personīgāku nozīmi autorei pašai. Zemgaliešu Biruta (īstajā vārdā Konstance Grencione) 1906. gadā, sākoties reakcijai pēc 1905. gada notikumiem, kuros iesaistījusies arī jaunā autore, dodas bēgļu gaitās “caur Pēterburgu uz Helsinkiem, marta vidū – pāri jūrai, mēneša beigās viņa jau ir Ķelnē un dodas tālāk uz Parīzi”, kas uz jauno autori atstāj neizdzēšamu iespaidu (Vērdiņš 2005: 87). Parīze atkārtoti parādās vairākos viņas īsprozas tekstos kā atbrīvojoša telpa un viena no nedaudzajām konkrēti minētajām darbības vietām.

Saistot autoras un viņas darbu ģeogrāfiju, jāpiemin arī Angelika Gailīte, kura no 1910. līdz 1916. gadam dzīvojusi Vologdā, Krievijā, kur tapusi daļa tekstu viņas pirmajam prozas krājumam “Ilgas un maldi” (1913). Lai arī šajā krājumā konkrēti vietvārdi nav minēti, vietas aspektam tekstos ir būtiska nozīme; parādās Lielais dārzs, dzelzceļa pietātnē, jūra, kapsēta, piejūras ciemats dienvidos, kalni, arī dzimtene, vieta, kas cieši saistīta ar vēstītājas melanholisko noskaņojumu. Šis piemērs parāda to, ka ne vienmēr būtiskas vietas tiek apzīmētas ar vietvārdiem un to nozīme atklājas kontekstā, kuru bez rūpīga tuvlasījuma nav iespējams atklāt. Pēcāk Gailīte daudz ceļojusi un viņas vēlākajos tekstos bieži ienāk sieviešu ceļojumi Rietumeiropā (stāstos “Kāzu ceļojums”; “Sapņotājas”, abi 1927). Autobiogrāfiski motīvi parādās stāstā “Cor ardens” (1927), kura sākumā Maija Līdumniece tikko atgriezusies Rīgā no Pēterpils, kur studējusi konservatorijā, “un pēc studiju beigšanas bij atstāta pie tās par pasniedzēju, kāpēc tik ilgi nebij gribējusi aizbraukt no šīs izmirstošās un izsalstošās pilsētas” (Gailīte 1927). Gailīte, kura kopš 1916. gada strādājusi par priekšnieces palīdzi Marijas meiteņu institūtā Maskavā, lai paralēli darbam iegūtu augstāko izglītību Filoloģisko un juridisko augstāko sieviešu kursu vēstures nodaļā, 1920. gada sākumā atgriežas Latvijā. Arī Alijas Baumanes prozas darbos kā norises vieta parādās Pēterburga (“Auļojošā pilsēta” 1924; “Lietaviete” 1935), kur rakstniece pati no 1914. līdz 1920. gadam dzīvojusi un strādājusi par ierēdni, kā arī pasniegusi privārstundas, lai nopelnītu naudu filoloģijas un filozofijas studijām, šajā laikā piedzīvojot arī vēsturiskos notikumus. Sieviešu īsprozas literārās kartes ir cieši saistītas ar pašu autoru pārvietošanās un dzīvesvietu ģeogrāfiju, kā arī bieži ar vietas ķermenisko pieredzi, ko iespējams atklāt tikai tuvlasījumā.

Romāni

Romānus latviešu rakstnieces sāk publicēt vēlū – pirmais romāns, Paulas Līcītes “Lida Erik”, publicēts 1909. gadā, tam seko Tīrzmaliertes “Zem dzimtenes debesīm” (1911) un Ivandes Kaijas “Iedzimtais grēks” (1913). Intensīvākais romānu rakstīšanas un publicēšanas laiks ir, sākot no 1927. gada, kad lasītājus sasniedz seši jauni romāni, 1928. – četri, 30. gados to skaits pieaug vēl intensīvāk, ik gadu parādās 4–7 romāni, 1938. gadā pat astoņi. Ražīgākās romānistes ar vismaz trim romāniem ir Ivande Kaija, Lūcija Zamaiča, Aīda Niedra, Emīlija Prūsa, Elita, Vera Ziemeļniece, Elvīra Kociņa un Paula Balode.

Kartēšanai atlasīti 30 romāni. Kopumā romānu korpusu veido apmēram 80 romāni, no kuriem 18 publicēti tikai periodikā. Tas daļēji izskaidrojams ar to, ka jau tapšanas gaitā romāni nodomāti publicēšanai turpinājumos sievietes auditorijai paredzētajos žurnālos kā “Sievietes Pasaule”, “Zeltene” vai populārāajā presē (“Atpūta”, “Pēdējā Brīdī”). Pirmais zināmais romāns, kura autore ir sieviete un kas literatūrvēsturē marķēts kā romāns, latviešu valodā publicēts 1909. gadā, tas ir rakstnieces un komponistes Paulas Līcītes romāns “Lida Erik” ar apakšvirsrakstu “Kāda cilvēka dienasgrāmata” (iezīmīgi, ka sieviete ir arī izdevēja Berta Ēķis, kas ir re-tums tā laika grāmatniecībā). Salīdzinot ar publicēto rakstnieku vīriešu romānu hronoloģiju, 1909. gads ir vēlīns, proti, kopš 1873. gada, kad publicēts pirmais latviešu valodā sarakstītais oriģinālromāns, grāmatās un presē latviešu lasītāju sasnieguši vairāk nekā 70 romāni, kuru autori ir vīrieši.

Pirmo trīs romānu (Līcītes “Lida Erik”, Tīrzmaliertes “Zem dzimtenes debesīm” un Kaijas “Iedzimtais grēks”) kartējums raksturo dažādo pasaules uztveri un sievietes pozīciju līdz Pirmajam pasaules karam. Tīrzmalierte iezīmē paaudžu atšķirības, uzsver savas paaudzes pie-ķeršanos dzimtenes zemei, pretstatot jauno paaudzi, kas dodas studēt uz Krieviju vai strādāt uz Ameriku bez sirdssāpēm par aiziešanu no dzimtās vietas, viņas romāna darbības vieta ir tikai Latvijas lauki. Savukārt Paulas Līcītes romānā Eiropas pilsētu pievilksanas spēku sevī iemieso varonis vīrietis, kura dienasgrāmatas ieraksti ir romāna pamatā. Viņam nemitīgi ir vilkme doties, ceļot, baudīt Parīzi, Berlīni, Venēciju, Florenci, Mīnheni, Romu. Savukārt Lida Erik, daudzsološa mūziķe un dziedātāja, klausās šī varoņa, sava mīlnieka, stāstos, bet pati lielākoties paliek inerta un savai pazīstamajai videi piesaistīta. Tā nav nejaušība, jo Paula Līcīte šo romānu uzraksta uzreiz pēc ģimnāzijas beigšanas 1907. gadā, atsaucoties uz žurnāla “Stari” izsludināto konkursu. Viņa vēl nav bijusi ārpus Latvijas, romāna pamatā ir iztēle un laikrakstos, žurnālos un dailliteratūrā iedvesmoti motīvi un vietu apraksti. Varoņa monologi ir tam spilgts piemērs: “Esmu bijis Berlīnes nacional-galerijā, kur nosēdēju veselām dienām krāsu karaļa Bocklina darbu priekšā... Pārstaigāju Harcu, ieelpoju svaigo sveķaino smaršu, apbrīnoju jautro Ilzi, par kuru Heine dzied tik brīnišķīgi... Biju Parīzē, – mans pirmais gājiens bij uz Luksemburgas muzeju, kur atrodas Rodēna darbi... Esmu iepazinies ar Parīzes nervozo, pārsmalcināto dzīvi, kurai pat dienā ir krēslaina nokrāsa, – bet vakarā viņa mirdz kā neīsts zelts elektriskā gaismā ... Esmu dažu labu nakti nosēdējis līdz ar biedriem restorānos, kafējnīcās un klausījies, ko man stāsta krāsotas lūpas un apvilktas acis... Biju Drezdenes ainu galerijā, kur sajūsminājos par Rafaela madonnām... Bet visilgāk uzkavējos Itālijā... Gribēju pilnīgi izjust viņas mirdzošo, pasakaino burvīgumu...” Paula Līcīte caur Līdu Erik un viņas mīļoto

parāda konfliktsituāciju, ko veido tēlu dzimumnosacītās, dažādās dzīves uztveres un vēlmes, dažādie ģeogrāfiskie un kultūrtelpas horizonti. Šo skaidro arī Sintija Enlo (*Cynthia Enloe*): “Daudzās sabiedrībās sievišķība ir definēta kā turēšanās pie mājām. Turpretim vīrišķība ir bijusi ceļošanas pase. Feminisma ģeogrāfi un etnogrāfi ir uzkrājuši pierādījumus, kas atklāj, ka galvenā atšķirība starp sievietēm un vīriešiem neskaitāmās sabiedrībās ir bijusi atļauja ceļot prom no vietas, kas tiek uzskatīta par “mājām” (Enloe, 2000: 21). Līcītes romānā sieviete, talanta ziņā līdzvērtīga, ir tēlota šajā pasīvajā lomā, un “ārzesmes” tekstā kļūst par lokāciju, kas iemieso vīrieša sajūsmu un piedzīvojumus.

Savukārt Ivande Kaija līdz pirmā romāna tapšanai savu redzesloku jau bija krietni paplašinājusi ārvalstīs, klausījusies filozofijas lekcijas Bernes, Leipcigas un Sorbonnas Universitātē, apceļojusi Eiropu. Šī pieredze ienāk arī viņas romānos, varones dodas studēt, ceļo uz ārzemēm ar alkām redzēt Eiropas kultūrvietas, arī kāzu ceļojumā. Tieši Ivandes Kaijas romāni nodrošina, ka biežāk piesaukto valstu un pilsētu saraksta augšgalā ir Itālija, Šveice, Austrija un attiecīgi Venēcija, Cīrihe, Tirole, turpat arī Vācija un Krievija ar Berlīni un Maskavu. Jauna sieviete, kas dodas studēt, ir būtisks tēls sieviešu prozā. Studente tiek uzskatīta par strīdīgu kultūras simbolu, jo viņa vienlaikus iemieso gan sieviešu emancipācijas centienus, gan vīriešu bažas par sociālajām pārmaiņām un modernitāti, kas transformē tradicionālo sievietes lomu, maina jaunas sievietes dzīves prioritātes (Mazon, 2000). Intelektuālās vienlīdzības centieni rada atbalsis ne tikai vīriešu autoru satīrās, studentes tēls tika sentimentalizēts vai trivializēts arī sieviešu autoru sižetos. Īpaši raksturīgi tas ir 20., 30. gados, kad sievietēm studijas līdz ar Latvijas Universitātes atvēršanu ir daudz pieejamākas, un stāstu/romānu galvenais fokuss no pašām studijām novirzās uz studentu dzīves piedzīvojumiem, iemīlēšanos, nejaušu grūtniecību, pašnāvības mēģinājumiem u. c.

Elvīra Kociņa, tāpat kā Aīda Niedra un vēl vairākas autore, netiecas romānos precīzi nosaukt vietas, iztiekot ar dalījumu “pilsēta” un “lauki” vai izdomājot savus, dabā neeksistējošus nosaukumus vietām, taču, ja romāna darbība iesniedzas ārpus Latvijas, autore kļūst konkrētāka un ģeogrāfiski detalizētāka. Kociņas romāns “Pēdējā laime” (1937) ir vēl viens piemērs sieviešu ceļojumu motivācijai – ceļot, lai mainītu dzīvi, lai aizbēgtu no ikdienas, vilšanās, rutīnas. Romāna galvenā varone Elga strādā monotonu mašīnrakstītājas darbu ministrijā un ir saderinājusies. Sarežģījums sākas līdz ar līgavaīņa vēstuli, kas vēstī, ka viņš Elgu pamet. Pārdzīvojumu iespaidā saasinās tuberkuloze, un ārsts iesaka doties uz sanatoriju. Varone, izmantojot kāzām un kāzu ceļojumam nodomātos ietaupījumus, dodas cauri Eiropai uz Bulgārijas pilsētu Varnu Melnās jūras krastā. Epizode, kurā Elga izlemj, ka nedosies uz konkrētu sanatoriju, bet brauks, kur ir saule, silta jūra, mūzika un smiekli: “Mana sirds iepukstas priekā. Man vēl drusku atlicis prieka. Drudzaini meklēju Eiropas karti. Uz Franciju, Itāliju, nē – gribas uz austrumiem, kur vairāk noslēpumu. Braucu ar pirkstu pa karti līdz Melnai jūrai. Bulgārijas krasts. Kūrorts Varna. Turp var braukt caur Bukaresti līdz Konstancei, ar kuģi pa Melno jūru uz Varnu. Pēc vasaras atgriezties caur Istambulu, Sofiju, Budapeštu.” Šo iecerēto maršrutu Elga arī veic, autore nosauktās vietas rūpīgi ievijusi romāna sižetā. Kā atzīmējis Jānis Klīdzējs, Elvīra Kociņa tajos gados pati ir daudz ceļojusi pa Eiropu, Sorbonnā kādu laiku studējusi mākslas vēsturi, un “savus romāna varoņus, latviešu cilvēkus Elvīra Kociņa ievēd Eiropas kultūras ainavā un domu līmenī. Daži no viņas tēlotiem cilvēkiem romānos ir citu tautu piederīgie, un tie tiek saistīti ar latviešu cilvēku gaitām pār dzīvojumu kopībā” (Klīdzējs, 1972).



1. attēls. Tekstu korpusa kopīgā karte.

Līdzīgi arī Hilda Vika, kura savos stāstos gandrīz nekad nav pārkāpusi Latvijas robežas, romānā “Laimīgā dziesma” (1939), tēlodama kādas latviešu ekskursantu grupas piedzīvojumus Itālijas ceļojumā, ievij daudz no pašas redzētā un piedzīvotā ceļojumā uz Itāliju 1938. gadā, no kura pārbrauc pilna jaunu ideju gan glezniecībā, gan rakstniecībā. Viens no rezultātiem ir šis romāns, kurā Itālijas dabas un mākslas iespaidi kalpo par fonu spilgtiem cilvēku tēliem. Kā norāda Lidija Pērļupe recenzijā: “Te varētu minēt smagos traģiskos pārdzīvojumus krāšņajā, zeltainajā Palermo pilsētā vai humorisko, viegli drastisko kauju Dženovas kapsētā, kur Varduks bildina Mariannu; neaizmirstama ir arī Anitras iekšējā pārtapšana, kļūstot pa Monrealas domu un raugoties slavenajos mākslas darbos” (Pērļupe 1942). Itālijas ceļojums iedvesmojis arī Emiliju Prūsu, romānā “Itāliete” (1927) viņa tēlojusi kādas Rīgā pārvestas itāliešu meitenes likteni un mīlas pārdzīvojumus, precīzi atsaucoties uz apmeklētajām vietām un objektiem. Gandrīz visos gadījumos sieviešu prozā ienāk lokācijas, kuras ir konkrētās valsts vai pilsētas apskatīšanai ieteicamo vietu sarakstā, kas saskan ar to, ko secinājusi Līsbeta Larsone (*Lisbeth Larsson*), analizējot Virdžīnijas Vulfas literāro ģeogrāfiju. Proti, Vulfa, atšķirībā no Čārlza Dikensa, izvēlas Londonu parādīt caur atvērtām, drošām, atpazīstam vietām, netēlojot netikamus un neuzticamus rajonus un nomales. Larsone to sasaista ar Jaunās sievietes konceptu un pilsētu kā radošu un uz jauninājumiem (inovāciju) vērstu, izvirzot tēzi, ka vieta var apturēt (*suspend*) iedibināto dzimumu kārtību, diskrimināciju un izslēgšanas mehānismu, piedāvājot dažādību un iespēju kaut kam jaunam (Larsson, 2017: 12). Var secināt, ka romānos rakstnieces izvērstāk, ļaujoties literārajai fantāzijai un nereti autobiogrāfiskajai pieredzei, apceļo dažādas vietas un objektus, detalizēti iedarinot sieviešu tēlos ar konkrēto telpu saistītās emocijas, fizisko un intelektuālo tvērumu.



↑ 2. attēls. Romānu korpusa karte.

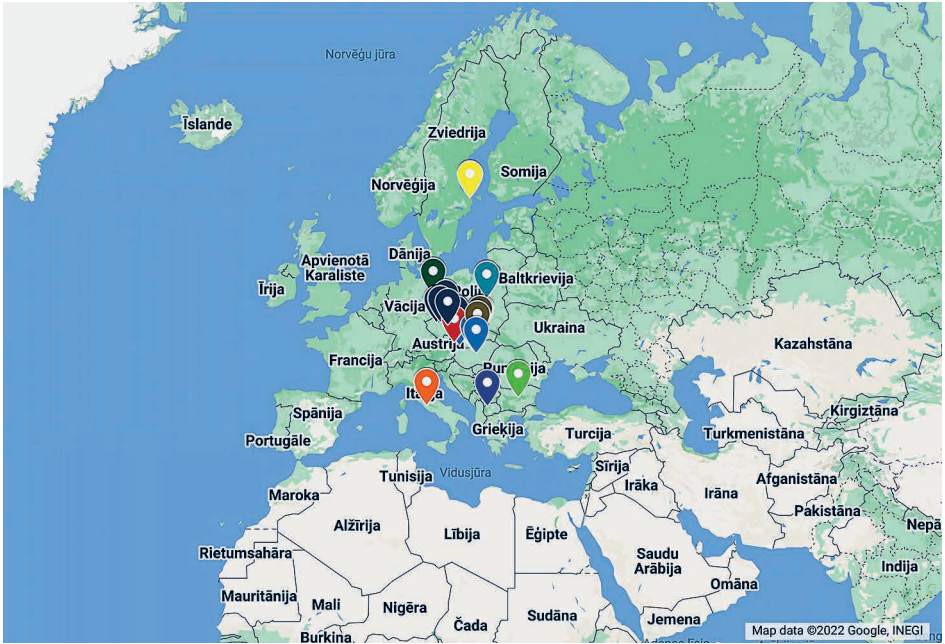
↓ 3. attēls. Īsprozas korpusa karte.

Mobilitāte, literārā ģeogrāfija un autobiogrāfiskums: Martas Grimmas piemērs

Izmantojot digitālās tehnoloģijas un rīkus, mēs varam veidot digitālas kartes un noteikt vietvārdu skaitu, taču tikpat būtiska ir spēja atklāt niansētās attiecības starp ģeogrāfiski telpisko modeli un teksta nozīmi – tādēļ būtiski ir interpretēt darbību vietu nozīmes tekstos. Piemēram, vietvārdi, veidojot stāstījuma un ģeogrāfisko telpu, vienlaikus rosina un veicina konotācijas, kas radušās plašākā kultūras aprītē (Heuser 2016: 27) (piem., Parīze, arī Berlīne, Sanktpēterburga) un kurām iespējams piekļūt ar tuvlasījumu, tāpat telpiskajiem elementiem un struktūrām ir dažādas funkcijas tekstā (piem., nosaucošā, emociju blīvumu raksturojošā), kuras iespējams atšķetināt vien tuvlasījumā.

30. gados Latvijas periodikā bieži redzams Martas Grimmas (1901–1983) vārds, viņa raksta daudz – gan dzeju un prozu, gan ceļojumu aprakstus un recenzijas. Grimma pazīstama arī kā pirmā latviešu un čehu kultūrsakaru īstenotāja (Zeile 1993), čehu literatūras tulkotāja, pētniece un popularizētāja (Štolls 2015: 119). Pēc studijām Latvijas Universitātes Filoloģijas un filozofijas fakultātē Romāņu un ģermāņu filoloģijas, vēlāk Baltu filoloģijas nodaļā (1920–1926), paralēls subasistentes darbs pie nodaļas dekāna profesora Jāņa Endzelīna, kam seko laulības ar Endzelīnu, un piedzimst meita Līvija. 1929. gadā Kultūras fonda piešķirtā 500 latu studiju stipendija Grimmai paver ceļu uz Prāgas Kārļa universitāti, kur viņa studē ģermāņu, romāņu un slāvu salīdzināmo literatūrzinātni (filozofijas maģistres grāds 1931. gadā). Vēlāk studiju stipendiju piešķir arī čehi. Grimma savās atmiņās raksta: “1929. gada maijā ar sažņaugtu sirdi iekāpu Zemgales vilcienā, lai dotos caur Varšavu uz Prāgu, jo Latvijā man palika vienīgā dvēseļīte visā pasaulē – maza saslimusi meitiņa. Tagad vajadzēja dienu un nakti strādāt, lai aizmirstu bērna prombūtni” (Grimma 1935). No Prāgas viņa apeļo Centrāleiropu, un gan Prāga, gan apmeklētās vietas ienāk Grimmas literārajos darbos – dzejā, stāstos un romānā. Šajā raksta daļā, pievēršot uzmanību Grimmas prozā attēlotajām vietām, salīdzinām rakstnieces Čehijas perioda biogrāfisko karti (skat. 4. attēlu), kuras pamatā ir ceļojumu aprakstu grāmata “Čehoslovākijā” (Autores izdevums, 1933), Grimmas atmiņas “Manas studijas ārzemēs” (Latviete, Nr. 5, 1935) un “Mans dzīves stāsts” (Karogs, Nr. 11, 1981) un viņas literāro darbu karti (skat. 5. attēlu), kāda tā parādās stāstos (krājumā “Kad Donava šalc”, autores izdevums, 1934) un romānā (“Sirds saulgrieži”, Sievietes Pasaule, 1938). Lai arī ģeogrāfija tiek uztverta kā reāla telpa, turpretī rakstnieces radītā telpa kā izdomāta, jo “telpas, ko rada rakstīti vai runāti vārdi, ir iedomātas pasaules, pat ja tās atsaucas uz fiziskām telpām” (Wells 2005: 80), ģeokritika aicina jautāt, kāpēc šīs vietas ir svarīgas. Salīdzinot biogrāfisko un literāro karti, sievietes tēlus un viņu attiecības ar vietām, kurās viņas darbojas un caur kurām viņas pārvietojas, iezīmējas paralēles, kas ļauj izdarīt secinājumus par konkrēto vietu nozīmi Grimmas tekstos. Tieši pārvietošanās, mobilitātes iespējas, attēlojot sievietes telpas paplašināšanos un vēlmi pēc sociālās vienlīdzības, Grimma ieraksta savu *geo-parler femme*.

Lasot Grimmas ceļojumu aprakstu grāmatu “Čehoslovākijā” un viņas publicētās atmiņas periodikā, iespējams izsekot vismaz daļai no vietām, kuras rakstniece apmeklējusi. Studiju laikā Prāgas Kārļa universitātē Grimma dzīvojusi Prāgas rajonā Žižkovā un labi pārzinājusi



↑ 4. attēls. Martas Grimmas biogrāfiskā karte.

↓ 5. attēls. Martas Grimmas prozas korpusa karte.

arī citus pilsētas rajonus un nozīmīgākās celtnes, kā arī vērojuši Prāgas izplešanos. Ceļojumu aprakstā Grimma raksta: “Kad maijā iebrauc Prāgā, tad vēl attālos kvartālos – Bubenčos, Dejvicē, Žižkovā u. c. bija tikai lauki un tīrumi. Tagad – atgriežoties – atradu tur veselās jaunās pilsētas, kas stiepās līdz attālai Zbraslavai u. c. pilsētām.” (Grimma 1933) Šajā laikā viņa Čehiju un Slovākiju izbraukājusi “krustām, šķērsām, apmeklējot kādas 20 pilsētas” (tai skaitā Turnovu, Pilzeni, Marienbadi u. c.) (Grimma 1933). Tāpat Grimma pabijusi gan Austrijā un Vācijā, gan Itālijā un Ungārijā. Taču, kā viņa raksta savās atmiņās, ceļojumu aprakstā nav bijis iespējams atklāt emocijas, kādas pārdzīvotas šajā laikā, atzīstot, ka emocionālos stāvokļus, lai arī ne pilnīgi, tomēr vieglāk bijis izteikt literatūrā (Grimma 1981). Grimma prozā izmanto arī ģeogrāfiskas vietas kā stāstījuma stratēģijas, atklājot emocijas un jūtas, kuras nav bijis iespējams izteikt citās formās.

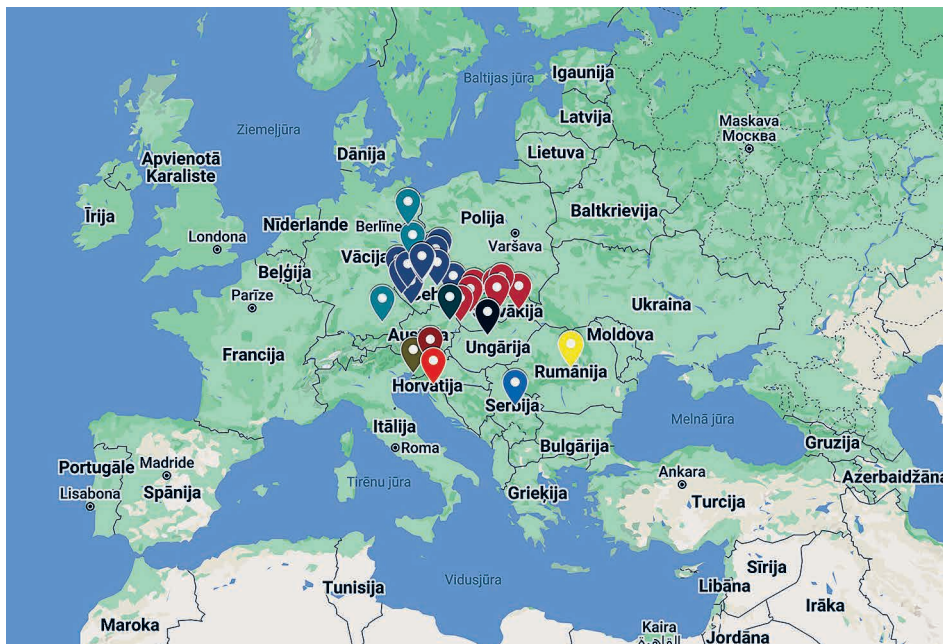
Krājumā “Kad Donava šalc” iekļautajos stāstos darbība lielākoties norisinās ārpus Latvijas. Elza Grāvīte recenzijā par krājumu raksta, ka svešo zemju izteiktā klātbūtne “pat it kā aizēno dzimtenes skaistumu un mātes valodas saulainību” (Grāvīte 1934). Galvenās varones ceļo ne vien maršrutā Rīga–Varšava–Prāga – divos no sešiem stāstiem tās līdzīgi kā pati autore brauc uz Prāgu papildināties studijās –, bet arī apmeklē Itāliju, dodas uz Alpiem un Tatriem Slovākijā, iezīmējot maršrutu, pa kuru pārvietojusies arī Grimma. Viņas radītie tēli atrodas pastāvīgā kustībā un, kā raksta Grāvīte: “ir trauksmaini ideālisti, intelektuālisti, kurus vislabāk spēj apmierināt – perpetuum mobile” (Grāvīte 1934). Galvenais tēloto sieviešu nemiera cēlonis ir mīlestības radītas ciešanas, kas liek viņām traukties uz dažādām pasaules vietām cerībā atrast sevi un savu vietu. Būtiska stāstos ir gan sievietes ierobežotās sociālās pozīcijas atklāšana, gan viņas vēlmes un kaislības, iekšējā emocionālā dzīve, kuras attēlojuma dēļ sava laika kritikā Grimmas stāsti anonīmi vērtēti kā “neapvaldītā valodā uzrakstīti”, “piebārstīti ar dažādiem spriedelējumiem, mīlestības skumju pārdzīvojumiem, erotiskiem sapņojumiem” (V.: 1934). Ja Prāgas un citu Centrāleiropas vietu karti Grimmas prozā, atsaucoties uz Velsu, iespējams interpretēt kā ģeogrāfiski kodētu valodu, kas iezīmē konkrētas ģeogrāfiskās vietas tieši sieviešu darbības izpaušmei (Wells 2005), literārās ģeogrāfijas un stāstu melodramatisko sižetu mijiedarbi iespējams interpretēt kā “slepenu valodu”, kas palīdz autorei izteikt arī autobiogrāfiskas jūtas un pieredzi.

Stāstā “Čehu meitenes stāsts” skartās tēmas saistītas ar sievietes pozīciju sabiedrībā pēc laulībām, kas nozīmē sievietes ierobežojumu privātajā telpā, grūtniecības pieredzi, kā arī sievietes izvēli starp savām vēlmēm un upurēšanos vīrieša labā. Tas ir sakāpinātu kaislību pilns melodramatisks stāsts par jaunu meiteni, kura labi apzinās savu ierobežoto stāvokli. Uztverot laulību kā “cilvēcības zaudēšanu”, viņa atklāj vidusšķiras un augstākās sabiedrības tikumus, kur ārēji valda galantas manieres, bet patiesībā – liekulība, netikumība un brutālitate. Stāstā smalki kartēta Prāga un tās apkārtnē – gan Kārļa universitāte un Žižkova (rajons, kur Grimma mitinājās studiju laikā), gan Vāclava laukums un Vltavas upe, gan apkārtnes parki un pilis. Čehu meitene apceļo arī Austriju (Vīni) un Ungāriju (Budapeštu). Tekstā ietverts arī ceļojuma apraksts, kurā atzīmētas aplūkotas vietas – katedrāles, pilis un mākslas muzeji –, uzsverot, ka vēstītājai, kura to redz un fiksē, viss ir vienaldzīgs lielo pārdzīvojumu dēļ, kas saistīti ar nākamajām laulībām, kuras viņa redz kā upurēšanos līgavaiņa labā. Stāsta melodramatiskais sižets ir sakāpinātu emociju piesātināts – skumjas, dusmas, izmisums,

prieks, laime, grūtsirdība, ķermeniskas gaidību moka u. tml. rāda galvenā tēla iekšējo emocionālo dzīvi, nopietni uztverot jūtas, tai skaitā arī neapmierinātību un trauksmi. Šāds tēlojums sasaucas ar feministisko kritiķu Sandras Gilbertas un Sūzenas Gubāras rakstīto par romānu “Džeina Eira”, kuras varoni autore apveltījusi “ar nepārvarami spēcīgām kaislībām, kas ne vienmēr ir racionālas un kuras bieži vien nav iespējams formulēt valodā” (Gilbert, Gubar 1977: 357). Feministiskās kritiķes uzsvērušas melodrāmas potenciālu uzdot jautājumus par sievietes pozīciju patriarhālajā kārtībā un patriarhālās kārtības atmakošanā, kas ir arī būtiska Grimmas stāsta iezīme.

Savukārt stāstā “Noslēpums” parādās jūtas mātei, kura aizbraukusi uz ārzemēm studēt, mājās atstājot mazu puisīti. Intensīva nodošanās darbam palīdz aizmirst ilgas pēc bērna: “Sākumā ārzemēs jutos kā pazudusi. Milzīgā, drūmā Ziemeļu Roma apraka mani kā kapā. Noīrēju mazu mēbelētu istabiņu Masarika kundzes namā. Dzīve šeit bija neparasti dārga, bet stipendija – maza. Tomēr ikdienas rūpes bija nieks, salīdzinot ar psihiskajām ciešanām. Tādēļ jo aizrautīgāk nodevos darbam, praktizējot Prāgas slimnīcās un klausoties Kārļa Universitātē medicīnas profesorus.” (Grimma 1934: 21) Arī šajā stāstā parādās konkrētas Prāgas vietas: Prāgas Kārļa universitāte, Stromovka – liels parks Prāgas Bubenečas rajonā, Vltavas upes palienē: “[v]akar, svētdien, iedomāju Tevi, manu mazo zīlīt, kad redzēju šeit Prāgā mazos bērniņus pa Stromovku skaidrām” (Grimma 1934: 22); baroka bibliotēka Klementīnums – “[v]elti viņš man seko visās auditorijās, laboratorijās, bibliotēkās, koridoros, akadēmiskā namā Strachovā, bibliotēkā Klementīnumā” (Grimma 1934: 23) u. c. Stāsta “Kad Donava šalc” vēstītāja, nelaimīgas mīlestības vajāta, dodas uz Prāgu, kur fanātiski nododas medicīnas studijām, cerot, ka darba radītais nogurums ļaus aizmirst ilgas pēc mīļotā. Ieskanas arī vēstītājas vēlme ar iegūtajām zināšanām būt sabiedrībā noderīgai: “Tagad augu dienu pavadu plašās slimnīcās un tvanīgās laboratorijās. Papildinos anatomikumā. Specializējos sieviešu ginekoloģijā. Varbūt kādreiz spēšu palīdzēt nabaga sievietēm – savām māsām. Darbs ir nogurdinošs līdz dullumam, bet sirds ilgas viņš nespēj apslāpēt. Naktīs studēju zinātniskās grāmatas, jo bieži neguļu.” (Grimma 1934: 58) Arī šajā tekstā kā būtiskas minētas konkrētas vietas Prāgā, piemēram, dzelzceļa stacija, atkārtoti parādās Kārļa universitāte, Rīģera dārzs. Grimma caur biogrāfisku un sievietes skatījumu kartē Prāgu kā sievietes apdzīvotu vietu, kurā sievietes brīvi pārvietojas un kas piesātināta ar emocionālu un ķermenisku pieredzi.

Centrāleiropas iespaidi Grimmas daiļradē ir ceļojošā tēma, tā ienāk arī viņas vienīgajā romānā “Sirds saulgrieži” (1938), atklājot virkni autobiogrāfisku iezīmju, kas nozīmē, ka vēl pēc desmit gadiem autore vēlas savus iespaidus transformēt fikcijā, ierakstīt sevi un sev pietuvinātos cilvēkus literāros tēlos, savā ziņā padarīt par leģendu laiku, kas pavadīts Čehijā un citās Centrāleiropas valstīs (skat. 6. att.). Vairāki teikumi un detaļas romānā ir precīzi pārņemti no rakstnieces autobiogrāfijas. Romāns publicēts žurnālā “Sievietes Pasaule”, tā arī nekad nav publicēts grāmatā, nav izpelnījies nevienu atsauksmi vai recenziju. Strukturēts 29 nodaļās, tas izvada cauri galveno varoņu dzīvei apmēram 15 gadu garumā. Divi galvenie sieviešu tēli ir Linda Selga un Silvija Akmeņkalne, katrā autore iemiesojusi vairākas autobiogrāfiskās līnijas. Caur divdesmit gadus vecās jaunietes Lindas tēlu autore izstāsta autobiogrāfiski blīvu vēstījumu par jaunu censoni Lindu Selgu, kurai studijas ir dzīves papildījums un nākotnes apsoļījums un kura no trūcīgas ģimenes provincē nāk ar saukli: “Uz Rīgu! Uz augstskolu! Ar pāris



6. attēls. Martas Grimmas prozas korpuse
karte: Centrāleiropa.

latiem kabatā un mazu sainīti rokā, bet toties ar spēcīgu gribu un sajūsmā kvēlošu sirdi — iekārot dzīvi!” Pat studiju virziens ir tieši tāds pats kā pašai Grimmai, proti, ģermāņu un romāņu filoloģija, ar laiku pievēršoties baltu filoloģijai, arī pieredzētais naudas trūkums un izsalkums, pārslodze, savienojot fizisku darbu ar intelektuālu piepūli, sakrīt ar Grimmas biogrāfiskām skicēm. Savukārt 28 gadus vecajā Viestura teātra primadonnā Silvijā Akmeņkalnē autore izvēlas ievīt savu dzīves posmu, kas saistās ar Kultūras fonda stipendiju un došanos izglītoties uz ārzemēm, romānā – vokālajām un aktiermākslas studijām Prāgā. Stipendija ir ļoti gaidīta, jo ļauj izrauties no personīgās dzīves sarežģījumiem, ierastās sabiedrības, sākt jaunu dzīves posmu. Pirms meistarklasēm viņa izmanto laiku, lai apceļotu Vakareiropu: “Silvijā traucās prom ar visu sirdi un dvēseli. Prom, prom kalnos, svešzemju dabā, kur neviens to nepazīst un neatgādina tai viņas sāpes. Kas par sāpēm?! Viņa smējās. Tomēr viņu pašlaik nevilināja cilvēki. Labāk prom mežonīgos kalnos, kur daba neviltota, kur cilvēku naida viņu nesniedz. Izbraukusi cauri Varšavai un mākslīgi apguņotai Prāgai, kur atkal bija dzirdams lielpilsētas troksnis, viņa devās, pa ceļam uz Vīni un Slovākiju, uz Bratislavu un Čehoslovākijas–Ungārijas robežjoslu, uz Augstajiem Tātriem, pa reizei izbraukdama uz Bratislavu, Vīni vai Budapeštu, kur noskatījās vasaras sezonas izrādes. Viņas rīcībā atradās otrās klases ērtās

ceļojumu kartes, visos virzienos jeb kurā laikā, kā Čehoslovākijas, tā Ungārijas robežās.” (Grimma 1938) Šis maršruts gandrīz precīzi atbilst pašas Grimmas ceļojumam. Katras valsts vai pilsētas pieminējums romānā nav izvērsts, izņemot Tatrus. Romāns tādējādi papildina un padziļina Tatra iespaidu uz Grimmeru, kas aprakstīts gan īsprozā, gan dzejoļos ar Tatra tematiku. Arī romānā ienāk ķermeniskais gan caur sieviešu tēlu vizuālo raksturojumu sasaistē ar sajūtām un notikumiem, gan salīdzinājumos, piemēram, starp Silvijas tipisko ziemeļnieces uzbūvi un dienvidniecēm čehietēm: “Gan trenējusies ritmikā, plastikā, zviedru un skolu sistēmā un dažādos dejas kursos, Silvija ar savu prāvo latvietes augumu tikko patapa līdz Slovēnijas nīprajām studentēm, kas kā kalnu kazas akrobātiski lēkāja no klints uz klinti. [...] Šis kāpiens nepieradušajai Silvijai prasīja pārcilvēcīgus spēkus, lai gan viņa bija sportiste. Seja un krūtis viņai dega dzīvās ugunīs, bet tūristu apģērbs viņas slaido ziemeļnieces augumu tik labi izcēla. Viņai bija Milosas Venēras stāvs un slaidās bezkraiņās kājas ar smalkajiem ceļgaliem nu bija redzamas visā pilnībā.” (Grimma 1938) Tatra klimats, tā neparedzamība, kam Silvija nav sagatavojusies un ikvienā epizodē salst “savos plānajos tērpos”, spilgti notēlots caur subtropisko lietu, kas “lija gāznēm, piecas dienas un nakts no vietas. Kalnu ceļi pārplūda kā upes. Laukā nevarēja spert ne kāju.”

Eiropas kultūras metropoles Prāga, Vīne, Bratislava ienāk caur šo pilsētu teātriem, kur Silvija gatavojas uzstāties vai debitēt. Iekļaušanās teātros prasa ne tikai smagu darbu ar savu vokālu un aktiermeistarību, bet arī valodas zināšanas, piemēram, Bratislavas teātrī “slovaku akcents tai nebija pārvarams, viņa savā lomā uzstājās franciski” vai, gatavojoties “izrādēm Prāgā, centīgi studējama franču, čehu un slovaku valodas; arī vācu, jo gribēja uzstāties Vīnē. Tikai ungāru valoda viņai nepadevās, tā bija par daudz grūta.” Ja valodiskā ziņā ir jāpieslīpējas, tad gluži pretēji ir ar temperamentu. Grimma romānā ieraksta savu redzējumu, ka dienvīdos nevienam uz skatuves neimponēja straujums, “te uzvarēja ar intuitīvo dvēselīgo tēlojumu. Visus saistīja Silvijas neparastais ziemeļnieciskais jaunavīgums un saldā daiļskanīgā balss”, savukārt vizuālajā tēlā “vienkāršība uzvarēja visstingrāko kritiku. Viņas frizūra bija gluda kā lakota ar celiņu vidū un zelta sprogām pār pakausi.” (Grimma 1938) Jāpiebilst, ka teātris allaž bijusi Grimmas aizraušanās, pati sapņojusi kļūt par aktrisi un piedalījusies dramatiskajos kursos, arī ceļojuma aprakstā tam veltītas vairākas nodaļas, jo viņai bijusi brīvbiļete, ar kuru var apmeklēt visas teātra izrādes. Romāns blīvi daiļdarba formātā atklāj Grimmas autobiogrāfiskās detaļas jau ar distancētāku skatījumu laikā un telpā, savā veidā traktējot kā leģendāru šo mobilitāti, intensīvo ģeogrāfisko vietu un sabiedrības tvērumu un sievišķo pieredzi, kas saskan jau ar iepriekš šajā rakstā nostiprināto *geo-parler femme* teorētisko un tekstuālo pieeju.

Noslēguma piezīmes

Rakstnieču biogrāfiju un literāro darbu ģeogrāfijas izpēte ir daudz plašāks un sarežģītāks uzdevums, nekā to iespējams izvērst vienā rakstā, pirmkārt, jau tādēļ, ka pats ģeogrāfijas jēdziens ir piesātināts ar daudzām niansēm, sākot no kartogrāfijas, telpiskās pārvietošanās, trimdas un ekspatriācijas līdz topogrāfiskām identitātes metaforām, kartēšanai un ķermenim. Šajā rakstā

iezīmēti vien daži no aspektiem, parādot, ka pārvietošanās, sieviešu mobilitāte ietekmē vietas, kas ienāk arī viņu literārajos darbos – aplūkotojos piemēros latviešu rakstnieču prozā biogrāfiskās un literārās kartes pārklājas.

Izmantojot digitālās kartēšanas rīkus un metodes, iespējams pārdomāt ne tikai telpiskās kop-sakarības, bet arī spriedzi, kas saista literāros tekstus ar rakstniecēm un materiālo pasauli, un stāstījuma stratēģijas, kurās rakstnieces izmanto ģeogrāfiju kā slepeni vai kodētu valodu, lai pārraidītu emocionālu vai simbolisku nozīmi, atklājot, ka prozas ģeogrāfija nozīmē vairāk nekā tikai garuma un platuma grādus. Balstoties ģeokritiskajā pieejā, konkrēti *geo-parler femme* skatījumā, pievērsta uzmanība autorei un viņas radīto sieviešu tēlu ģeogrāfiskajai pieredzei, kas atklāj fiziskās ģeogrāfijas un literārās ģeogrāfijas (Martas Grimmas gadījumā Centrāleiropas, īpaši Prāgas) pārklāšanos.

Lai arī salīdzinājumā ar īsprozu romāni ir šķietami pateicīgāki, lai ievītu plašāku ģeogrāfiju sava apjoma un dažādo iespējamo sižeta līniju dēļ, tomēr autorei sievietes ne tik bieži to izmantojušas. Sieviešu romāniem nav arī tipiski, ka autore izvēlētos tēlot vietas, kur pašas nav bijušas. Izņēmumi, protams, ir, bet tad tās mēdz būt vai nu vietas, kas tiek pieminētas, bet nav romāna darbības vietas (Aīdas Niedras romāni), vai arī romāns rekonstruē senvēsturi, no-tiek laikā un vietās, kas ir vairākus gadsimtus pirms autore dzīves laika (Kaija, Dzintarzeme). Salīdzinājums ir viena no galvenajām priekšrocībām, ko piedāvā teksta kartēšana, veidojot konkrētas ģeotelpas izvilckumu dažādos viena vai vairāku autoru daiļliteratūras darbos.

- Brousseau, Marc (2017). In, of, out, with, and through. New perspectives in literary geography. *The Routledge Handbook of Literature and Space*. London and New York: Routledge, pp. 9–28.
- Daija, Pauls; Kalnačs, Benedikts (2019). The Geographical Imagination in Early Twentieth-Century Latvian Novels. *Primerjalna Književnost*, No. 42(2), pp. 153–171.
- Daija, Pauls; Kalnačs, Benedikts (2019). Exploring Space and Place in Literature: Towards Digital Analysis and Literary Mapping in the Context of Latvian Cultural History. *Colloquia*, No. 42, pp. 67–83.
- Daukste-Silasproģe, Inguna (2022). Jelgavas meitene pasaulē: ieskats rakstnieces Indras Gubiņas dzīves ģeogrāfijā. *Aktuālas problēmas literatūras un kultūras pētniecībā*, 27. Liepāja: LiePA, 253.–268. lpp.
- Daukste-Silasproģe, Inguna (2020). Rīga Andreja Johansona dzīves un daiļrades kartē. *Letonica*, Nr. 42, 50.–65. lpp.
- Enloe, Cynthia (2000). *Bananas, Beaches and Bases: Making Feminist Sense of International Politics*. 1989. Berkeley: University of California Press.
- Hones, Sheila (2010). Literary geography and the short story: setting and narrative style. *Cultural Geographies*, No. 17(4), pp. 473–485.
- Klīdzējs, Jānis (1972). Rakstniece ar iekšēju redzīgumu. *Latvija*, 11.11., 3. lpp.
- Gailīte, Angelika (1927). *Grēcinieces*. Rīga: Jāņa Rozes apgāds.
- Gailīte, Angelika (1936). Vienpusīgs vēstījums. Ķelpe, Jānis (sast.).
- Sieviete latvju rakstniecībā*. Jelgava, 80.–89. lpp.
- Gilbert, Sandra, Gubar, Susan (1979). *The Madwoman in the Attic: The Woman Writer and the Nineteenth-Century Literary Imagination*. New Haven, CT: Yale University Press.
- Grāvēite, Elza (1934). Kritika. *Izglītības Ministrijas Mēnešraksts*, Nr. 3.
- Grimma, Marta (1933). *Čehoslovākijā*. Autore izdevums.
- Grimma, Marta (1934). *Kad Donava šalc*. Autore izdevums.
- Grimma, Marta (1935). Manas studijas ārzemēs. *Latviete*, Nr. 5, 87., 88. lpp.
- Grimma, Marta (1981). Mans dzīves stāsts. *Karogs*, Nr. 11, 148.–152. lpp.
- Grimma, Marta (1938–1939). Sirds Saulgrieži. *Sievietes Pasaule*, Nr. 1–12, 1938, Nr. 1–2, 1939.
- Larsson, Lisbeth (2017). *Walking Virginia Woolf's London: An investigation in literary geography*. London: Palgrave Macmillan.
- Mazón, Patricia M (2000). “Fräulein Doktor”: Literary Images of the First Female University Students in Fin-de-Siècle Germany. *Women in German Yearbook: Feminist Studies in German Literature & Culture*, 16.1, pp. 129–150.
- Moretti, Franco (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. London & New York: Verso.
- Pērļupe, Lidijs (1942). Hildas Vikas romāns “Laimīgā dziesma”.
- Daugavas Vēstnesis Literatūrai un Mākslai*, 09.08., 1. lpp.
- Piatti, Barbara, Bär Hans Rudolf, Reuschel, Anne-Kathrin, Hurni, Lorenz, Cartwright, William (2009). Mapping Literature: Towards a Geography of Fiction. William Cartwright, Georg Gartner, Antje Lehn (eds.). *Cartography and Art*. Springer-Verlag Berlin Heidelberg, pp. 177–193.
- Rūmane, Anna (1902). Iz Dienvidiem. Jūras spēkā. *Pēterburgas Avižu Literāriskais pielikums*, 01.05., 275., 276. lpp.
- Scriber, Mary Suzanne (1997). *Writing Home: American Women Abroad, 1830–1920*. Charlottesville and London: University Press of Virginia.
- Štolls, Pāvels (2015). Marta Grimma un čehu rakstnieču daiļrade. *Dzimums, literārā konvencija un jaunrade. No Baroka līdz postmodernismam*. LU Akadēmiskais apgāds, 119.–126. lpp.
- Tally Jr, Robert T (2017). Introduction. The reassertion of space in literary studies. *The Routledge Handbook of Literature and Space*. London and New York: Routledge, pp. 1–7.
- Tally Jr, Robert T (2013). *Spatiality*. London: Routledge.
- Tally Jr, Robert T (2011). On literary cartography: narrative as a spatially symbolic act. *NANO: New American notes online 1*.
- Tally Jr, Robert T. (ed.) (2011). *Geocritical Explorations: Space, Place, and Mapping in Literary and Cultural Studies*. New York: Palgrave MacMillan.
- Upīts, Andrejs (1927). Latviešu literatūra. *Domas*, Nr. 6.

V. (1934). Marta Grimma. Kad Donava šalc. *Mūsu Nākotne*, Nr. 2, 148. lpp.

Vērdušs, Kārlis (2005). Dzejnieces dzimšana. Zemgaliešu Biruta. *Sārtais skūpstis*. Rīga: Pils, 83.–91. lpp.

Zeile, Pēteris (1993). Latvijas Kultūras fonds (1920–1940). *LZA Vēstis, A daļa*, Nr. 1, 17. –23. lpp.

Wells–Lynn, Amy (2005). The Intertextual, Sexually-Coded Rue

Jacob: A Geocritical Approach to Djuna Barnes, Natalie Barney, and Radclyffe Hall. *South Central Review*, No. 22(3), pp. 78–112.

Wells, Amy (2014). Geocriticism, Gender, and Genre: Literary Geographies and Female Narrative Strategies. *Géocritique: État des lieux/Geocriticism: A Survey, Limoges, Pulim*, pp. 146–154.

Wells, Amy (2017). Feminist Geocritical Activism: Natalie

Barney's writing of women's spaces into women's places. Tally, Robert T. (ed.). *The Routledge Handbook of Literature and Space*. London and New York: Routledge, pp. 349–361.

Lyrical Geographies: Mapping Literary Space in Novels and Short Stories by Latvian Women Writers

Zita Kārkla, Eva Eglāja-Kristsons

Keywords: digital mapping, literary geography, geocriticism, *geo-parler femme*, history of women's writing

The aim of the article is to explore the use of digital mapping tools in the study of history of women's writing. Considering that space is never neutral and women are positioned in the world differently than men, women's relationship to space also differs. Applying digital mapping tools (*Google Maps*, GIS) to the corpus of 110 prose texts (sketches, short stories, novels) by Latvian women writers, published between 1872 and 1940, we are particularly interested in the geography of women's literary space and the relationship between women's writing and social history, which are closely related. In the first part of the article, we look at the application of data visualization and digital mapping methods and tools in literary research, as well as the concepts of geocriticism and *geo-parler femme* for the interpretation of literary texts. Furthermore, using the results of literary digital mapping and combining distant and close reading approaches, we question the territory of women's literary space and the representations of local and global geography in women's writing, arriving at a conclusion that women writers' biographical and literary maps often overlap. In the final part of the article we analyse Marta Grimma's autobiographical and literary works as a case study of the practice of integrating biographical cartography into fiction, proposing that Grimma uses the *geo-parler femme* narrative strategy to inscribe bodily and emotional experiences in her literary works.

Ceļojumu kartēšana un sociālpolitiskās tendences: Latvijas Padomju rakstnieku savienības biedru braucieni 20. gadsimta 70.–80. gados

Madara Eversone, Jānis Oga

Publikācija tapusi Valsts pētījuma programmas “Humanitāro zinātņu digitālie resursi” projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001) un Jāņa Ogas pēcdoktorantūras pētījumā “Latviešu prozisti padomju okupācijā: sadarbība un nevardarbīgā pretošanās (1968–1991)” (pētniecības pieteikums Nr.1.1.1.2/VIAA/3/19/482).

Padomju otrreizējās okupācijas laikā (1944–1990) Latvijas iedzīvotāju ceļošanas iespējas ārpus PSRS bija ierobežotas, reglamentēta bija lielvalsts iedzīvotāju profesionālā dzīve, brīvā laika pavadīšanas iespējas un ceļojumu galamērķi. Tomēr noteikta sabiedrības daļa varēja baudīt daudz plašākas sadzīves un materiālās privilēģijas, tostarp ceļošanu, regulārus komandējumus, privātus un ārstniecības braucienus. Viena no šādām grupām bija radošā inteliģence, tostarp rakstnieki. Latvijas Padomju rakstnieku savienības (turpmāk – RS) biedra statuss deva iespējas doties radošos komandējumos un atpūtas braucienos gan pa Latviju un citām Padomju Savienības republikām, gan braucienos ārpus PSRS – uz sociālistiskā bloka, eksotiskajām un Rietumu valstīm.

Tūrisms Aukstā kara laikā bija kļuvis par sašķeltās pasaules *kaujas lauku* – politiskajā, ekonomiskajā, kultūras un drošības jautājumā. Abu antagonistisko lielvaru – PSRS un Amerikas Savienoto Valstu – sacensība, kas ietvēra abu konkurējošo kultūru savstarpēju izpēti, veicināja arī kultūras sakaru dibināšanu un apmaiņu starp Rietumiem un sociālistiskajām valstīm (Bechmann Pedersen, Noack 2019: 3). PSRS radošā inteliģence bija viens no propagandas līdzekļiem kā iekšzemē, tā latviešu sabiedrībā ārzemēs. Valstij bija nepieciešama tāda inteliģence, kurai uzticējās sabiedrība un kura bija lojāla padomju iekārtai (Zālīte 1998). Pētot radošās inteliģences privilēģijas, tostarp brīvākas ceļošanas iespējas, iespējams atklāt Komunistiskās partijas (turpmāk – KP) politiku attiecībā uz radošo inteliģenci – braucienu sociālos un politiskos aspektus.

Raksta mērķis ir, izmantojot digitālās kartēšanas iespējas Latvijas Universitātes Literatūras, folkloras un mākslas institūta digitālajā resursā vortālā *literatura.lv* (sk. Eglāja-Kristsone 2021) un *Google Maps*, apkopot un analizēt RS biedru ārpus Padomju Savienības veikto braucienu sociālpolitiskos aspektus un tendences 20. gadsimta 70. un 80. gados, kā arī iezīmēt šo braucienu atspoguļojumu minētā laika posma publicistikā.

Raksta pirmajā daļā analizēts RS biedru braucienu (1970–1989) datu korpuss un tā vizualizācija kartēs. Raksta otrajā daļā izvērstāk analizēti RS biedres rakstnieces Zentas Ērgles (1920–1998) braucieni uz Rietumu kapitālistiskajām un eksotiskajām valstīm un to atspoguļojums viņas publicistikā¹.

Pētījumā izmantota digitālo humanitāro zinātņu pētniecības metode – digitālā kartēšana (*digital mapping*), atklājot, kā rakstnieku braucienu korpusa izveide, iegūto datu

1 Daija un Kalnačs pētījumā par 20. gadsimta sākuma romānos minētajām vietām atzina, ka “papildus visam korpusam pielietotajai kvantitatīvajai analīzei, kas ļauj spekulēt par konkrētāku dinamiku atsevišķu ģeogrāfisko apgabalu popularitātes kāpumā un kritumā, iegūtie dati ļauj detalizētāk raksturot konkrētos romānos aplūkojamo ģeogrāfisko telpu. To var panākt, vai nu izveidojot attiecīgo tekstu digitālās kartes, vai rūpīgi pārbaudot konkrētas vietas dažādos romānos.” (Daija, Kalnačs, 2019a: 71)

ģeomarkēšana vortālā *literatura.lv* datubāzē un datu digitālā vizualizācija (karšu izveide) palīdz rakstnieku braucienu tendenču atklāsmē un analizē, atspoguļojot biežāk ceļojošās personas, populārākos braucienu galamērķus (sistemizējot valstis pēc to ideoloģiski politiskā stāvokļa) un braucienu nolūkus. RS biedru braucienu gadījuma analīzei izvēlētas Zentas Ērgles ceļojuma piezīmes “No Plutona līdz Neptūnam” (1975) par tūrisma braucienu uz Āfrikas valstīm un “Gads ar diviem pavasariem” (1979) par braucienu uz Jaungvineju, Austrāliju un Jaunzēlandi. Detalizētai analīzei veikta šo darbu literārā kartēšana (*literary mapping*), attālais lasījums (*distant reading*) un tuvlasījums (*close reading*), atklājot to literāro ģeogrāfiju, sociopolitiskās tendences un īpatnības. Ar digitālo humanitāro zinātņu metodēm izveidotas darbos minēto vietu kartes.

Aplūkojot ārzemju braucienu izpausmes stagnācijas perioda literatūrā, lietderīgi izmantot Alekseja Jurčaka (*Alexei Yurchak*) aprakstīto unikālo izpratni par jēdzienu “ārzemes” (*za-granica*) Padomju Savienībā. Proti, *zagranica* nav tiešs sinonīms izpratnei par jēdzienu “aiz robežām” (*za granicej*) ar akcentu uz konkrētu valsts robežu un teritoriju, kas atrodas aiz tās, vai jēdzienu “ārvalstis” ar akcentu uz citām, reāli pastāvošām valstīm. Padomju telpā “ārzemes” apzīmē nevis robežu un nevis reālo teritoriju, bet gan iedomātu telpu – vienlaikus reālu un abstraktu, pazīstamu un nerasniedzamu, ikdienišķu un eksotisku, tādu, kas atrodas gan šeit, gan tur” (Yurchak 2014: 312). Būtiska nianse, kas iezīmē Latvijas un pārējo Baltijas valstu atšķirības no pārējām Padomju Savienības republikām, saistīta ar Latvijas ģeogrāfisko novietojumu. Igaņu zinātniece Epa Annusa (*Epp Annus*) formulē Baltijas valstis kā Rietumu pierobežu un Baltijas valstu ciešo saistību ar kaimiņvalstīm: “Šeit Rietumu pierobežā tika baudīta joprojām aktīvās pirmspadomju kultūras atmiņas priekšrocība, salīdzinot ar tām zemēm, kurās sovietizācijas process sākās tūlīt pēc revolūcijas gados: Rietumu pierobežās dzīva atmiņa uzturēja nepārtrauktību ar pirmspadomju laikiem.” (Annus 2018: 240) Tuvāku un tālāku Rietumvalstu apmeklējums savā ziņā palīdzēja šai nepārtrauktībai piešķirt nianšes un ar 70. un 80. gados publicēto darbu starpniecību veicināja nevardarbīgo pretošanos un atgriešanos pie pirmskara Latvijas vērtību sistēmas.

Raksta autori ar pietuvinājumu konkrētam periodam, autoriem un darbiem pretendē uz turpinājumu pētījumiem gan latviešu literatūras kartēšanas, gan ceļojumu piezīmju izpētes jomā. Līdzšinējo nedaudzo kartēšanas pētījumu vidū jāizceļ Paula Daijas un Benedikta Kalnača atziņas (Daija un Kalnačs 2019a; 2019b), kas gūtas, pētot un kartējot 20. gadsimta sākumā izdotos latviešu romānus, Zitas Kārklas komentāri Minnas Freimanis ceļojumu aprakstiem (Kārkla 2021) un Ingūnas Daukstes-Silasproģes raksts “Rīga Andreja Johansona dzīves un daiļrades kartē” (Daukste-Silasproģe 2020). Latviešu rakstnieku ceļojumu piezīmes un tajās atklāto robežas simbolizāciju pētījusi Eva Eglāja-Kristsone (Eglāja-Kristsone 2012), savukārt orientālisma izpausmes un citādības aspektus aplūkojusi Maija Burima, kura uzsver ceļojuma piezīmju ciešo saistību ar koloniālisma paradigmu, norādot, ka ceļojumu piezīmju rakstīšana bijusi nozīmīga identitātes naratīva veidošanas sastāvdaļa koloniālajās impērijās (Burima 2018: 65). Trūkst plašāku pētījumu par latviešu rakstnieku ceļojumiem padomju periodā².

2 Par ārzemju ceļojumu iespaidiem latviešu rakstnieku darbos stagnācijas laikā un tajos rodamajām liecībām par varas atļauto un liegto skat. Oga 2021.

Raksta autori seko Daijas un Kalnača izteiktajai atziņai, ka “šāda tipa pētījumos iegūtās zināšanas var tikt izmantotas gan analītiskā veidā, lai definētu galvenos vietu atlasē principus, gan nodrošināt teorētisko datu analīzi no dažādiem aspektiem, kā arī ilustratīvā nolūkā, piemēram, apkopojot kvantitatīvu informāciju par vietvārdiem vai pielāgot pētījumus tūristu ceļvežiem” (Daija, Kalnačs 2019b: 156).

Ceļošana Padomju Savienībā un radošās inteliģences privilēģijas

Braucieni uz ārzemēm atradās pilnīgā PSRS KP kontrolē. Gan izejošā, gan ienākošā tūrisma organizācija notika centralizēti, pamatojoties uz KP lēmumiem un direktīvām. Par tūrisma organizēšanu ārpus PSRS atbildēja Vissavienības akcionāru sabiedrības “Intūrists” un Vissavienības arodbiedrību centrālās padomes Tūrisma un ekskursiju pārvaldes reģionālās nodaļas³.

Kanādiešu vēsturniece Anne Gorsača (*Anne E. Gorsuch*) izpētījusi, ka daudzi no tiem, kuri ceļoja 50. un 60. gados, izmantojot Vissavienības akcionāru sabiedrības “Intūrists” pakalpojumus, bija privilēģētas grupas, kurās noteicošā, lai gan ar izņēmumiem, bija partijas piederība: akadēmiķi, kultūras darbinieki, rūpnīcu vadītāji, partijas funkcionāri (Gorsuch 2011: 82). Tomēr arī piederība privilēģētai grupai nenodrošināja ceļošanai ārpus PSRS būtiskos finanšu līdzekļus. Vispirmām kārtām, lai ceļotu (īpaši dārgi bijuši eksotiskie ceļojumi), rakstniekam bija jābūt turīgam, taču, nonākot galamērķī un apjaušot reālās iespējas, jāpiedzīvo pazemojums un atklāsme – rakstnieks, lai arī ir viens no visbrīvākajiem cilvēkiem no visbrīvākās valsts pasaulē, nonācis tiešām brīvā pasaulē, joprojām paliek ieslodzījumā (Oga 2021: 133–134).

Lai arī šajā laikā KP politika attiecībā uz ceļošanu ārpus Padomju Savienības kļuva liberālāka, pastāvēja rīkojumi un instrukcijas, nosakot kārtību, kā jākomplektē tūristu grupas braucieniem uz ārzemēm, šo grupu vadītāju pienākumus un privilēģijas, atskaitīšanās procedūru pēc atgriešanās no tūrisma brauciena.

Ceļošanu ietekmēja Padomju Savienības starptautiskā un ārpolitiskā situācija. 20. gadsimta 70. gadi ir detantes (*Détente*) periods: aukstā kara saspīlējums starp Padomju Savienību un ASV mazinājās, pozitīvi ietekmējot tūrismu un kultūras sakaru apmaiņu (Eglāja-Kristsons 2016: 28). Zinātniski tehniskā un kultūras sadarbība starp Rietumvāciju un Padomju Savienību aizsākās 50. gadu beigās un uzplauka Brežņeva laikā, kad jau 1970. gada augustā

3 Vissavienības arodbiedrību centrālās padomes (VACP) Centrālās tūrisma un ekskursiju padomes Latvijas republikāniskā tūrisma un ekskursiju padome ar mainīgiem nosaukumiem (VACP Tūrisma un ekskursiju Latvijas PSR pārvalde, VACP tūrisma un ekskursiju Latvijas pārvalde, Latvijas tūrisma un ekskursiju pārvalde) pastāvēja no 1946. līdz 1990. gadam.

Maskavā tika parakstīts Vācijas un Padomju Savienības līgums. Tas aizsāka Vācijas politiku attiecībā uz Austrumeiropu (*Ostpolitik*), paverot ceļu diplomātisko attiecību normalizēšanai un apstiprinot miermīlīgo teritoriālo *status quo* starp Padomju Savienību un Vācijas Federatīvo Republiku (turpmāk – VFR).

Padomju Savienības tūrisma attīstību ar Rietumvalstīm veicināja arī 1975. gada 1. augustā Helsinkos parakstītais Eiropas drošības un sadarbības konferences akts. Saskaņā ar šo aktu Padomju Savienības un Rietumu valstis apņēmās veicināt brīvāku starpvalstu tirdzniecību, kopuzņēmumus, ražošanas uzņēmumus, tehnoloģiju nodošanu, transportu, tūrismu un darbaspēka migrāciju, kā arī atbalstīja cilvēku un ideju brīvu kustību un cilvēktiesību ievērošanu (Bechmann Pedersen, Noack 2019: 6). Taču Padomju Savienība nebija gatava neierobežotam, brīvi pieejamam tūrismam uz Rietumiem, jo tas varētu apdraudēt valsts totalitāro režīmu un ideoloģiju. Padomju Savienības un Rietumu attiecības atkal kļuva saspīlētas līdz ar PSRS iebrukumu Afganistānā 1979. gadā, un saspīlējums turpinājās līdz pat Mihaila Gorbačova nākšanai pie varas 1985. gadā. Krievu vēsturnieks Igors Orlovs (*Igor Orlov*) secinājis, ka Gorbačova laikā politiskās nostādnēs ceļošanas jautājumos bija liberālākas un attiecības ar kapitalistiskām valstīm šajā laikā ievērojami paplašinājās ne tikai kvantitatīvā, bet arī kvalitatīvā ziņā (Orlov 2014: 26).

Kopumā 20. gadsimta 70. un 80. gadi ir tūrisma un dažādu ārvalstu braucienu uzplaukuma laiks. Šajā laikā pieaug arī padomju pilsoņu sociālā labklājība – bija vairāk iespēju un brīvā laika ceļošanai. Anne Gorsača norāda, ka 1970. gadā vairāk nekā 1,8 miljoni padomju pilsoņu ceļoja uz ārzemēm, savukārt 1985. gadā šis skaitlis bija pieaudzis līdz pat 4,5 miljoniem (Gorsuch 2011: 5).

Tomēr līdz pat 80. gadu beigām ceļošanas jautājumā ārpus PSRS pastāvēja politiskie un ideoloģiskie ierobežojumi, kas attiecās uz jebkuru sabiedrības grupu. Līdz ar to arī piederība t. s. radošajai inteliģencei neatbrīvoja no stingrās izceļošanas kārtības un sarežģītās birokrātiskās sistēmas, ar ko nācās saskarties RS biedriem. Tai pašā laikā profesionāla rakstnieka statuss un piederība RS radīja daudz lielākas ceļošanas iespējas ārpus PSRS, ko uzskatāmi atspoguļo savāktie dati par RS biedru braucieniem 70. un 80. gados, kas turpmākajā tekstā aplūkoti un analizēti tuvāk.

Latvijas Padomju rakstnieku savienības biedru braucienu datu korpuss, digitālā kartēšana un datu vizualizācija

Viens no pētniecības avotiem digitālajās humanitārajās zinātnēs ir strukturēti un vizuāli dati (kartes, diagrammas u. tml.), kas palīdz atbildēt uz pētījumā izvirzītajiem jautājumiem, izmantojot arī starpdisciplināras pieejas un metodes (Anderson et al. 2017: 248–249), piemēram, arhīva studijas, kultūrvēsturisko metodi, tuvlasījumu un digitālo humanitāro zinātņu metodes, kā šī pētījuma gadījumā. Lai aptvertu RS biedru braucienu ģeogrāfiju un intensitāti, kā arī biedrus, kuri ceļojuši biežāk, šajā pētījumā izveidots strukturēts RS biedru braucienu (1970–1989) datu korpuss (sk. 1. attēlu). Korpussā iekļautie dati iegūti no RS sekretariāta sēžu (LNA LVA 473. fonds) un RS KP pirmorganizācijas (LNA LVA PA-7263. fonds) sēžu protokoliem

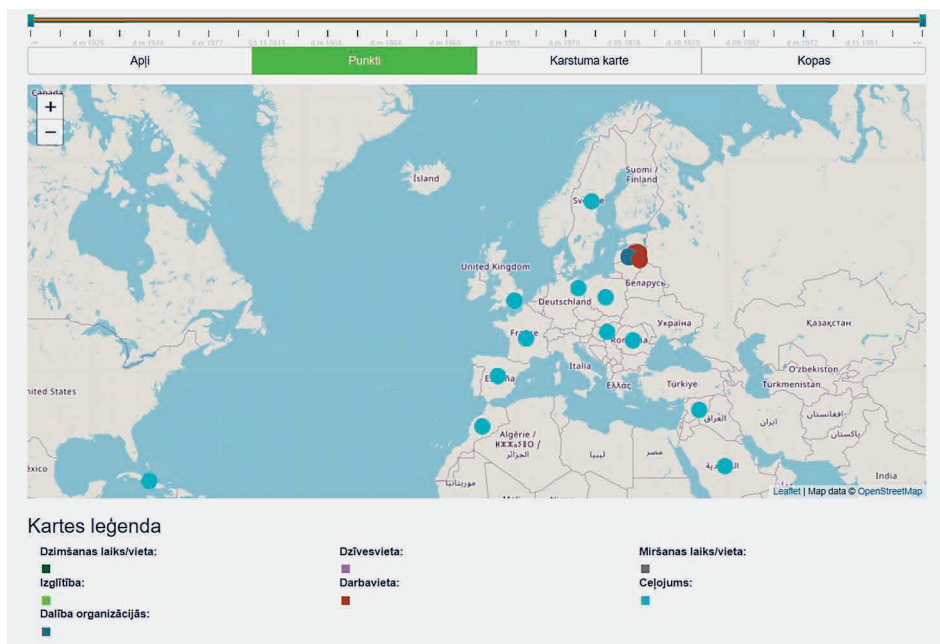
Persona	Vieta	Vietas statuss	Veids	Datums	Komentārs [teksta lauks]	Organizācijas, kas sūta
Andrejs Dripe	Tuvie Austrumi	Eksotiska valsts	komandējums	1982	Brauciens materiāla vākšanai.	
Andrejs Dripe	Ungārija	Satelītvalsts	tūrisma brauciens	11.1981.	Brauciens kopā ar sievu.	
Andris Jakubāns	Bulgārija	Satelītvalsts	komandējums	04.1982	Brauciens uz 10 dienām.	
Andris Jakubāns	Čehoslovākija	Satelītvalsts	komandējums	03.1984.	Brauciens PSRS Rakstnieku savienības delegācijas sastāvā.	PSRS Rakstnieku savienība
Andris Jakubāns	Turcija	Eksotiska valsts	tūrisma brauciens	09.1985.	Brauciens KPFSR RS tūrisma grupas sastāvā.	KPFSR Rakstnieku savienība
Andris Jakubāns	Dienvidslāvija	Sociālistiska valsts	komandējums	10.1980.	Brauciens PSRS Rakstnieku savienības delegācijas sastāvā.	PSRS Rakstnieku savienība
Andris Jakubāns	Čehoslovākija	Satelītvalsts	komandējums	03.1984.	Brauciens PSRS Rakstnieku savienības delegācijas sastāvā.	
Andris Kolbergs	VFR	Rietumi	komandējums	09.1982.	Brauciens Padomju Miera komitejas specgrupas sastāvā uz 14 dienām.	Padomju Miera komiteja
Andris Kolbergs	Ruse	Satelītvalsts	komandējums	09.1983.	Brauciens uz 14 dienām dienām Rusē.	
Andris Kolbergs	Turcija	Eksotiska valsts	tūrisma brauciens	09.1985.	Brauciens uz 14 dienām Krievijas Padomju Federatīvās Sociālistiskās Republikas Rakstnieku savienības specgrupas sastāvā.	

1. attēls. Ekrānuzņēmums no Madaras Eversones izveidotā Latvijas Padomju rakstnieku savienības biedru braucienu (1970–1989) datu korpusa *MS Excel* formātā. Korpusa fragments.

Latvijas Nacionālā arhīva Latvijas Valsts arhīvā (turpmāk – LNA LVA). Datu korpusā apkopoti tikai tie braucieni, kas kārtoti caur RS un fiksēti tās lietvedības dokumentos. Šādā veidā datu apkopojums ir ērti pārskatāms, analizējams, kā arī to iespējams importēt dažādās datubāzēs. Izveidotajā datu korpusā RS biedru braucieni sadalīti pa desmitgadēm, atsevišķi izdalot ģeogrāfiskās vietas pēc to ideoloģiski politiskā statusa: 1) citas PSRS republikas, 2) sociālistiskās valstis (piemēram, Bulgārija, Čehoslovākija, Vācijas Demokrātiskā Republika (turpmāk – VDR), Polija u. c.), 3) padomju pilsoņa izpratnē eksotiskās un jaunattīstības valstis (piemēram, Filipīnas, Turcija, Indija, Āfrikas valstis, Laosa, Kambodža u. c.) un 4) Rietumi jeb t. s. kapitālistiskās valstis (piemēram, VFR, Lielbritānija, ASV, Kanāda, Austrālija u. c.).

Lai izvērtētu braucienu nolūkus, atsevišķi izdalīts brauciena veids un organizācija, ar kuras starpniecību RS biedrs varēja izbraukt ārpus Padomju Savienības. Papildu informācijai izveidots komentāru lauks.

Pētījumā iegūtie dati par RS biedru braucieniem 20. gadsimta 70.–80. gados ir pieejami LU LFMI virtuālajā resursā – vortālā *literatura.lv*. Dati šajā resursā ievadīti manuāli, tādējādi

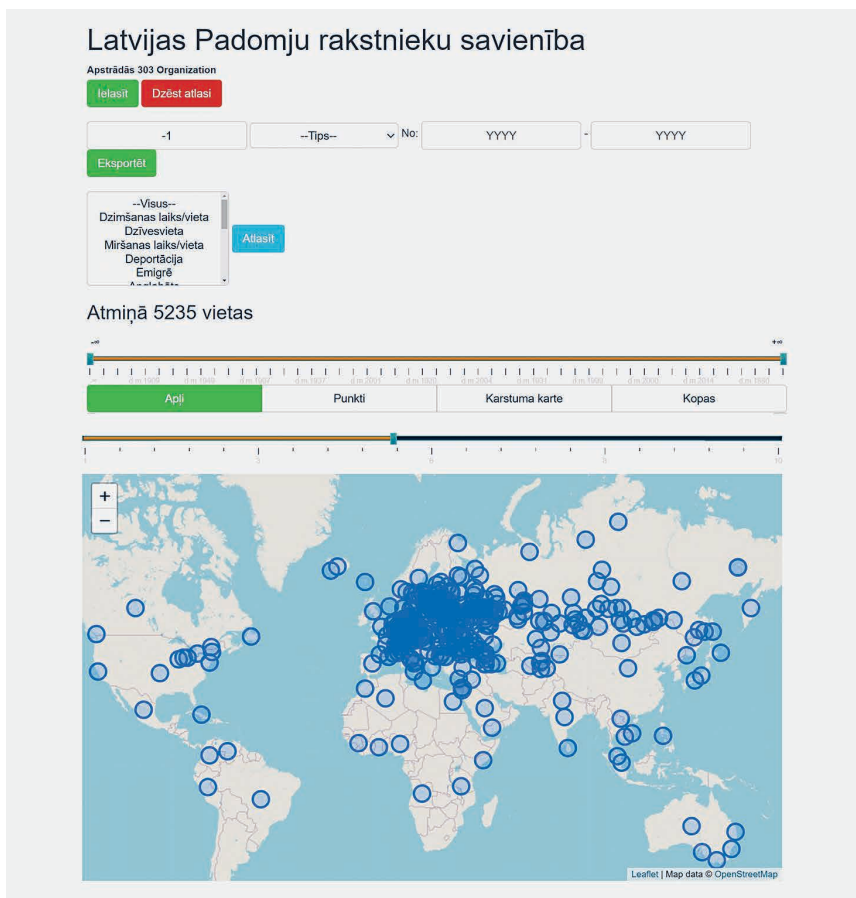


2. attēls. Ekrānuzņēmums no Andreja Dripes reģistrētajiem ceļojumiem vortālā *literatura.lv*.

izvairoties no varbūtējām datu ievades kļūdām un ļaujot katru gadījumu izvērtēt individuāli. Brīvpieejā iespējams aplūkot katra RS biedra individuālos braucienus, apmeklējot konkrētu personas profilu vortālā, piemēram, Imants Ziedonis, Žanis Grīva, Andrejs Dripe, Zenta Ērgle⁴ u. c. Dati par braucieniem pieejami un vizualizēti katras personas profilā, atlasot sadaļu “Vietas”. Šajā sadaļā pieejama karte un tās leģenda, kurā var atlasīt interesējošo laika posmu un notikuma tipu (pētījuma gadījumā “ceļojums”), iegūstot datus, kas saistīti tikai ar konkrēto notikumu. Dati digitālajā kartē vizualizēti kā apli, punkti, “karstuma karte” un kopas (sk. 2. attēlu).

Vienkopus pieejamus datus par visiem RS biedru braucieniem atlasītajā laika posmā šobrīd iespējams izgūt tikai vortālā *literatura.lv* administratīvajā vidē *admin.garamantas.lv* (sk. 3. attēlu). Piekļuves tiesības administratīvajam resursam ir LU LFMI digitālo resursu redaktoriem un pētniekiem.

4 Skatīt: Imants Ziedonis <https://www.literatura.lv/lv/person/Imants-Ziedonis/867820>, Žanis Grīva <https://www.literatura.lv/lv/person/Zanis-Griva/872941>, Andrejs Dripe <https://www.literatura.lv/lv/person/Andrejs-Dripe/871700>, Zenta Ērgle <https://www.literatura.lv/lv/person/Zenta-Ergle/871710>.



3. attēls. Ekrānuņēmums no Latvijas Padomju rakstnieku savienības biedru ceļojumu kartes LU LFMI vortāla *literatura.lv* administratīvajā vidē *admin.garamantas.lv*.

Administratīvajā vidē datus var izgūt, atlasot konkrētas organizācijas (Latvijas Padomju rakstnieku savienība) profilu, tajā ar speciāli izveidotu ģeogrāfiskās kartēšanas izpēti tiek atlasīta visa informācija par RS kā organizācijas notikumiem (darbavietas, ceļojumi, izglītības vietas u. c.), kas ievadīti datubāzē. Pēc visu notikumu atlasē jāizvēlas interesējošais notikuma tips (šajā gadījumā “Ceļojums”), lai izgūtu datus turpmākai pētniecībai un datu vizualizācijai. Atlasītos datus (ģeogrāfiskās vietas, to koordinātas, informācija par personām u. c.) administratīvajā vidē var izgūt *MS Excel* dokumenta formātā, tos iespējams importēt dažādos digitālo karšu rīkos, piemēram, *Google Maps*. Vietu koordinātas LU LFMI digitālo resursu administratīvajā vidē tiek iegūtas no *Lursoft* datubāzēm, *OpenStreetMap* un *Google Maps* digitālo



4. attēls. Ekrānuņēmums no Madaras Eversones izveidotās Latvijas Padomju rakstnieku savienības biedru ārzemju braucienu (1970–1989) digitālās kartes *Google My Maps*.

karšu koordinātām. Raksta tapšanas laikā turpinās darbs pie publiski pieejama rīka izstrādes, ko veic LU LFMI IT speciālisti.

Pētījumā iegūto datu (braucienu skaitu, maršrutu, mērķu u. tml.) izvērtēšanai un prezentācijai izveidota arī digitālā karte (sk. 4. attēlu). Tūrisma vēsturnieki jau kādu laiku izmanto šo metodi, lai spētu precīzāk saskatīt pagātnes tūrisma maršrutu relācijas laikā un telpā (Anderson et al. 2017: 247). Kartēšanas praksi, sevišķi digitālo kartēšanu, plaši izmanto arī citu zinātņu pētnieki. Tas ir ērts rīks, jo digitālajās ģeogrāfiskajās kartēs ļauj vienkopus savie- tot vairākus datu slāņus, tādējādi izveidojot t. s. dziļās kartes (*deep mapping*) – tās vienlaicīgi funkcionē gan kā datubāzes un krātuves, gan datu vizualizācijas un darbojas kā platformas, būdamas gan datu apstrādes process, gan produkts (Foka et al. 2020: 203). Digitālās kartes nodrošina iespēju viegli atjaunināt datus un pilnveidot interaktivitāti. Tomēr tas nenotiek automātiski, un digitālās kartes bieži tiek veidotas, lai attēlotu konkrētā objekta telpiskā rakstura statisku vizualizāciju (turpat: 205).

RS biedru braucienu karte (1970–1989) izveidota, gan izmantojot ģeogrāfiskās kartēšanas un ģeodatu izguves iespējas LU LFMI digitālajā resursā *literatura.lv*, gan *Google Maps* datu vizualizācijas iespējas⁵. Šajā kartē reģistrēts, cik biedru kopumā devušies braucienos ārpus

- 5 *Google Maps* ir ērts un populārs ģeodatu vizualizācijas rīks, ko izmanto daudzos digitālos projektos, piemēram, apjomīgu literāro tekstu korpusu kartēšanā u. c. (Cooper et al. 2016: 12).

PSRS, cik daudz un cik reižu apmeklētas valstis, reģioni, pilsētas u. tml. Kartēšanas procesu sarežģī PSRS un sociālistiskā bloka sabrukuma izraisītās ģeopolitiskās pārmaiņas – daļa no sociālistiskajām valstīm (Čehoslovākija, Dienvidslāvija, VDR, VFR u. tml.) un pilsētām (Rietumberlīne, Austrumberlīne) mūsdienās vairs neeksistē, līdz ar to nav iespējams precīzi noteikt šo vietu ģeogrāfiskās koordinātas. Tās digitālās kartēšanas procesā tika noteiktas aptuveni un ievadītas manuāli.

Šāda tipa projekti ir papildu veids, kā saskatīt plašāku ainu, tendences vai detaļas gan braucieni, gan ceļojumu aprakstu izpētē (Belgum et al. 2018: 320). RS biedru braucieni datu korpusa un digitālo karšu izveide *Google Maps* un vortālā *literatura.lv* ir noderīgs resurss ne tikai latviešu literatūras vēstures, bet arī citu nozaru pētniekiem.

Latvijas Padomju rakstnieku savienības biedru braucieni sociālpolitiskās tendences

Izpētot savāktu datu korpusu par braucieniem no 1970. līdz 1989. gadam, redzams, ka RS biedri ārpus Padomju Savienības braucienos devās PSRS Rakstnieku savienības, Latvijas un ārzemju kultūras sakaru biedrības, Latvijas Komitejas kultūras sakariem ar tautiešiem ārzemēs, Latvijas Miera aizstāvēšanas komitejas delegāciju sastāvā, kā arī dažādu starpvalstu draudzības biedrību uzdevumā, plašākās kultūras delegācijās un grupu tūrisma braucienos, izmantojot “Inturist” piedāvātos maršrūtus. Īpašos gadījumos notika arī braucieni uz personīga ielūguma pamata.

Apskatāmajā laika posmā RS biedri visvairāk apmeklēja sociālistiskās valstis jeb t. s. satelītvalstis, kas arī bija Padomju Savienības starptautiskā tūrisma pamatā. Šo valstu struktūras, kas bija atbildīgas par starptautisko tūrismu, bija institucionāli līdzīgas Padomju Savienībai (Orlov, Popov 2016). Rakstnieki devušies gan tūrisma braucienos, gan komandējumos ar nolūku stažēties, piedalīties semināros, vasaras skolās, kultūras pasākumos, arī kongresos, plēnumos. Kvalifikācijas celšanas nolūkos visvairāk ceļojuši tulkotāji. Kopumā uz VDR veikti vismaz 88 braucieni, uz Čehoslovākiju – 56 braucieni, Bulgāriju – 40 braucieni, Poliju – 38 braucieni, Rumāniju – 35 braucieni, Dienvidslāviju – 34 braucieni, Ungāriju – 26 braucieni.

Ceļojumiem uz visvairāk apmeklēto satelītvalsti – VDR –, kā izpētījuši vēsturnieki Igors Orlovs un Aleksejs Popovs (*Aleksej Popov*), 70. gados tūristiem bija pieejami vairāk nekā 50 maršruti, savukārt 80. gados tādu bija jau 60 (turpat). Tas izskaidro arī braucieni popularitāti RS biedru vidū, kā arī jāņem vērā ar Otrā pasaules kara saistītā tematiskā līnija – populāra bija memoriālo kompleksu (piemēram, Būhenvalde), muzeju u. c. politiski ideoloģisku objektu apmeklēšana (turpat). Piemēram, rakstnieks Miervaldis Birze, kurš Otrā pasaules kara laikā atradās koncentrācijas nometnēs Salaspilī un Būhenvaldē, vairākkārt aicināts un apmeklējis dažādus Otrā pasaules kara piemiņas pasākumus VDR.

Populārs braucieni mērķis bija arī Čehoslovākija, sevišķi tās kūrortpilsēta Karlovi Vari. Turp ārstniecības nolūkos devās vairāki Rakstnieku savienības biedri (piemēram, Imants

Ziedonis, Jāzeps Osmanis, Ērika Lūse). Būtiski arī atzīmēt biežos braucienus uz sadraudzības pilsētām – Rostoku VDR, Ščecinu Polijā un Rusi Bulgārijā –, tādējādi ierindojošot mīnētās valstis visvairāk apmeklēto vidū.

RS biedru vidū tūrisma braucienu uz Rietumiem ir mazāk. Pamatojums tam rodams ierobežotajā braucienu pieejamībā, ko apstiprina arī dramaturga Pētera Pētersona paustais mutiskā atskaitē par braucienu uz dramaturģijas semināru VFR 1978. gadā: “Tā bija reta iespēja, jo mūsu zemes pārstāvji uz semināriem ārzemēs brauc reti. [...] Kas attiecas uz mūsu Rakstnieku savienību, es nesaprotu, kāpēc viens var braukt ārzemju komandējumos un ceļojumos, bet cits nē.” (LNA LVA 473-1-504: 157) RS sēžu protokolos nav rakstisku atskaišu par komandējumiem uz sociālistiskajām, t. s. kapitālistiskajām un eksotiskajām valstīm, kas liedz gūt pilnīgu priekšstatu par veiktajiem braucieniem. Dažkārt sniegtas tikai īsas, mutiskas atskaites, sēdē lakoniski informējot pārējos sekretariāta vai KP pirmorganizācijas biedrus par braucienu norisi.

Pēc dokumentu analīzes secināms, ka braucieni uz Rietumvalstīm – Zviedriju (aptuveni 31 brauciens), VFR (28 braucieni), Somiju (22 braucieni), Kanādu (14 braucieni), Lielbritāniju (12 braucieni), ASV (11 braucieni), Itāliju, Grieķiju, Franciju (uz visām 10 braucieni), Dāniju (7 braucieni) u. c. valstīm – šajā laikā notikuši dažādu organizāciju delegāciju sastāvā. RS biedri uz Rietumvalstīm devās pārsvarā profesionālos nolūkos – piedalīties konferencēs, semināros, simpozijos u. tml. Piemēram, 1985. gada jūnijā Stokholmā notika Krišjāņa Barona 100. gadadienai veltīta konference. Šajā pasākumā piedalījās 11 RS biedri (Imants Auziņš, Jānis Baltvilks, Jevsejs Bareinboms, Harijs Gulbis, Jānis Kalniņš, Olga Lisovska, Māra Misiņa, Jānis Peters, Nikolajs Zadornovs, Imants Ziedonis un Saulcerīte Viese). Taču šādiem braucieniem bija arī ideoloģiski politisks mērķis – propagandēt Padomju Savienības sasniegumus, kultūru un padomju literatūru, kā arī dibināt kontaktus un izpētīt situāciju trimdā, tātad sava veida spiegošana. Rakstnieki bija viens no padomju propagandas instrumentiem. Piemēram, dzejniece Lija Bīdaka RS KP pirmorganizācijas sēdē par braucienu uz Dāniju PSRS Rakstnieku savienības delegācijas sastāvā 1975. gada septembrī teikusi: “Mūsu ceļojums radīja zināmu satraukumu, jo jau gadu desmitiem neviena padomju rakstnieku grupa Dānijā nebija bijusi. Mūs brīdināja, ka varam tikt iejaukti kādā pretpadomju pasākumā. Braucām turp saspringti un ideoloģiski bruņojušies. [...] Kopenhāģenu redzējām steidzīgu ekskursantu acīm. [...] Nekādu provokāciju nebija, interese par mums bija draudzīga.” (LNA LVA 473-1-341: 87) Savukārt rakstniece un tulkotāja, tolaik arī viena no RS valdes sekretārēm Anna Žīgure, kura arī bija atbildīga par ārzemju braucieniem un bija daudz ceļojusi, RS atklātajā KP sapulcē 1986. gada 29. oktobrī uzsvērusi: “Šie braucieni ir bāze jebkuram kontrpropagandas darbam. Ja nav zināšanu, ne pret ko nevar cīnīties. Pretinieks ir ļoti jāpazīst. Mēs bieži nezinām, pret ko mēs īsti cīnāmies. 50to un 60to g. pieņēmumi vairs neder! Jābrauc pašiem! [...] Rakstnieka liels un svēts pienākums ir pastāstīt par redzēto, un rakstniekiem tic.” (LNA LVA PA-7263-1-44: 144, 145) Šajā sanāksmē tika nolemts, ka rakstniekiem, kuri ir bijuši ārzemēs, regulāri jālasa lekcijas par savu ceļojumu pieredzi un iespaidiem citiem rakstniekiem un dažādiem darba kolektīviem. Savukārt tā brīža RS valdes pirmais sekretārs Jānis Peters pauda atšķirīgu viedokli, norādot, ka literatūrai, arī trimdas, nevajadzētu būt nekādam sakaram ar politiku (turpat: 144).

Ārzemju braucienu tendenču atklāsmē jāizceļ vēl plašāki ceļojumi pa Eiropu vai kruīzi pa Baltijas un Melno jūru, Vidusjūru un Ziemeļjūru. Šādos braucienos gan varēja piedalīties ne visi biedri, bet īpašāka kategorija, piemēram, rakstnieks Žanis Grīva, kurš ieņēma nozīmīgus amatus padomju nomenklatūrā – viņš bija PSRS un Latvijas PSR Augstākās Padomes deputāts, LKP CK locekļa kandidāts un loceklis, Latvijas PSR Miera aizstāvēšanas komitejas priekšsēdētājs, Padomju Miera aizstāvēšanas komitejas prezidijs loceklis un Vispasaules Miera padomes loceklis. Grīvas gadījumā akcepts izbraukšanai no RS vadības nebija vajadzīgs⁶. Padomju Savienība centās uzturēt un izveidot attiecības ar jaunatīstības jeb t. s. Trešās pasaules valstīm (piemēram, Āfrikas valstīm), ar kurām Padomju Savienībai bija noslēgti sadraudzības līgumi, tirdzniecības sakari u. tml. Visvairāk apmeklētā bija Indija, ar kuru Padomju Savienībai bija ciešas attiecības. RS biedri visbiežāk uz eksotiskajām valstīm devās kā tūristi, nereti brauciena nolūks bija sagatavot arī reportāžu, aprakstu kādam preses izdevumam.

Apkopojot kvantitatīvos datus par RS biedru braucieniem 70. un 80. gados, redzams, ka kopumā ārpus Latvijas PSR, tostarp arī PSRS iekšienē, veikti vismaz 606 braucieni, savukārt 80. gados – 575. Visvairāk braucienos šajos gados devušies: Imants Ziedonis (32 braucieni), Gunārs Cīrulis, Pēteris Zirņītis (17 braucieni), Jānis Peters, Māris Čaklais (15 braucieni), Andrejs Dripe (13 braucieni), Andris Vējāns (13 braucieni), Ēvalds Strods (11 braucieni), Anna Žigure, Lija Brīdaka (10 braucieni). Braucienu kopskaits ir krietni lielāks nekā ceļotāju skaits. RS KP organizācijas atklātajā sapulcē 1986. gada 29. oktobrī Žigure ziņoja, ka 80. gados no visām Padomju Savienības republikām visvairāk uz ārzemēm ceļojuši tieši RS biedri; ik gadu ap 30–40 rakstnieki vairākas reizes gadā varējuši izbraukt uz ārzemēm (LNA LVA PA-7263-1-44: 144).

Tomēr jāatzīmē, ka iegūtie dati no RS un tās KP pirmorganizācijas dokumentiem RS biedru ceļošanas tendences neatspoguļo pilnībā, jo dokumentos fiksēti tikai tie biedru braucieni, kuriem bijusi nepieciešama RS vadītāja rekomendācija, kā arī daļa rakstnieku ārzemju braucienos devās arī no citām organizācijām (tostarp no citām radošajām savienībām, kurās viņi bija arī biedri), piemēram, no LPSR Komponistu savienības – Marģeris Zariņš, LPSR Mākslinieku savienības – Uldis Zemzaris. Turklāt informācija par ārzemju braucieniem RS dokumentos ir samērā regulāra tikai līdz 1985. gadam. Vēlākā posma dokumentos informācija ir fragmentāra un skopa; tas saistāms ar politiskajām un ekonomiskajām pārmaiņām valstī līdz ar Gorbačova nākšanu pie varas, kad iespējas ceļot un doties ārpus Padomju Savienības kļuva pakāpeniski arvien pieejamākas un šim nolūkam vairs nebija nepieciešams saskaņojums KP struktūrās.

Visaptverošam Latvijas PSR rakstnieku ceļošanas tendenču apskatam nepieciešams izpētīt vēl citus LNA LVA fondus, kas saistīti ar tūrismu un izbraukšanu ārpus Latvijas PSR⁷. Taču datu korpusā un digitālajās kartēs apkopotajos datos redzams, ka braucienu biežums pētījumā apskatītajā laikā posmā ir pietiekami liels, kas norāda gan uz RS biedru privilēģēto

6 Plašāk par Žaņa Grīvas statusu skat. Eversone 2021.

7 Jāņem vērā arī, ka visi dokumenti nav pieejami, jo iznīcināti vai vispār nav arhivēti.

statusu, gan ceļošanas pieejamību, paplašinot apmeklēto valstu ģeogrāfiju, kas redzama RS biedru braucienų digitālajās kartēs.

Ceļojumu atspoguļojums publicistikā

Padomju periodā rakstītās ceļojumu piezīmes var iedalīt četrās tematiskās grupās: ceļojumi pa Latviju, ceļojumi pa Baltijas valstīm, ceļojumi pa PSRS teritoriju un ceļojumi ārpus PSRS uz sociālistiskajām un citām pasaules valstīm (Burima 2018: 67), savukārt tūrisma braucienus var iedalīt vismaz trīs grupās: braucieni, par kuriem rakstītas publicistiskas vai beletrizētas ceļojumu piezīmes; braucieni, kuros gūtā pieredze pārkausēta literāros darbos; braucieni, par kuriem piezīmes nav publicētas vai arī tās publicētas pēcpadomju periodā (Oga 2021: 128). Gan publicistisko, gan beletrizēto ceļojuma piezīmju klāstā pārstāvēti gan apraksti par braucieniem pa PSRS teritoriju, gan ceļojumi uz sociālistiskajām jeb t. s. satelītvalstīm, Rietumu kapitālistiskajām un eksotiskajām valstīm.

Atsauces uz ceļojumiem plašāk jaušamas dzejā un publicistikā. Piemēram, braucieniem pa PSRS veltītas Imanta Ziedoņa, ukraiņu rakstnieka Vitālija Korotiča (*Vitalij Korotich*) un fotogrāfa Gunāra Janaiša Tadžikijas ceļojuma (1971) piezīmes “Perpendikulārā karote” (1972); ceļojumiem uz satelītvalstīm – Ritas Luginskas “iespaidu mozaika” par Polijas apmeklējumu 1968. un 1970. gadā “Polonēzē pa Poliju” (daļēji periodikā 1968, grāmatā 1973) un Miervalža Birzes apraksti par VDR (1970) un Čehoslovākijas (1972) apmeklējumu rakstos “Ceļojums pa Harcu” un “Prāga, kastaņu ziedā” (periodikā attiecīgi 1970 un 1972, grāmatā “Lai atgrieztos”, 1984). Rietumvalstu apciemojumi fiksēti Birzes piezīmēs par ceļojumiem uz Dāniju, Zviedriju un Somiju (1967), VFR (1980) un Franciju (1984) (periodikā attiecīgi 1967, 1980 un 1984, grāmatās “Lai atgrieztos”, 1984, un “Tāli gadi, tuvi laiki”, 1987), Māra Čaklā Dānijas ceļojuma aprakstos (grāmatā “Nozagta gliemežnīca”, 1980), Andreja Dripes aprakstos par 1981. un 1984. gada braucieniem uz VFR (periodikā 1981 un 1984, grāmatā “Līdzatvestais”, 1986). Piezīmes par kopīgu viesošanos ASV (1975) rakstījis gan Jānis Anerauds (grāmatā “Zem saules – cilvēku zemē”, 1984), gan Jānis Peters (grāmatā “Kalējs kala debesīs”, 1981), savus iespaidus pēc ASV apmeklējuma publicējuši arī Arvīds Grigulis (ceļojums 1971. gadā; dzeja periodikā 1971 un krājumā “Marginālijas angļu valodas vārdnīcā Amerikas ceļojuma laikā”, 1972) un Marģeris Zariņš (ceļojums 1975. gadā; publikācija periodikā 1976, grāmatā “ppf: Pārdomas. Pieraksti. Feļetoni”, 1980). Ceļojumus uz eksotiskajām valstīm atspoguļo, piemēram, Zentas Ērgles piezīmes “No Plutona līdz Neptūnam” par ceļojumu uz Āfriku 1973. gadā (periodikā daļēji 1973, grāmatā 1975) un “Gads ar diviem pavasariem” par 1976. gada ceļojumu uz Jaungvineju, Austrāliju un Jaunzēlandi (periodikā 1977 un 1978, grāmatā 1979), Artūra Heniņa apraksts par Islandi “Atlantijas savādniece” (1973), Jāņa Anerauda piezīmes “Kontrasti” par 1971. gada braucienų uz Latīņameriku grāmatā “Zem saules – cilvēku zemē” (1984). Andra Vējāna piezīmes par braucieniem Miera komitejas delegāciju sastāvā uz Kanādu, ASV, Kubu, Somiju, Zviedriju, Dāniju, Franciju, Lielbritāniju, Nikaragvu publicētas grāmatā “Vulkāns un asara” (1987). Ierobežotā informācijas

pieejamība stagnācijas apstākļos pieļauj, ka ceļojumu iespaidus kā nenovecojušus iespējams lasītājiem pasniegt grāmatas formātā arī 10–15 un vairāk gadus pēc brauciena norises brīža.

Turpmāk rakstā analizēti Latvijas Padomju rakstnieku savienības biedres, rakstnieces Zentas Ērgles ārzemju braucieni 70. un 80. gados, izvēloties fokusu uz valstīm aiz dzelzs priekškara un atklājot tendences šo ceļojumu norisē un iespaidu atklāšanā publicistikā.

Zentas Ērgles piezīmes par ceļojumiem uz eksotiskajām un Rietumu valstīm

Rakstniece un arhitekte Zenta Ērgle (1920–1998) bija viena no pazīstamākajām latviešu bērnu un pusaudžu rakstniecēm 20. gadsimta 50.–80. gados. Līdzās bērniem un jauniešiem rakstītajiem darbiem publicējusi arī ceļojumu piezīmes – “No Plutona līdz Neptūnam” (1975) un “Gads ar diviem pavasariem” (1979).

Ērgle savu Āfrikas piezīmju ievadā rakstījusi: “Tālajā bērībā blakus manai gultai karājās sienas sega. Tajā bija attēloti džungļi. [...] Vakaros aizmigdama es prātoju: kad izaugšu liela, aizbraukšu uz Āfriku un pastaigāšos pa īstiem džungļiem. [...] Bērnu dienu piedzīvojumu literatūru nomainīja nopietnās grāmatas – Gržimeka, Hantera un citu autoru stāsti par Āfriku. Ilgas pēc dienviņu eksotikas nespēja apslāpēt pat divkārsa darba slodze un ģimenes rūpes.” (Ērgle 1975: 6, 7) Līdzīgi izteicies rakstnieks Zigmunds Skujiņš, kurš nodēvējis to par “sapņu revīziju” (Skujiņš 2006: 15), apliecinot Jurčaka tēzi par ārzemēm kā iedomātu telpu.

Zenta Ērgle tūrisma ceļojumos dodas, jo viņas grāmatas ir populāras un tiek tulkotas svešvalodās, viņa ir to rakstnieku vidū, kuri var atļauties dārgos ceļojumus; paralēli rakstniecībai strādā par inženieri arhitekti Zemes ierīcības projektēšanas institūtā (1949–1975), ir gan RS, gan LPSR Arhitektu savienības biedre, nav bijusi LKP biedru vidū.

“– Vai tev nepietiek? – piederīgie brīnās, redzēdami uz galda izklātās kartes un labi zinādami, ko tas nozīmē. Ko lai dara, pret ceļošanas slimību, tāpat kā pret cece mušu un miega slimību, pagaidām vēl jāļū nav,” (Ērgle 1975: 85) tā Ērgle noslēgusi Āfrikas ceļojuma piezīmes, ceļošanu salīdzinot ar atkarību, pret kuru nav iespējams cīnīties, bet kuru, tieši otrādi, viņai ir iespējams izbaudīt statusa un privilēģiju dēļ.

Ērgles apmeklētās valstis liecina arī par zināmu atkāpi no pieņemtās tendences padomju inteliģencei saņemt atļauju doties uz eksotiskajām valstīm vai Rietumvalstīm tikai pēc tam, kad apmeklētas t. s. satelītvalstis. RS dokumenti par 60.–80. gadiem uzrāda šādus Ērgles ceļojumus: Armēnija (1965), Tālie Austrumi, Kamčatka un Kuriļu salas (1973), Gruzija (1974), Azerbaidžāna (1976), Bulgārija (1982, jūlijs–augusts, LPSR un ārzemju draudzības un kultūras sakaru biedrības speciālās tūristu grupas sastāvā uz 14 dienām). Publicētā informācija liecina, ka 1967. gadā Ērgle apmeklējusi arī Dāniju, Zviedriju un Somiju (Auziņš et al.

1971: 75). Šāda atļauja, iespējams, dota tāpēc, ka jau “atkušņa” laikā, 1956. gadā, viņa kopā ar citiem kolēģiem arhitektiem apmeklējusi Poliju (Ērgle 1956). Tādējādi Ērglei 70. gados nav formālu šķēršļu braucienam uz eksotiskajām un Rietumvalstīm.

Grāmatās “No Plutona līdz Neptūnam” un “Gads ar diviem pavasariem” (katra grāmata iespiesta 15 000 eksemplāru metienā) aprakstītie ceļojumi⁸ uz Āfrikas valstīm (Malagasijas Republiku, Reinjonu, Maurīcijas Republiku, Tanzānijas Apvienoto Republiku) 1973. gadā un Jaungvineju, Austrāliju un Jaunzēlandi 1976. gadā RS arhīva dokumentos nav minēti. Braucienus organizējusi PSRS un ārzemju draudzības un kultūras sakaru biedrība, un šī informācija minēta publicētajos ceļojumu aprakstos, attiecīgi grāmatas “No Plutona līdz Neptūnam” ievadā (Ērgle 1975: 8) un grāmatas “Gads ar diviem pavasariem” izdevēja anotācijā (Ērgle 1979: [120]).

Šajos PSRS un ārzemju draudzības un kultūras sakaru biedrības organizētajos ceļojumos dodas nelielas tā saucamās specgrupas, piemēram, par braucienu uz Āfriku Ērgle rakstījusi, ka “[e]sam deviņi “laimīgie”. Bez vadītāja un PSRS un Āfrikas kultūras sakaru biedrības pārstāvja tūristu grupā vēl ir divas Maskavas pasniedzējas, ģeoloģijas zinātņu doktors, radiožurnālists un kinooperators. No Rīgas esam divas – māksliniece M. Stāraste un es.” (Ērgle 1975: 8) Ceļojumā uz Jaungvineju, Austrāliju un Jaunzēlandi devušies pavisam 19 delegāti (Ērgle 1979: 11), no Latvijas arī LPSR un ārzemju draudzības un kultūras sakaru biedrības prezidents (1970–1977) Žanis Zakenfelds, kurš bijis delegācijas vadītājs (turpat: 101), un komponists Raimonds Pauls (turpat: 5), savukārt pārējo ceļotāju vidū bijuši nomenklatūras un inteliģences pārstāvji: PSRS Augstākās Padomes deputāts Postņikovs, “Ļeņingradas skolotāja Valentīna Talalaikina” (turpat), “Šuvalovs, kādreizējais sacīkšu braucējs” (turpat: 111), “Sociālistiskā Darba Varonis Luka Ivaņiščenko, Hersonas rajona Radzjankas kolhoza priekšsēdētājs” (turpat: 118), armēņu mākslinieks Zulans Grigorjans (turpat) un citi, kuru biogrāfijas un uzvedība iepriekšējos ceļojumos uz satelītvalstīm ļāvuši saņemt braukšanas atļauju un kuri finansiāli varējuši šādu ceļojumu atļauties. Kādā intervijā Ērgle Austrālijas brauciena sakarā izteikusies: “Ceļošana ir ļoti dārgs prieks. Bet man nemaz nav žēl!” (Ločmelis 1980: 169)

Brauciena dalībniekiem ceļojuma laikā bijuši arī oficiāli pienākumi – tikšanās ar Jaunzēlandes un PSRS draudzības biedrības biedriem, PSRS vēstniecības pārstāvjiem –, tādējādi šie padomju tūristi nav bijuši brīvdienās, bet pildījuši tādu pašu propagandas darbu kā PSRS mūziķi vai dejotāji, kuri ar koncertiem viesojušies Jaunzēlandē kultūras apmaiņas nolūkos (1970. gadā PSRS viesojies arī maoru ansamblis un baritons Inia Te Viata (*Inia Te Wiata*)). Ērgle, atsaucoties uz biedrības viceprezidentu Levisu, Jaunzēlandes radio un televīzijas ministru Templtonu un pašu delegācijas vadītāju Zakenfeldu, atzīst, ka “ir labi, ja dažādu zemju un pārliecību ļaudis satiekas, iepazīst cits citu un sadraudzējas” (Ērgle 1979: 101), taču neslēpj arī sarūgtinājumu, ka brauciena dalībnieki “[i]espiesti jau iepriekš sastādītā un apstiprinātā

8 Ar digitālo humanitāro zinātņu metodēm vortālā *literatura.lv* datubāzē izveidotas grāmatās minēto vietu kartes: <https://www.literatura.lv/lv/work/No-Plutona-lidz-Neptunam/406593>; <https://www.literatura.lv/lv/work/Gads-ar-diviem-pavasariem/406594>. Minēto daiļdarbu teksti nav ievietoti sistēmā autortiesību ierobežojumu dēļ.

maršruta Prokrusta gultā” (turpat: 104), šo tendenci gan skaidrojot ar to, ka “dabiska un saprotama ir kapitālistisko valstu tūrisma firmu cenšanās iebrāucējiem, vēl jo vairāk Padomju zemes pārstāvjiem, parādīt spožu savas valsts parādes fasādi”, un uzsverot, ka “[t]ikšanās reizēs ar Jaunzēlandes un PSRS draudzības biedrības biedriem, tāpat sarunās ar latviešu un krievu emigrantiem uzzinājām arī par medaļas otro – ēnas pusi” (turpat: 116).

Ērgles ceļojumu piezīmes “No Plutona līdz Neptūnam” un “Gads ar diviem pavasariem” adresētas pieaugušajiem lasītājiem, un tās raksturo blīvs ģeogrāfisko un politisko faktu un personīgo iespaidu kopums, ko veidojusi intensīvā gatavošanās ceļojumiem un to laikā fiksētā informācija un emocijas. Par autores erudīciju un avotu apguvi (“Pirms braukšanas salasījos visādus briesmu stāstus.” (Ērgle 1979: 5)) liecina grāmatā izmantotā zinātniskā un populārzinātniskā literatūra – čehu etnologa Miloslava Stingla (*Miloslav Stingl*), norvēģu ceļotāja Tūra Heijerdāla (*Thor Heyerdahl*), vācu zoologa Bernharda Gržimeka (*Bernhard Grzimek*) u. c. autoru darbi, Latvijā un PSRS pieejamā satelītvalstu prese, arī Padomju Latvijā oficiāli nepieejami izdevumi, kā, piemēram, “Zeme un tautas” (1929–1931) –, kā arī oficiāli statistikas un propagandas materiāli, piemēram, Papua-Jaungvinejas premjerministra raksti un runas (turpat: 49–51) un pārskati par jaunattīstības valstu ekonomiku.

Gan Āfrikas, gan Austrālijas un Jaunzēlandes ceļojumu piezīmēs uzsvērts, ka vietējie iedzīvotāji padomju tūristus sastop reti vai šīs grupas ierašanās ir pirmā reize: “Padomju cilvēki Madagaskarā ir reti viesi. Aerodroma kalpotāji mūs aplūko ar vāji slēptu interesi. Mēs darām to pašu.” (Ērgle 1975: 9); “Mūsu sarunā piedalās košās drānās tērptie papuasau jaunieši. Baltie zobī mirdz draudzīgos smaidos. Viņi pirmo reizi redz padomju cilvēkus. Skan balsis: – Lenin, dražba, Moskau!” (Ērgle 1979: 28) Šādi tiek īstenota savdabīga dubultā koloniālisma pieredze, satiekoties jaunatklājējiem no abām pusēm, taču šīs attiecības neraksturo bailes vai pakļaušanās, vien abpusēja interese, kuru neietekmē pastāvošas vai nākotnē paredzamas saistības.

Austrālijas un Jaunzēlandes stāstījuma otrā daļa saucas “Pa kapteiņa Granta bērnu pēdām”, un Ērgle vairākkārt citē latviešu lasītājam pazīstamo franču rakstnieka Žila Verna (*Jules Verne*, 1828–1905) romānu “Kapteiņa Granta bērni” (*Les Enfants du capitaine Grant*, 1865–1867, latviski 1926) un atsaucas uz tā tekstu, piemēram: “Bet atgriezīsimies pie Vaikato upes. No šejienes sākas Verna varoņu ceļojums 20 metru garajās pirogās pret straumi uz dienvidiem – uz Taupo ezeru. Mēs to pašu ceļu veicam dažās stundās ar autobusu.” (turpat: 63) Pētījuma izvērsumā iespējams apsvērt Verna romāna tulkojuma teksta kartēšanu un, izmantojot datu slāņu pievienošanas un atslēgšanas metodi, salīdzināt un analizēt Verna romāna varoņu maršrutu Jaunzēlandē 19. gadsimta 60. gados un Ērgles ceļojuma laikā apmeklētās vietas pēc vairāk nekā simt gadiem – 1976. gadā.

Ceļojumu piezīmēs minētās vietas iedalītas piecās kategorijās, rakstnieku ceļojumu maršrutu un apmeklēto vietu attēlošanai nedaudz transformējot Barbaras Pjati (*Barbara Piatti*) piedāvāto literārajos darbos minēto vietu reģistrēšanas un kartēšanas piecu pakāpju modeli (Piatti 2016: 92–93), kuru Daija un Kalnačs reducējuši trīspakāpju modelī (skat. Daija un Kalnačs 2019b: 159). Attiecīgi Ērgles piezīmēs minētās vietas grupētas šādās kategorijās:

1. vietas, kas ietilpst vispārējā brauciena maršrutā, piemēram, Āfrikas ceļojumā, ietverot apmeklētās valstis vai pilsētas: Rīga–Maskava–Parīze–Džibuti–Antananariva–Sendenī–Portluī–Daressalama–Entebe–Kaira–Beiruta–Erevāna–Maskava–Rīga (Ērgle 1975: 85);
2. katrā valstī un pilsētā apmeklēto vai minēto vietu detalizēts uzskaitījums, piemēram, Jaunzēlandē: Oklenda, Kvīnstrīta un Domeina parks Oklendā, Vaitematas osta; arī viesnīcas un restorāni u. tml. vietas (tostarp vairs neeksistējošas (piemēram, Malagasijas Republika), norādot to atrašanās vietu tieši vai aptuveni);
3. piezīmju tekstā minētās vietas, kas saistītas ar brauciena maršrutu, ne visas no tām tikušas apmeklētas. Šī kategorija pilnībā vai daļēji pārklājas ar otro kategoriju;
4. piezīmju tekstā minētās vietas, kas nav saistītas ar brauciena maršrutu, taču atspoguļo autora erudīciju un tēlaino domāšanu, piemēram, salīdzinot apceļotās vietas ar Latvijas ģeogrāfiju vai citos braucienos apmeklētajām vietām; piemēram sniegti citāti no Ērgles piezīmēm: “Blakus tirgum paceļas dzīvojamo ēku jaunceltņu rajons, gandrīz tāds pats kā pie mums Āgenskalnā, tikai priežu vietā kokospalmas” (turpat: 49); “[n]olaidenie, kokiem apaugušie kalni vietām atgādina Krimu, stāvas klinšu nogāzes, pa kurām putodami un šalkdami gāžas lejup ūdenskritumi, – Kaukāzu, bet vietām ar savu skarbumu un varenību – Tjanšanu.” (Turpat: 21) Šajā kategorijā ietilpst vietas, kas raksturo kraso nošķirumu “pazīstamā/vietējā” un “nepazīstamā/svešā” telpa Rietumvalstu un eksotisko valstu apmeklējumu pieredzē. Kā atzinuši Daija un Kalnačs, šo atšķirību “izpētes rezultātus var pamatot, katalogizējot asociācijas un stereotipus, kas pārstāvēti korpusa lasīšanas procesā apkopotajos [romānu] fragmentos” (Daija, Kalnačs, 2019a: 76). Šai kategorijā ietilpst arī piezīmju tapšanas laikā pastāvējušās, bet mūsdienās vairs neeksistējošās valstis, kā Čehoslovākija, Dienvidslāvija u.tml.;
5. piezīmju tekstā minētās, mūsdienās neeksistējošās (zudušās) vēsturiskās, mītiskās, leģendās tēlotās vietas, piemēram, “[a]pmēram 160 kilometru uz ziemeļrietumiem no Tahiti atrodas Raiatea saliņa, kādreizējā leģendārā Havaiki – senās Polinēzijas reliģiskais centrs” (Ērgle 1979: 65); “[s]enajā pasaulē bija pazīstami septiņi brīnumi: Ēģiptes piramīdas, Semiramīdas gaisa dārzi Bābelē [Babilonā], Artemīdas templis Efesā, Zeva statuja Olimpijā, Halikarnasas mauzolejs, Rodas koloss un Aleksandrijas bāka.” (Ērgle 1979: 74)

Pirmās divas kategorijas apkopo ceļojuma dalībnieka apmeklētās vietas, kā arī padomju perioda ceļojumu maršrutu tendences, piemēram, obligātā izlidošana no Maskavas, pārsēšanās un ekskursijas maršruts Parīzē ceļā uz Āfriku (“Ko tūrists var ieraudzīt Parīzē dažu stundu laikā? [...] Atstājot Parīzi, sapratām, kā jūtas izbadēties cilvēks, ja tam no bagātīgi klāta galda atļauj nogaršot vienu vienīgu kumosu.” (Ērgle 1975: 8)) Šīs kategorijas, vizualizētas kā atsevišķi datu un karšu slāņi, ļauj veidot papildu slāņus, transformējot vai papildinot pastāvošos, piemēram, ar minēto Verna romāna “Kapteiņa Granta bērni” varoņu pārvietošanās karti, citu autore lasīto autoru (Heijerdāla, Gržimeka) maršrutu kartēm u. tml., fiksējot kopīgi minētās un apmeklētās vietas un maršrutu paralēlās daļas. Skatot trešās, ceturtais un piektās kategorijas slāņus apvienojumā, tiek iegūta visu ceļojuma piezīmju tekstā minēto vietu karte.

Izmantojot *Google Maps* programmatūru, iespējams arī izkalkulēt visa ceļojuma vai tā daļu kilometrāžu un salīdzināt to ar ceļotāju kalkulācijas datiem, ja tādi sniegti (Āfrikas ceļojuma piezīmēs Ērgle minējusi, ka “ar dažādu valstu lidmašīnām un automašīnām nobraukti ap 33 000 kilometru” (turpat: 85)).

Ceļojuma piezīmēs minēts plašs vietu skaits⁹: “No Plutona līdz Neptūnam” – 265 unikālas vietas vai objekti, “Gads ar diviem pavasariem” – 322 unikālas vietas vai objekti, no kurām attiecīgi 163 (62 %) un 216 (67 %) saistītas ar konkrēto ceļojumu maršrutiem. Atšķirīgu paņēmieni autore izmanto bērniem adresētajās grāmatās, kuras iedvesmojuši ceļojumi, – šajos darbos minētas tikai vietas Āfrikā un Austrālijā, ar izņēmumu stāsta “Ceļojošais cirkus “Matilde”” ievadā un nobeigumā, tēlojot autorei nokļūšanu Austrālijā un atgriešanos no tās: “Šeit nelielā Zelta piekrastes pilsētiņā mūsu ceļi šķīrās. Ceļojošā cirka “Matilde” ļaudis devās jaunā braucienā pa Austrāliju, bet mani lidmašīna nesa pāri teiksmainajām Indonēzijas salām un Indijai atpakaļ uz Dzimteni.” (Ērgle 1981: 19)

Zentas Ērgles ceļojumu aprakstus gan periodikā, gan grāmatās papildina fotogrāfijas, krāsainie diapozitīvi, kurus rakstniece demonstrējusi arī tikšanās reizēs skolās un bibliotēkās, kā arī īpašos stāstu vakaros šaurā lokā, tādējādi īstenodama savu iespaidu pārnesi vairākās pakāpēs un pielāgodama to dažādām auditorijām – piezīmju lasītājiem periodikā un grāmatās, bērniem viņiem adresētajos darbos un, visbeidzot, klātienē tikšanos apmeklētājiem.

Nobeigums

Latvijas Padomju rakstnieku savienības biedru 70.–80. gadu braucieni un ar tiem saistītā publicistika ir turpmāk plašāk pētāms jautājums. Rakstā ar digitālo humanitāro zinātņu metodēm apkopotie dati, karšu izveide vortālā *literatura.lv* un *Google Maps*, piemēri par rakstnieku braucieniem tendencēm un atsevišķu rakstnieku (Zentas Ērgles) ceļojumu un darbu padziļinātā analīze ir viens no veidiem radošās inteliģences tūrisma dinamikas izpētē periodā, kad šīs iespējas Latvijas PSR bija limitētas un ekskluzīvas. Tādējādi rakstā atspoguļotas arī radošās inteliģences privilēģijas un loma padomju ideoloģijā. Pētījuma procesā aizsāktā rakstnieku braucieniem un to darbu digitālā kartēšana un datu vizualizācija ir viens no mēģinājumiem latviešu literatūras vēstures izpētē izmantot digitālo humanitāro zinātņu metožu iespējas, kas palīdz sistematizēt un pārskatāmāk analizēt ceļojumu maršrutus un to intensitāti, kā arī ļauj precīzāk saskatīt braucieniem sociālpolitiskos aspektus – braucieniem mērķus, braucieniem norisē iesaistītās organizācijas, politisko situāciju un sociālos apstākļus u. tml. Publiski pieejamie dati un digitālās kartes par rakstnieku veiktajiem braucieniem LU LFMI vortālā

9 Šajā pētījumā, veidojot ceļojumu piezīmēs un dailliteratūrā minēto vietu sarakstu un kartes, šo vietu atkārtoti pieminējumi tiek reģistrēti, taču aprēķinā netiek ņemti vērā, atšķirībā no Daijas un Kalnača pieejas dailliteratūras darbu kartēšanā (Daija, Kalnačs 2019b: 155). Atkārtoti vietu pieminējumi tiks analizēti rakstnieku un viņu daiļdarbu varoņu ceļojumu maršrutu kartēšanā pētījuma turpinājumā.

literatura.lv ir vērtīgs avots un vizualizācijas rīks dažādu nozaru pētniekiem, kas pēta, piemēram, daiļliteratūru un publicistiku, postkoloniālo situāciju Baltijas valstīs, tūrismu un tā vēsturi, ceļojumu aprakstus u. tml. Vortālā *literatura.lv* pieejamā rakstnieku ceļojumu datubāze un administratīvajā vidē pieejamais kartēšanas rīks sniedz iespēju latviešu literatūras vēstures un padomju literārā procesa pētniecībā izmantot jaunas pieejas un metodes, aktualizējot iepriekš nepētītus jautājumus. Darbs pie kartēšanas rīka pilnveides vortālā *literatura.lv* un resursa administratīvajā vidē turpinās, lai publiski būtu pieejamas datu vizualizācijas ne tikai par atsevišķu personu ceļojumiem, bet tos varētu atlasīt arī pēc piederības konkrētai organizācijai, piemēram, Rakstnieku savienībai.

Pēc iegūto faktu un datu analīzes secināms – lai arī vairāki tūrisma vēsturnieki atzīst, ka tieši arhīvos iespējams kvalitatīvāk izsekot tūrisma institucionālajai attīstībai un tekstuālajam un galamērķu vizuālajam attēlojumam, nekā pētot atsevišķu tūristu individuālo pieredzi (Bechmann Pedersen, Noack 2019: 8), tomēr RS biedru braucienu izpēti sakarā ne vienmēr arhīva dokumentos iespējams iegūt informāciju par rakstnieku ceļošanas tendencēm, kā arī gūt apstiprinājumu ceļojumu īstenošanai, it īpaši gadījumos, kad pieejami tikai dokumenti par braucienu atļauju izsniegšanu, bet nav pieejamas komandējumu atskaites, kas ļautu gūt pilnīgāku priekšstatu par braucienu norisi¹⁰. Tādēļ tēmas atklāsmē būtiska ir atsevišķu Rakstnieku savienības biedru individuālo gadījumu izpēte, padziļinātu uzmanību pievēršot šo rakstnieku personības, daiļrades, publicistikas, memuārliteratūras, kā arī periodikas un citu arhīvu fondu izpētei.

10 Latvijas Padomju rakstnieku savienības sekretariāta un komunistiskās partijas pirmorganizācijas sēžu dokumentos no 1970. līdz 1989. gadam LNA LVA krājumā nav nevienas rakstiskas RS biedra ārzemju komandējuma vai tūrisma brauciena atskaites.

- Anderson, Carrie, Ceserani, Giovanna, Donaldson, Christopher, Gregory, Ian N., Hall, Melanie, Rosenbaum, Adam T., Taylor, Joanna E. (2017). Digital Humanities and Tourism History. *Journal of Tourism History*, No. 9(2–3), pp. 246–269. DOI: 10.1080/1755182X.2017.1419455
- Annus, Epp (2018). *Soviet Postcolonial Studies. A View from the Western Borderlands*. London, New York: Routledge, Taylor & Francis Group.
- Auziņš, Imants, Avotiņa, Daina, Kindzulis, Česlavs (sast.) (1971). *Latvijas Padomju rakstnieku savienība pīcos gados (1966–1970)*. Izdevums rakstnieku VI kongresam. Rīga.
- Bechmann Pedersen, Sune, Noack, Christian (2019). Crossing the Iron Curtain. An Introduction. Bechmann Pedersen, Sune, Noack, Christian (eds.). *Tourism and Travel During the Cold War: Negotiating Tourist Experiences Across the Iron Curtain*. 1st ed. London: Routledge, pp. 1–20.
- Belgum, Kirsten, Handley, Keith, Bott, Rachel (2018). Mapping Travel Writing: a Digital Humanities Project to Visualise Change in Nineteenth-Century Published Travel Texts. *Studies in Travel Writing*, No. 22(3), pp. 306–324. DOI: 10.1080/13645145.2019.1575765
- Burima, Maija (2018). Orientalism, otherness, and the Soviet empire: travelogues by Latvian writers of the Soviet period. Annus, E. (ed.). *Coloniality, Nationality, Modernity. A Postcolonial View on Baltic Cultures Under Soviet Rule*. New York: Routledge, pp. 65–75.
- Cooper, David, Donaldson, Christopher, Murrieta-Flores, Patricia (2016). Introduction: Rethinking Literary Mapping. Cooper, David, Donaldson, Christopher, Murrieta-Flores, Patricia (eds.). *Literary Mapping in the Digital Age*. London, New York: Routledge, pp. 1–22.
- Daija, Pauls, Kalnačs, Benedikts (2019a). Exploring Space and Place in Literature: Towards Digital Analysis and Literary Mapping in the Context of Latvian Cultural History. *Colloquia*, No. 42, pp. 67–83.
- Daija, Pauls, Kalnačs, Benedikts (2019b). The Geographical Imagination in Early Twentieth-Century Latvian Novels. *Primerjalna Književnost*, No. 42(2), pp. 153–171.
- Daukste-Silasproģe, Inguna (2020). Rīga Andreja Johansona dzīves un daiļrades kartē. *Letonica*, Nr. 42, 50.–65. lpp. DOI: 10.35539/LTNC.2021.0042.I.D.S.0004
- Eglāja-Kristsone, Eva (2012). Robežas simbolizācija latviešu rakstnieku ceļojuma aprakstos par tūrisma braucieniem ārpus PSRS. *Letonica*, Nr. 24, 9.–21. lpp.
- Eglāja-Kristsone, Eva (2016). *Dzelzsgriežņi: Latvijas un Rietumu trimdas rakstnieku kontakti*. Rīga: Latvijas Universitātes Literatūras, folkloras un mākslas institūts.
- Eglāja-Kristsone, Eva (2021). The Digital Resource Literatura.lv: An Introduction. *Colloquia*, No. 47, pp. 149–159.
- Ērgle, Zenta (1956). Polija šodien. *Zvaigzne*, Nr. 17, 22.–24. lpp., Nr. 18, 12.–14. lpp.
- Ērgle, Zenta (1975). *No Plutona līdz Neptūnam. Ceļojuma piezīmes*. Rīga: Liesma.
- Ērgle, Zenta (1979). *Gads ar diviem pavasariem. Ceļojuma piezīmes*. Rīga: Liesma.
- Ērgle, Zenta (1981). Ceļojošais cirks “Matilde” (nobeigums). *Zīlīte*, Nr. 7, 18.–19. lpp.
- Eversone, Madara (2021). Komunistiskās partijas kontroles mehānisms Latvijas Padomju rakstnieku savienībā: Zaņa Grīvas piemērs. *Aktuālas problēmas literatūrā un kultūras pētniecībā*, 26. Liepāja: LiePA, 110.–118. lpp. DOI: 10.37384/APLKP.2021.26-1.110
- Foka Anna, Cocq Coppélie, Buckland Phillip I., Gelfgren Stefan (2020). Mapping Socio-Ecological Landscapes: Geovisualization as method. Schuster, Kristen, Dunn, Stuart. *Routledge International Handbook of Research Methods in Digital Humanities*. 1st ed. London: Routledge, pp. 203–217.
- Gorsuch, Anne E. (2011). *All This is Your World. Soviet Tourism at Home and Abroad After Stalin*. New York: Oxford University Press.
- Kārkla, Zita (2021). Ievads, komentāri. *Par piemiņu. Minnas Freimanes ceļa apraksti*. Rīga: LU LFMI.
- LNA LVA 473. f., 1. apr., 504. l. Latvijas Padomju rakstnieku savienības valdes sekretariāta sēžu protokoli 1978. g.
- LNA LVA 473. f., 1. apr., 341. l. Latvijas Padomju rakstnieku savienības valdes sekretariāta sēžu protokoli (08.01.1975.–25.12.1975.)
- LNA LVA PA-7263. f., 1. apr., 44. l. Latvijas Padomju Rakstnieku savienības komunistiskās pirmorganizācijas Rīgas pilsētas Kirova rajona partijas pirmorganizācijas sapulču

- protokoli Nr. 3–12 (31.01.1986.–24.12.1986.)
- LNA LVA PA-7263. f., 1. apr., 36. l. Latvijas Padomju rakstnieku savienības komunistiskās partijas pirmorganizācijas biroja sēžu protokoli (04.02.1981.–09.12.1981.)
- LNA LVA PA-7263. f., 1. apr., 39. l. Latvijas Padomju rakstnieku savienības komunistiskās partijas pirmorganizācijas biroja sēžu protokoli (26.01.1983.–28.12.1983.)
- Ločmelis, Alberts (1980). Atkal pie darba. *Karogs*, Nr. 12, 168.–169. lpp.
- Oga, Jānis (2021). Ārzemju ceļojumu iespāidi latviešu rakstnieku darbos t. s. “Brežņeva laikmetā”: liecības par atļauto un liegto. *Aktuālas problēmas literatūrās un kultūras pētniecībā*, 26(1). Liepāja: LiePA, 119.–136. lpp. DOI:10.37384/APLPK.2021.26-1.119
- Orlov, Igor (2014). *The Soviet Union Outgoing Tourism in 1955–1985: Volume, Geography, Organizational Forms*. Higher School of Economics Research Paper No. WP BRP 50/HUM/2014. Available: <https://www.hse.ru/data/2014/04/29/1322754646/50HUM2014.pdf> [accessed 13.11.2021.].
- Orlov, Igor, Popov, Aleksej (2016). *Skvoz’ “zheleznyj zhanaves”. Russo turista: sovetskij vyezdnij turizm. 1955–1991*. Moskva: Izdatel’skij dom Vysshej shkoly jekonomiki. Kindle versija.
- Piatti, Barbara (2016). Mapping Fiction. The Theories, Tools and Potentials of Literary Cartography. Cooper, David, Donaldson, Christopher, Murrieta-Flores, Patricia (eds.). *Literary Mapping in the Digital Age*. London, New York: Routledge, pp. 88–101.
- Skujiņš, Zigmunds (2006). *Jūtu ceļojumi. Dažādu dimensiju apsēstības. Rakstu 3. sējums*. Rīga: Mansards.
- Yurchak, Aleksej (2014). *Eto bylo navsegda, poka ne konchilos’. Poslednee sovetskoe pokolenie*. Moskva: Novoe literaturnoe obozrenie.
- Zālīte, Indulis (1998). Radošā inteliģence – VDK kontroles objekts un propagandas ierocis (70., 80. g. atspoguļojums VDK dokumentos). *Latvijas Vēsture*, Nr. 3, 55.–61. lpp.

Travel Mapping and Sociopolitical Tendencies: Travels of the Members of the Latvian Soviet Writers' Union in the 1970s and 1980s

Madara Eversone, Jānis Oga

Keywords: digital mapping, Latvian Soviet Writers' Union, writers of the Latvian SSR, Soviet tourism

In the period of Soviet occupation, travel opportunities were limited, thus radically affecting citizens' travel habits and destinations. However, privileged groups of society, such as *intelligentsia*, including members of the Latvian Soviet Writers' Union, enjoyed material privileges and regular trips for trade or cultural missions, tourist and health improvement to other Soviet republics, satellite states, Western states, and "exotic" places with Soviet significance.

The focus of this research is aimed at exploring and analysing social and political aspects, as well as travel tendencies by members of the Latvian Soviet Writers' Union outside the Soviet Union in the 1970s and 1980s. Within the study, a digital data collection of the travels has been set up to cover the geographical area and examine the intensity of trips, as well as to identify the members who were frequent travellers, the most popular destinations (systematizing countries by ideologically political position), and the purposes of trips. Geographical mapping methods have been used to visualize the data. The database and maps are available in the digital resource *literatura.lv* of the Institute of Literature, Folklore and Art of the University of Latvia.

The case study method was also applied, namely, by exploring Zenta Ērgle's travelogues *From Pluto to Neptune* (1975) about a tourist trip to African countries and *A Year with Two Springtimes* (1979) about a trip to New Guinea, Australia, and New Zealand.

DH methods, such as digital mapping and data visualization, have been used in the study of Latvian literature to achieve a systematized and in-depth analysis of travel routes and intensity, providing an opportunity to investigate the data on the socio-political situation affecting the trips, their purposes, as well as the persons and organizations involved.

Latento Dirihlē sadalījumu modeļa izmantojums laikraksta *Latvijas Kareivis* tematu analīzē: Oskara Kalpaka gadījuma izpēte

Anda Baklāne, Valdis Saulespurēns

Publikācija tapusi Valsts pētījuma programmas "Humanitāro zinātņu digitālie resursi" projektā "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" (Nr. VPP-IZM-DH-2020/1-0001).

Ievads

Jau kopš 1999. gada Latvijas Nacionālā bibliotēka (LNB) veic vēsturisko laikrakstu, grāmatu, attēlu, audio un video kolekciju digitalizāciju (Krūmiņa 2012; Zariņš 2014). Teksta kolekcijām tikusi pievērsta vislielākā vērība; tiek lēsts, ka digitalizēto laikrakstu kolekcijas ietver vairāk nekā 80 % periodikas materiālu, kas publicēti līdz 20. gadsimta 90. gadu vidum¹. Materiāli tikuši segmentēti un optiski atpazīti, tādējādi to lietotāji var izmantot iespējas, kuras sniedz iespēja meklēt vārdus pilnajā tekstā. Tomēr, sekojot valodas tehnoloģiju attīstībai un pašreizējām tendencēm digitālo humanitāro zinātņu izpētē, pastāv pieprasījums pēc jaunu pakalpojumu izveides, kas sniegtu vēl vairāk iespēju padziļinātai teksta dokumentu izpētei (Ehrmann et al. 2020; Ūdre et al. 2019).

Šajā rakstā gadījuma izpēte veikta ar mērķi pētīt un pārbaudīt, kā automatiskas tematu noteikšanas metodes varētu tikt izmantotas LNB digitalizēto periodisko izdevumu kolekciju pārlūkošanā un izpētē. Tematu noteikšana (*topic detection*) jeb tematu modelēšana (*topic modelling*) bijusi viena no biežāk izmantotajām teksta digitālas analīzes tehnikām sociālajās un humanitārajās zinātnēs 21. gadsimta pirmajās desmitgadēs, savukārt latentu Dirihlē sadalījumu (*latent Dirichlet allocation*, LDA) metodoloģija – viena no biežāk izmantotajām tematu noteikšanas metodēm; tātad LDA jau ir labi izprasta un pārbaudīta dažādos lietojumgadījumos.

LDA izmantojums vēsturisko laikrakstu kolekciju pētniecībai latviešu valodā tomēr uzskatāms par novatorisku. Dabīgās valodas apstrādes tehnoloģijas visstraujāk attīstījušās lielajās valodās, savukārt mazās valodās ar sliktu valodas tehnoloģiju un resursu nodrošinājumu jauninājumi tiek ieviesti novēloti (Abney 2010). Saskaņā ar META-NET aplēsi 2012. gadā latviešu valodas atbalsts teksta un valodas apstrādes uzdevumu veikšanai tika vērtēts kā nepilnīgs vai neesošs (Skadiņa et al. 2012). Laikā, kad top šī publikācija, ir pieejami daudzi no dabīgās valodas apstrādes resursiem un rīkiem, kas 2012. gadā vēl nebija izstrādāti, to vidū apjomīgi anotēti korpusi², latviešu valodas teksta automatiskas morfoloģiskas un sintaktiskas marķēšanas rīki³, tiek veidota latviešu valodas leksēmu semantisko

- 1 Latvijas Nacionālās digitālās bibliotēkas Periodikas portāls. Pieejams: <http://www.periodika.lv/> [skatīts 18.06.2022.].
- 2 Latviešu valodas teksta un runas korpusi. Pieejams: korpusl.lv [skatīts 18.06.2022.].
- 3 NLP-PIPE: Latvian NLP Pipeline as a Service. Pieejams: <http://nlp.aialab.lv/> [skatīts 18.06.2022.].

attiecību datubāze *WordNet*⁴. Tomēr, attīstoties jaunām tehnoloģijām, piemēram, vārdlietojumu kartēšanai (*word embeddings*) un īpaši lielu valodas modeļu veidošanai, valodas, kurām vēsturiski bijis zems tehnoloģiju un valodas resursu nodrošinājums, joprojām atpaliek (Alabi et al. 2020). 24 oficiālo Eiropas Savienības valodu vidū 15 valodas (t. sk. latviešu valoda) var tikt uzskatītas par valodām ar nepietiekamu resursu nodrošinājumu (Alves et al. 2020). Šajā publikācijā aplūkots pētījums ir pirmais LDA metodoloģijas izmantojums latviešu vēsturisko laikrakstu analizē⁵ un kopumā viens no pirmajiem LDA izmantojumiem, analizējot tekstu latviešu valodā⁶.

Dabīgās valodas apstrādes tehnoloģiju ieviešana vēsturisko dokumentu kolekcijās saistīta ar dažādiem šķēršļiem, ar kuriem nenākas saskarties, analizējot digitāli radītas, liela apjoma, stilistiski viendabīgas un standartizētas teksta datu kopas. Digitalizētu senu tekstu kopas nereti ietver relatīvi mazus korpusus, kas nav viendabīgi un līdzsvaroti, tādējādi ierobežojot mašīnmācīšanās metožu izmantošanas iespējas (McGillivray 2021). Laika posms no 19. gs. otrās puses līdz 20. gs. sākumam jau piedāvā salīdzinoši bagātīgu teksta dokumentu klāstu, taču digitālo analīzi apgrūtina dažādi vārdu pieraksta varianti, ar ortogrāfijas reformām saistītās rakstības izmaiņas un optiskās atpazīšanas kļūdas (Bollmann 2019). Oskara Kalpaka gadījuma izpēte apliecina, ka šī problemātika aktuāla, arī analizējot digitalizēto laikrakstu kolekcijas latviešu valodā.

Rakstā sadaļā “LDA metodoloģija humanitāro zinātņu pētniecībā” skaidrots LDA jēdziens, aplūkotas atšķirīgas pieejas LDA rezultātu interpretācijā un iezīmēti vairāki scenāriji, kā tematu modeļus varētu integrēt digitalizēto laikrakstu kolekciju saskarnēs. Sadaļā “Datu kopa” atrodama informācija par izmēģinājumā izmantotās datu kopas parametriem un datu atlasē pamatojums. Sadaļā “Gadījuma izpētes rezultāti” izklāstīti un interpretēti datu modeļa apmācības rezultāti. Secinājumu daļā apkopotas pētījuma gaitā gūtās atziņas un norādīti iespējamie turpmākie soļi.

LDA metodoloģija humanitāro zinātņu pētniecībā

Tematu modelēšanas metodoloģijas ietver dažādas dabīgās valodas apstrādes un mašīnmācīšanās tehnikas, kas tiek izmantotas, lai atrastu liela apjoma dokumentu kolekcijās ietvertas saturs struktūras jeb tematus. Tematu modelis ir matemātiskas varbūtības sadalījums no kādas dokumentu kolekcijas izgūtā vārdu kopā, kas paredz, kādiem tematiem pieder šie vārdi un kuri temati ir klātesoši katrā dokumentā.

4 Latvian WordNet. Pieejams: <https://wordnet.ailab.lv/> [skatīts 18.06.2022.].

5 Provizoriskie Oskara Kalpaka gadījuma izpētes rezultāti aplūkoti rakstā: Baklāne, Saulespurēns 2022.

6 R. Viksna, M. Kirikova un D. Kiopa veikuši Latvijas tiesību aktu tematu analīzi (kopumā vairāk nekā 50 000 dokumentu). Pētījumā salīdzinātas trīs tematu analīzes metodes, t.sk. LDA (Viksna et al. 2020).

Tematu modeli lietotājs parasti var aplūkot kā vārdu sarakstus, kuros katram vārdam piešķirta atbilstošā varbūtības vērtība. Tiek sagaidīts, ka šiem vārdu sarakstiem jābūt semantiski saskanīgiem, taču ne visi tekstu žanri un ne visas dokumentu kolekcijas var kalpot par pamatu tematiski vienotu vārdu sarakstu radīšanai. Piemēram, mākslinieciski teksti vai nevienmērīgas kolekcijas, kurās sajaukti dažāda garuma un dažādu žanru teksti, var nenodrošināt pamatu saskanīgu un viegli interpretējamu atslēgvārdu sarakstu veidošanai. Tomēr arī neviennozīmīgi interpretējami vārdu saraksti var tikt izmantoti korpusu izpētes vajadzībām, t. i., tematu modelēšanas algoritmu darba rezultāti var tikt izmantoti, ne vien lai lakoniski un nepārprotami atspoguļotu saturu, bet arī plašāk – lai vizualizētu, pētītu, izvirzītu hipotēzes par korpusu (Blei 2012).

Latento Dirihlē sadalījumu metode tematu atrašanai pirmo reizi aplūkota publikācijā 2003. gadā (Blei et al. 2003) un pašlaik tiek uzskatīta par vienu no biežāk izmantotajām tematu modelēšanas tehnikām (Marjanen et al. 2020; Pääkkönen, Ylikoski 2020). LDA ir varbūtības modelis, kas balstīts divos pieņēmumos: (1) pastāv noteikts skaits dotajos dokumentos bieži kopā lietotu vārdu kopumu (tematu); (2) katrs dokuments korpusā var saturēt vairākus tematus dažādās nozīmīguma pakāpēs (Blei 2012). Formāli LDA var definēt kā varbūtības $P(\theta:1:M, z:1:M, \beta:1:k \mid D; \alpha:1:M, \eta:1:k)$ atrašanu. Tas nozīmē kopīgas varbūtības atrašanu M dokumentiem ar šādiem nezināmajiem: θ – tematu sadalījums pa vienam uz katru dokumentu; z – temati katrā dokumentā; β – vārdu sadalījums katrā tematā, k – kopējais tematu skaits visos dokumentos. Kā kopēja dotā varbūtība ir dots korpus D un parametri: α – parametru vektors katram dokumentam, η – parametru vektors katram tematam. Šādu varbūtību nevar atrisināt ar standarta aritmētiskām metodēm (Blei et al. 2003). LDA atrisina šo varbūtību ar iteratīvu algoritmu, kas turpina modeļa apmācību, līdz tiek sasniegts zināms konverģences līmenis.

Varbūtiskās tematu modelēšanas lietojums var atšķirties atkarībā no pētniecības jomas un pētnieka mērķiem. Kopš 2003. gada tikuši veikti daudzi pētījumi, kuros LDA metodoloģija un tās atvasinājumi izmantoti, lai veidotu tematu modeļus zinātnisku publikāciju kolekcijām (Blei, Lafferty 2007; Newman et al. 2006; Hall et al. 2008), kā arī lai pētītu vēsturiskos laikrakstus un žurnālus (Block 2006; Nelson 2011; Templeton et al. 2011; Hengchen 2017). Spriežot pēc piemēriem, kurus piedāvā varbūtisko tematu modelēšanas metožu izstrādātāji, metode primāri tikusi veidota lietišķu teksta žanru, nevis daļlīdētāru izpētei, taču laika gaitā tikusi analizēta arī mākslinieciskā proza un dzeja (Rhody 2012).

Koherentu, uzticamu tematu izveide, lai tos varētu izmantot lielu akadēmisku publikāciju repozitoriju vai ziņu avotu digitālo kolekciju izpētē, ir viens no virzieniem, kas mērķtiecīgi tiek attīstīts tematu modelēšanas jomā (Chang 2009). Akadēmisko repozitoriju kontekstā neiederīgo vārdu (*intrusion words*) klātbūtne un tematu sajaukšanās (*mixing of topics*) ir nevēlama, jo mazina lietotāju pašāvēību, ka modelis darbojas pareizi. Turpretī, pētot literāru darbu datus, temati nereti nav viegli interpretējami, semantiski saistītu atslēgvārdu saraksti, tomēr tie netiek noraidīti, jo var sniegt cita veida informāciju par darbu leksiku un stilistiku. Literatūras pētnieku vidū sastopams viedoklis, ka literāru tekstu izpētē ambivalenti temati var sniegt pat vairāk informācijas nekā viennozīmīgie tematu saraksti, kuriem priekšroku dotu vēsturnieki:

gadījumos, kad vārdu saraksti neveido tematiskas satura vienības, kas apraksta konkrētu referentu, tie var reprezentēt diskursu, sociolektu vai noteiktu poētiskās retorikas veidu (Underwood 2012). Citiem vārdiem sakot, tematu modelis var tikt izmantots, ne vien lai noskaidrotu, *par ko* cilvēki raksta, bet arī to – *kā* viņi raksta (Goldstone 2012).

Pieejā, kurā uz tematu saprotamību un semantisko saskanīgumu tiek likts mazāks uzsvars, pazīstama ne vien daļēji literatūras, bet arī vēstures avotu pētniecībā. Daudzos gadījumos modelēšanas rezultāti ir neviennozīmīgi un grūti interpretējami, un daži pētnieki uzsver, ka, lai arī tematu modelēšanas rezultāti ne vienmēr izmantojami kā lietošanai gatavi pierādījumi par kāda tekstu kopuma tematisko aptvērumu, tie tomēr ir lietderīgs palīgīdzeklis teksta izpētes procesā (Brett 2012). Tematu modelis var vērst uzmanību uz tēmām, kas varējušas palikt nepamanītas kvalitatīvas analīzes procesā, jo nav iespējams izlasīt visus tekstus vai izlasīt tos vienlīdz uzmanīgi (Kurvinen 2020). Tādējādi, īpaši vēsturisko un literāro avotu analīzē, tematu modeļi var papildināt hermeneitikas un tuvaslājsjuma pētniecības metodes, savukārt, veicot digitālu kvantitatīvu pētījumu, būtiska var būt iespēja atgriezties pie pirmavota, lai novērtētu modeļa atbilstību vai labāk izprastu, kā interpretēt rezultātus (Rhody 2012; Kurvinen 2020; Viola, Verheul 2019). Minētās pieejas tematu analīzei var klasificēt kā tematu realīsmu (*topic realism*) un tematu instrumentālismu (*topic instrumentalism*). Tematu realisms ir skatījums, kurā tiek atzīts, ka modelēšanas process var tvert reprezentācijas vai teorētiskus konstruktus (rāmējumus, diskursus, naratīvus), kas reāli eksistē tekstos. Savukārt tematu instrumentālisms ir skatījums, kurā tiek atzīts, ka modeļi vienkārši sniedz informāciju par sakarībām, kas var būt noderīgas tekstu interpretācijā, neizvairot prasību nonākt pie precīza un patiesā teksta satura realitātes atspoguļotāja modeļa (Pääkkönen, Ylikoski 2020).

Tātad ne visu veidu teksti vienlīdz labi pakļaujas mēģinājumiem formalizēt to saturu, lakoniski iekodēt jēgu dažos nozīmīgos vārdos. Tomēr neapšaubāmi tematu kvalitāte ir atkarīga arī no pārdomātas datu atlasē, no teksta priekšapstrādes kvalitātes un apmācības parametriem, kas uzstādīti modelim (Wallach et al. 2009). Proti, ne katrs modelis ir uzskatāms par metodoloģiski korekti izveidotu pat tādā gadījumā, ja sniedz iedvesmojošas idejas pētniekam.

Šajā rakstā skatītā piemēra izstrādes procesā tika apsvērti vairāki scenāriji, kā LDA metodoloģija var tikt izmantota vēsturisko laikrakstu digitālās bibliotēkas izpētē. Pirmkārt, tematu modeļa apmācībai un pielāgošanai varētu tikt izmantots viss LNB digitalizētās periodikas krājums, izveidojot tematiskus blokus, kas papildinātu periodikas portāla pārlūkošanas funkcionalitāti un piedāvātu ieteikumus lietotājiem. Otrkārt, varētu tikt apmācīti un pielāgoti vairāki modeļi dažādiem LNB periodisko izdevumu segmentiem – atsevišķiem izdevumiem, izdevumu veidiem vai laika periodiem, integrējot šo informāciju kā periodikas portāla papildu funkcionalitāti. Treškārt, papildus tradicionālajiem digitālās bibliotēkas pakalpojumiem varētu tikt izveidota atsevišķa saskarne, kas ļautu lietotājam pašam atlasīt datus korpusa izveidei un veidot tematu modeli šim korpusam. Šāda pieeja sniegtu iespēju analizēt arī tikai, piemēram, kādai konkrētai tēmai vai personai veltītus materiālus. Oskara Kalpaka gadījuma izpētē lielā mērā izmantoti trešā scenārija elementi, taču piemērs sniedz noderīgas atziņas arī pārējo scenāriju īstenošanai.

Tematiska korpusa izveidei ir gan priekšrocības, gan trūkumi. No vienas puses, iepriekšēja atlase ļauj daudz detalizētāk izpētīt konkrētu interesējošo tematu. Tā, piemēram, izveidojot tematu modeli visam *Latvijas Kareivja* korpusam, tikai viens no 50 tematiem saturēja vārdu “kalpaks”, savukārt Oskara Kalpaka apakškorpusā modelis ar augstāko koherences rādītāju ietver sešus tematus. No otras puses, jāpatur prātā, ka tematiska apakškorpusa izmantojums ir ierobežots, ļauj secināt tikai sākotnēji izvēlēta temata ietvaros. Izpēte, kas veikta, analizējot visu laikraksta numuru komplektu vai varbūtīgu izlasi, uzskatāma par lielākā mērā datu virzītu (*data driven*), savukārt iepriekšēja tematu atlase var palielināt pētnieka subjektivitātes (t. sk. iepriekš pieņemtu spriedumu un aizspriedumu) ietekmi modeļu veidošanā.

Oskara Kalpaka gadījuma izpētē netika ņemta vērā tematu modelēšanas laika dimensija. LDA pamata metode neņem vērā laika aspektu, t. i., laiks nav iekļauts modelī kā mainīgais, taču humanitāro zinātņu pētniecībā bieži ir nepieciešamība pētīt arhīvus un bibliotēkas, kuru materiāli publicēti vairāku gadu desmitu vai pat simtu gaitā (Marjanen 2020). Lai mazinātu šī ierobežojošā faktora ietekmi, datu kopa var tikt sadalīta secīgos segmentos, apmācot atsevišķu modeli katram segmentam. Papildinot LDA, ir tikušas izstrādātas arī citas tehnikas, kas ņem vērā laika dimensiju, piemēram, dinamisko tematu modeļu metode (*dynamic topic models*) (Blei, Lafferty 2006).

Datu kopa

Gadījuma izpētei izmantots laikraksta *Latvijas Kareivis* korpuss un no tā atvasināts Oskara Kalpaka apakškorpus. *Latvijas Kareivis* ir oficiālais Latvijas Bruņoto spēku štāba dienas laikraksts, kas tika izdots no 1920. līdz 1940. gadam (Pētersone 1999). Līdz 1925. gadam tas tika iespiests vecajā drukā, vēlāk pāriets uz jauno ortogrāfiju. Ortogrāfijas nekonsekvenču, pakāpeniskas izmaiņas un rakstības reformas ir grūtības, ar kurām bieži jāsasopas vēsturisko laikrakstu pētniekiem. Lai neradītu papildu nenoteiktību izmēģinājumu korpusā, pētījumā tika izmantota tikai *Latvijas Kareivja* modernajā ortogrāfijā iespiestā daļa. No kopējās datu kopas tika atlasīts apakškorpus – raksti, kas satur vārdu “kalpaks”.

Pulkvedis Oskars Kalpaks (1882–1919) bija Latvijas Pagaidu valdības bruņoto spēku komandieris, Landesvēra latviešu vienību un Pirmā atsevišķā latviešu bataljona komandieris (Jēkabsons 2022), kurš tiek uzskatīts par vienu no Latvijas armijas pamatlicējiem. Veicot gadījuma izpēti, hipotētiski tika pieņemts, ka ar Kalpaka vārdu varētu būt saistīti vairāki temati un tiem varētu būt mainīga aktualitāte analizētajā laika periodā. Atslēgais vārds “kalpaks” ļauj pārbaudīt arī vārdu daudznozīmības problēmas ietekmi: korpusā pieminēts Kalpaka bulvāris, Kalpaka iela, Kalpaka tilts, kas vairumā kontekstu nav tieši saistīti ar Oskara Kalpaka tematiku. Tajā pašā laikā uzvārds “Kalpaks” ir salīdzinoši rets, tādējādi nerada papildu sarežģījumus izmēģinājuma modeļa veidošanas procesā. Darba gaitā, pirmkārt, tika izveidots tematu modelis visam *Latvijas Kareivja* korpusam, otrkārt, atsevišķs tematu modelis Oskara Kalpaka apakškorpusam – pēdējais tika analizēts sīkāk.

Latvijas Kareivja korpuss satur 55,9 milj. vārdformu; Oskara Kalpaka apakškorpuss satur 1,3 milj. vārdformu. Korpusa dati ir segmentēti rakstu līmenī, tomēr rubrikas, kas satur īsas ziņas un paziņojumus, segmentēšanas procesā tikušas konsolidētas, veidojot sadaļas, kas ietver vairākas tēmas.

Teksta segmentēšanas īpatnības vai kļūdas teorētiski var kļūt par šķērslī koherenta tematu modeļa izveidei, taču, šķiet, šajā gadījumā īso ziņu savienošana nav radījusi nevēlamu ietekmi. Konsolidētās rubrikas satur daudz vienādu vārdu, saīsinājumu un skaitļu, tādējādi algoritms šos tekstus grupēja vienkopus kā atsevišķu tematu – šādi automātiski tika nošķirts materiāls, kas nesatur izvērstus pārspriedumus par tēmām, kas saistītas ar Oskara Kalpaka dzīves gaitu vai piemiņas pasākumiem. Tas vedina domāt, ka tematu modelis var būt noderīgs instrumentis arī pētāmā materiāla atlases un filtrēšanas procesā – lai atbrīvotos no liekajiem datiem. Vienlaikus bija novērojams, ka dažkārt atsevišķi garāki raksti savienoti kopā kļūdas dēļ: šiem tekstiem raksturīgi augsti vairāku tematu procentuālie rādītāji (tā vietā, lai izteikti dominētu viens temats).

Korpuss tika lemmatizēts, izmantojot dabīgās valodas apstrādes rīku ķēdi NLP-PIPE (Znotiņš, Cīrulle 2018). Sākotnējā korpusa vārdnīca tika samazināta, atmetot vārdlietojumus, kas satur tikai vienu simbolu, – šis solis ļāva samazināt arī optiskās atpazīšanas kļūdu ietekmi. Var argumentēt, ka vārdlietojumi, kas satur tikai divus simbolus, arī var tikt atmetti, jo to vidū nav semantiski nozīmīgu vārdu, tomēr Kalpaka gadījuma izpēte parādīja, ka vismaz atsevišķos gadījumos divciparu skaitļiem bija sava loma jēgpilnas dokumentu grupēšanas procesā.

Gadījuma izpētes rezultāti

Gadījuma izpēte tikai veikta, izmantojot *Python* atvērtā koda bibliotēku *Gensim*⁷. Uzstādot modeļa apmācības parametrus, izmantoti sistēmas noklusējuma parametri un ņemti vērā *Gensim* izstrādātāju ieteikumi (Řehůřek, Sojka 2010). Tematu veidošanai izmantotais vārdu krājums (vārdnīca) tikai veidots, ņemot vērā atsevišķus vārdus, bigramas un trigramas; izmantots *bag-of-words* modelis (t. i., nav ņemta vērā vārdu secība); izveidotā vārdnīca sastāv no 5030 tekstvienībām. Sekojot *Gensim* izstrādātāju ieteikumiem, vārdnīcā tika iekļautas tekstvienības, kas lietotas vismaz 20 reižu, un netika iekļautas tekstvienības, kas sastopamas vairāk nekā 50 % dokumentu (attiecīgi vārdnīcā iekļūst maz bieži lietoto vārdu bez patstāvīgas nozīmes). Katrs LDA modelis tika apmācīts 400 iterāciju (*iterations*), 20 epohu (*epochs*) gaitā. Optimāla tematu skaita kalkulācijas balstītas (CV) koherences mērījumā.

Koherences mērījumi (*coherence measurements*) tiek izmantoti, lai novērtētu izvēlēto tematu modeļu precizitāti – piemēram, lai noteiktu, kāds tematu skaits jāizvēlas, lai

7 *Gensim: Topic Modelling for Humans*. Pieejams: <https://radimrehurek.com/gensim/>

veidotos koherents modelis. Šie mērījumi raksturo modeļa izvēlēto tematu saturošo vienību saskaņotību jeb vienota veseluma veidošanu starp šīm vienībām (t. i., cik lielā mērā kādā tematā ietilpstošie vārdi ir savstarpēji saistīti). Modeļu novērtēšanai tika izmantoti vairāki iepriekš zināmi un dotajā darba vidē (*Gensim* bibliotēkā) pieejami koherences mērījumi; šajā rakstā izmantoti (CV) koherences rādītāji. (CV) mērījums ir balstīts slīdošā loga (*moving window*) principā – apstrādājot kādu tematu pa segmentiem un virzoties uz priekšu pa vienam vārdam. Tiek segmentēti visbiežāk sastopamie vārdi, aprēķināta konkrētā termina atrašanās varbūtība, un izveidots mērījuma apstiprinājums; visbeidzot tiek izveidots apstrādāto segmentu (logu) rezultātu kopsavilkums. Kā parādīts pētījumos, augstākie (CV) koherences rādītāji lielā mērā sakrīt ar cilvēku subjektīvo tematu kvalitātes novērtējumu (Röder et al. 2015). Oskara Kalpaka tematu modelī augstākais (CV) rādītājs – 0,61 – tika sasniegts Oskara Kalpaka apakškorpusa modelim, kas sastāv no sešiem tematiem. Šī raksta autori subjektīvi izvērtēja četru, piecu un sešu tematu modeļu kvalitāti. 0,61 nav uzskatāms par īpaši augstu modeļa koherences rādītāju, un arī subjektīvā vērtējumā nevarēja apgalvot, ka visi apakškorpusa raksti pārliecinoši tika sagrupēti pa tematiem.

Piemēra izstrāde bija balstīta hipotētiskos pētnieciskos jautājumos: kādi temati ir saistīti ar Oskara Kalpaka vārdu laikrakstā *Latvijas Kareivis* laika posmā no 1925. līdz 1940. gadam? Kādos kontekstos tiek minēts Kalpaka vārds? Cik daudz dažādu tematu ir saistīti ar Kalpaku? Kā šo tematu popularitāte mainās laika gaitā?

Pārlūkojot *Latvijas Kareivja* korpusa 50 tematu modeli (tematu izlasi skat. 1. tabulā), varam secināt, ka tajā spilgti iezīmējas temati, kas saistīti, piemēram, ar sportu, izglītību, transporta pārvadājumiem. Lielā skaitā atrodami temati, kas satur valstu un tautību nosaukumus, turklāt nereti vairākas valstis tiek grupētas viena temata ietvaros. Šķiet, šāds tematiskais dalījums varētu būt noderīgs, lai virzītu lasītāju pie viņam aktuāliem rakstiem, tomēr būtu nepieciešams padziļināts pētījums, lai secinātu, piemēram, cik lielā mērā tematos, kuros minētas vairākas valstis, vērojama tematu sajaukšanās, kā arī vai sastopami neiederīgie vārdi. Rūpīgāka modeļa kvalitatīva izpēte varētu sniegt atbildi uz jautājumu, vai šie tematiskie ietvari (tematu atslēgvārdu saraksti paši par sevi) var kalpot par informācijas avotu, pētot, kādā kontekstā tiek runāts par dažādām valstīm: vai būtu pamatoti sacīt, ka Igaunija un Somija galvenokārt tiek saistītas ar ciešām diplomātiskām attiecībām un vizītēm, savukārt Holande un Anglija – ar finanšu jautājumiem utt.

Vārds “Kalpaks” 50 tematu modelī parādās vienā no tematiem. Jāņem vērā, ka pēc nozīmīguma Kalpaks šeit nav viens no galvenajiem temata vārdiem – tā, piemēram, ja lietotājam, strādājot ar tekstiem digitālā kolekcijā, būtu redzami tikai pieci vai septiņi temata nozīmīgākie vārdi, Kalpaka to vidū nebūtu vispār. Pastāvošajā rāmējumā varam spriest drīzāk par to, ka tādi jēdzieni kā *Rīga*, *latvietis*, *vēsture*, *ordenis* un *novembris* saistās arī ar Kalpaka vārdu, nevis pretēji. Kāpinot tematu skaitu, Kalpaka vārds varētu parādīties vairākos tematos, taču, izvēloties lielāku tematu skaitu šim korpusam, samazinās modeļa (CV) koherences rādītājs.

2. temats	armija, aizsardzība, vienība, kars, apmācība, zirgs, militārs, valsts_ aizsardzība, laiks, organizācija, darbība, uzdevums, dienests, karaspēks, ierocis, pulkst_19, grupa, jātnieks, tikt, vadība, kā, sastāvs, daļa, sagatavošana, kauja, manevrs, viss, rezerve, sakars, kurš
3. temats	savienība, padome, tauta, kongress, Rumānija, tauta_ savienība, Turcija, turks, Austrija, ungārs, Dienvidslāvija, antante, Danciga, Bulgārija, starptautisks, Grieķija, loceklis, tēvzeme_ mīlestība, Vīne, Kronvalds, maza, delegāts, Bukaresta, grieķis, bulgārs, pārstāvis, ukraiņi, sanākt, Eiropa, ārlietas
4. temats	Latvija, Igaunija, Somija, ārlietas, Baltija, mēs, igauņi, vakar, sūtnis, valsts, pārstāvis, ministrija, ārlietas_ ministri, Tallina, Rīga, Baltija_ valsts, ierasties, prese, izbraukt, direktors, vadītājs, delegācija, iepazīties, ārlietas_ ministrija, piedalīties, konference, sūtniecība, apmeklēt, iepazīties_ ar, notikt
6. temats	tanks, varēt, mašīna, ātrums, ceļš, mm, izmantot, vai, automobilis, degviela, čemberlens, gāze, smags, šāds, līdzeklis, ierocis, viegls, lietot, auto, svars, iespēt, šis, veids, transports, nafta, aparāts, cm, kustība, katrs, dienvidaustrumi
9. temats	zvēders, Holande, Beļģija, Zviedrija, soms, Šveice, beļģis, soma, dānis, Vācija, marka, Francija, vicepriekšsēdētājs, lēdija, angļu_ frančs, birža, vāci, Anglija, Itālija, angļu_ vēstnieks, franks, Rīga_ birža, ražība, darbs_ ražība, tīrs_ peļņa, kurss, Amsterdama, padome_ sesija, viceadmirālis, beigas
12. temats	Rīga, satiksme, vilciens, stacija, līnija, pasts, pasažieris, vagonis, akc, akc_ sab, autobuss, 13_05, sab, ceļš, līdz, virsvalde, biļete, pa, telegrāfs, pienākt, prece, brauciens, braukt, starp, tarifs, jūrmala, Jelgava, aiziet, Lielupe, pārvadāt
16. temats	sports, vienība, sacīkste, sacensība, spēle, sek, Latvija, uzvarēt, notikt, pirmais, labs, cīņa, mēs, min, vieta, uzvara, Rīga, svars, laukums, punkts, bet, futbols, savienība, balva, vārti, US, gūt, valsts, LSB, sportists
17. temats	skola, pamatskola, izglītība, skolotājs, kurss, mācība, jaunatne, skolēns, ģimnāzija, audzēknis, direktors, valsts, beigt, izglītība_ ministri, institūts, ministrija, izglītība_ ministrija, lauksaimniecība, vidusskola, akadēmija, klase, jauns, audzināšana, pārbaudījums, skola_ jaunatne, skolnieks, bērns, darbs, arodskola, praktisks
32. temats	gads, dzimt, Latvija, Rīga, latvietis, vēsture, ordenis, novembris, mirt, viņš, kars, pirmais, cīņa, 1919, gaidīt_ laiks, atbrīvošana, janvāris, maijs, pēc, kā, krievs, līdz, pie, marts, laiks, oktobris, strēlnieks, februāris, Kalpaks

1. tabula. Tematu izlase no *Latvijas Karcivja* korpusa 50 tematu modeļa;
norādītas katra temata 30 raksturīgākās tekstvienības.

Oskara Kalpaka apakškorpusa sešu tematu modeļi 1., 3. un 4. temata vārdu saraksti apstiprināja gaidas, ka temati būs saistīti ar militāriem terminiem, proti, tiek pieminēta armija, pulks, cīņa, ģenerālis, pulkvedis u. c. (skat. 2. tabulu). Savukārt 2., 5. un 6. temats rāda jēdzienus, kas mazākā mērā saistās ar kara lietām: *icla*, *pilsēta*, *pulkst*, *skola* u. c. Novērtējot subjektīvi, katrs no tematiem šķiet veidojam kādu konkrētu tematisku identitāti, tomēr īpaši 2., 5. un 6. temata gadījumā nav skaidrības, kā šie jēdzieni saistīti ar pulkvedi Kalpaku, un ir nepieciešams vērsties pie rakstu pilnajiem tekstiem, lai saprastu, kā šos tematus interpretēt.

1. temats	tauta, arī, šī, savs, ciņa, varēt, armija, vēl, tad, kad, jūs, tikai, zeme, jo, jau, karavīrs, spēks, daudz, latvietis, visa, Latvija, vai, es, labs, pats, Aris ⁸ , viņa, ja, dzīve, viens
2. temats	iela, pilsēta, Aris, valde, vakar, vieta, pa, ministrija, biedrība, 10, paredzēt, ls, ministrs, jau, nodaļa, notikt, daļa, galva, pikstēt ⁹ , kāds, vēl, telpa, varēt, nolemt, vai, nams, 000, atrast, vakars, policija
3. temats	ministrs, ģenerālis, svētki, pilsēta, aizsargs, prezidents, pulks, armija, karavīrs, priekšnieks, piemiņa, kops, komandieris, pieminēklis, pulkv, pulkvedis, arī, notikt, valsts_prezidents, krist, garnizons, vieta, baznīca, biedrība, svinība, Liepāja, svinīgs, bataljons, organizācija, dievkalpojums
4. temats	rota, pulks, bataljons, kauja, armija, pulkvedis, marts, 1919, lielnieks, komandieris, janvāris, uzbrukums, pulka, Cēsis, jātnieks, virsnieks, karavīrs, vāci, tikt, daļa, kājnieks, karaspēks, ienaidnieks, ieņemt, vienība, atsevišķs, ciņa, eskadrons, jau, muiža
5. temats	pulkst, 30, pikstēt, 20, koncerts, 19, ziņa, 15, 18, 12, 10, 17, 00, šodien, skaņa, 22, 16, plate, rīts, vakars, 13, iela, mūzika, dziesma, pl, pilsēta, 21, piedalīties, opera, koris
6. temats	ls, pag, 10, skola, pamatskola, 10_ls, 50, kl, 25, 20, 100, 50_ls, pagasts, ba, valde, grāmata, darbinieks, skolotājs, 000, mazpulk, skolēns, 20_ls, 15, 100_ls, sab, 25_ls, 30, pils, pilsēta, biedrība

2. tabula. Oskara Kalpaka apakškorpusa sešu tematu modelis;
norādītas katra temata 30 raksturīgākās tekstvienības.

Subjektīvi pārskatot atbilstošos rakstus¹⁰, ir novērojams, ka 1., 3. un 4. temats dominē rakstos, kas veltīti pulkvedim Oskaram Kalpakam un ar viņa personību saistītām norisēm. Šie temati nošķir vairākus atšķirīgus kontekstus: Oskara Kalpaka atceres dienām tapuši raksti, kuros paustas pārdomas par Kalpaka nozīmi Latvijas vēsturē (1. temats); Oskara Kalpaka atceres dienu notikumu pārskatī, kuros aprakstītas dažādas svinības un ceremonijas (3. temats); raksti, kuros tiek pārspriesti 1919. gada notikumi – Latvijas Bruņoto spēku izveide un kaujas pret lieliniekiem (4. temats).

5. temats dominē paziņojumos un reklāmās par koncertiem un citiem kultūras pasākumiem. Atsevišķos gadījumos šeit var būt pieminēti Kalpakam veltīti pasākumi, taču vairumā gadījumu tiek minēts Kalpaka bulvāris un Kalpaka iela kā norises vietas. Šajā tematā sastopamie

- 8 Vārds "Aris" tematu vārdnīcā ieviesies sistematiskās optiskās atpazīšanas kļūdas dēļ. Vārds "arī" nereti atpazīts kā "ari", turklāt "Ari" bieži sastopams teikuma sākumā, rakstīts ar lielo sākumburtu. Morfoloģiskās marķēšanas rīka interpretācijā tas kļuvis par īpašvārdu – "Aris".
- 9 Vārds "pikstēt" tematu vārdnīcā ieviesies sistematiskās optiskās atpazīšanas kļūdas dēļ. Saīsinājums "plkst." neprecīzi atpazīts kā "pikst", savukārt morfoloģiskās marķēšanas rīks to pārveidojis pamatformā – "pikstēt".
- 10 Oskara Kalpaka apakškorpusa rakstu pilnie teksti pieejami šeit: <https://doi.org/10.5281/zenodo.6569249>. Raksti sagrupēti pa tematiem, ņemot vērā tematu ar visaugstāko rādītāju.

skaitļi ir norises datumi un laiki; vairākās variācijās sastopams vārds “pulksten” (skat. arī 9. beigu piezīmi). 6. temats dominē galvenokārt paziņojumos, kas saistīti ar skolām, kuras nosauktas Oskara Kalpaka vārdā. Parasti šis ziņas saistītas ar ziedojumu vākšanu un dāvinājumiem skolām. Kopš 1939. gada šis temats sastopams arī paziņojumos par ziedojumiem Latvijas aizsardzības spēkiem – skaitļi ir ziedotās summas. 2. temats satur vislielāko atšķirīgas tematikas atslēgvārdu sajaukumu. Šis temats dominē rakstos, kas dažkārt atsaucas uz Oskaru Kalpaku, bet lielā daļā gadījumu tiek pieminēts tvaikonis “Kalpaks”. Vienā no rakstiem minēta persona ar uzvārdu “Kalpaks”, kura nav Oskars Kalpaks.

1. attēlā redzamā vizualizācija apstiprina novērojumus, kas gūti tekstu un tematu subjektīvā analizē – raksti, kuros dominē 1., 3. un 4. temats, saturiski ir vairāk saistīti, un tie kā līdzīgi tuvāk sagrupēti arī daudzdimensiju kartējumā, savukārt 5. un 6. temats saturiski nav saistīti ar Oskaru Kalpaku, un arī kartējumā tie atrodas perifērijā.

LDA modelī katrs raksts var ietvert vairākus tematus dažādās nozīmīguma pakāpēs. Raksti, kuru tematiskajā kompozīcijā sajaukti vairāki temati ar lielu nozīmīguma pakāpi, biežāk izrādās vai nu Kalpaka tēmai nepiederīgi, vai tādi, kuros garāki raksti savienoti kopā segmentēšanas kļūdas dēļ. Daudzos gadījumos tomēr viena temata īpatsvars ir izteikti dominējošs – virs 80 %. Tā, piemēram, Kalpaka piemiņai veltīta publikācija ar 83 % 1. temata īpatsvaru vēsta:

Svinot ik gadus Kalpaka bataljona gada svētkus mēs pieminam to laikmetu, kurš mūsu armijas un līdz ar to mūsu tautas vēsture ierakstīts neizdzēšamiem nacionālo varoņu asinīm slacītiem burtiem. Katras tautas vēsture ir šāds laikmets, atmiņa par kuru iet no paaudzes uz paaudzi, paužot par to spēku, kurš atsvabināja tautu no svešas varas važām¹¹.

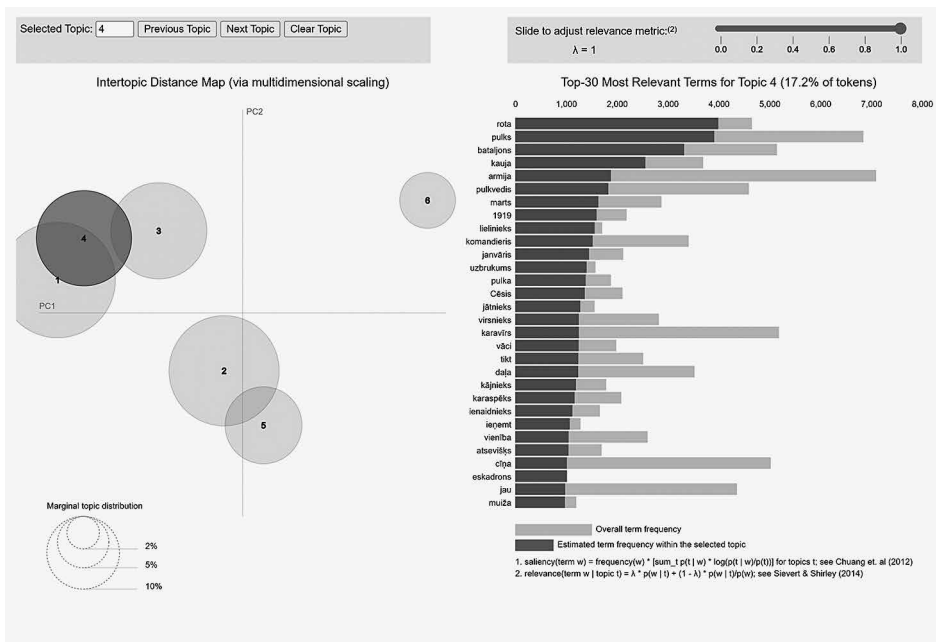
Raksti ar lielu 1., 3. vai 4. temata īpatsvaru drīzāk izvērsti vēsta par jautājumiem, kas saistīti ar Oskaru Kalpaku vai viņam veltītiem pasākumiem, tomēr tā nav vienmēr. Tā, piemēram, kāda publikācija ar 85 % 1. temata īpatsvaru veltīta Zigfrīda Meierovica (1887–1925) atcelei, un Kalpaks tajā ir tikai pieminēts (tajā pašā laikā tematiski šis raksts tuvs patriotiskajiem Kalpakam veltītajiem rakstiem):

Īstā brīdī laime mums sūtīja Kalpaku kara frontē un Meierovicu vēl grūtāka ārējās politikas frontē. Tās abus ņēma tad, kad viņu darbs vēl ārkārtīgi vajadzīgs. Mums atliek tikai ticēt Latvijas zvaigznēm kā to darīja šie divi Latvijas krietnākie dēli¹².

Šie novērojumi vedina domāt, ka, veidojot kādai tēmai veltītu korpusu, varētu būt lietderīgi atlasīt rakstus, kuros interesējošais vārds minēts vairāk nekā vienu reizi – lai atspoguļotu maznozīmīgākus pieminējumus.

11 E.G. Kalpaku pieminot. *Latvijas Kareivis*, 03.03.1925., 1. lpp. Rakstus, kuros dominē pirmais temats, datu kopā sk. pie “Topic 0”: <https://doi.org/10.5281/zenodo.6569249>

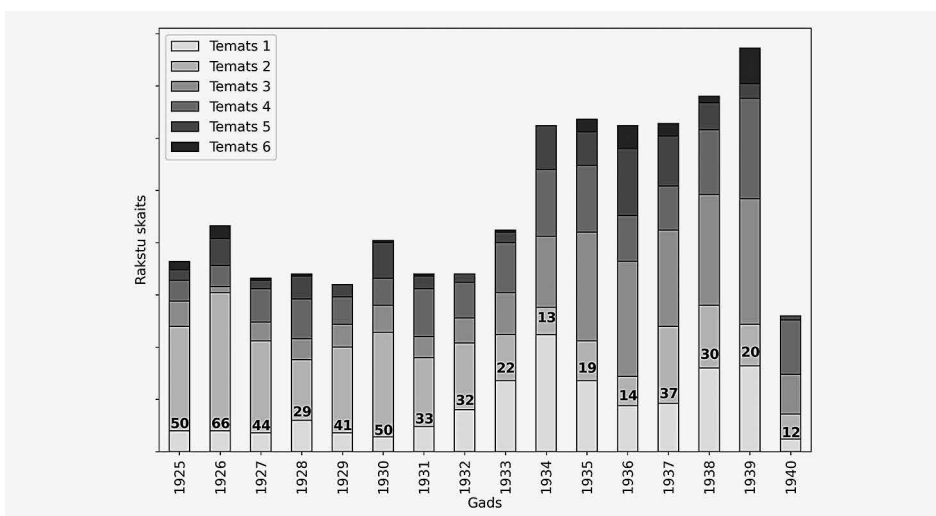
12 Gailītis, M. Valsts vīri par pirmā diplomāta nāvi. *Latvijas Kareivis*, 26.08.1925., 3. lpp.



↑ 1. attēls. Kreisajā pusē: sešu Kalpaka apakškorpusa tematu daudzdimensiju kartējums. Labajā pusē: 30 nozīmīgākie ceturrtā temata jēdzieni.

Vizualizācija veidota, izmantojot *pyLDAvis Python* bibliotēku (Sievert, Shirley 2014).

↓ 2. attēls. Sešu tematu sadalījums Oskara Kalpaka apakškorpusa rakstos no 1925. līdz 1940. g.



2. attēlā aplūkojams katram tematam veltīto rakstu skaits laikposmā no 1925. līdz 1940. gadam¹³. Kopumā Kalpaka pieminējumu skaits izteikti palielinās, sākot ar 1934. gadu, īpaši liels pieaugums vērojams 1. un 3. tematam, savukārt 2. tematam vērojams samazinājums; 6. temats īpaši aktuāls 1939. gadā. Šīs izmaiņas, iespējams, ir interpretējamas 1934. gada valsts apvērsuma kontekstā – lai to noskaidrotu, būtu jāveic kvalitatīva rakstu izpēte. Jāpatur prātā, ka gadījuma izpētēs netika nošķirti un saskaitīti īstie un neīstie Kalpaka pieminējumi, tādējādi, lai nonāktu pie precīzākiem datiem par Kalpaka tematu popularitāti, būtu jāturpina pieminējumu analīzes un filtrēšanas darbs.

Secinājumi

Gadījuma izpēte liecina, ka LDA tematu modelēšanas metodoloģija ir noderīga vēsturiskās periodikas pētniecībai un varētu būt piemērota, lai to integrētu kā jaunu funkcionalitāti vai papildu saskarni LNB digitālajās kolekcijās. LDA metodoloģija ir daudzkārt pārbaudīta lietojumiem citās valodās, un tā tiek uzskatīta par īpaši piemērotu akadēmisku publikāciju, žurnālu un laikrakstu tematu modelēšanai. Papildus šajā rakstā skatītajam LDA variantam pasaulē tikuši izstrādāti risinājumi, kas ļauj precīzāk tvert tematu izmaiņas ilgstošā laika periodā.

Gadījuma izpētes gaitā tika definēta darbplūsma, kas nepieciešama, lai izveidotu uzticamu tematu modeli: datu priekšapstrāde (korpusa kompilēšana, tīrīšana, morfoloģiskā marķēšana), modeļa iteratīva apmācība, provizoriski izveidojot modeļus ar dažādu tematu skaitu, tematu modeļa koherences mērīšana, tematu sadalījuma subjektīva izvērtēšana, optimālā modeļa izvēle, vizualizāciju un citu modeļa reprezentāciju veidošana. No lietotāja viedokļa, izstrādājot atbilstošu pētniecības pakalpojuma risinājumu, vēlams nodrošināt, lai lietotājs var piekļūt modeļa avottekstiem.

Oskara Kalpaka apakškorpusa tematu modelis apliecināja, ka LDA ļauj veidot semantiski saskanīgus, noderīgus tematus, tomēr, lai pilnībā interpretētu rezultātus, nepieciešams vērsties pie pašiem rakstiem. Rakstu turpmāka subjektīva pārbaude atklāja, ka tematu dalījums ļāvis jēgpilni nošķirt rakstus, tomēr sastopama arī neatbilstoša attiecināšana. Secināms, ka Oskara Kalpaka gadījuma izpētē lietotā metode veiksmīgi izmantojama tematu instrumentālisma pieejā, savienojot gan kvalitatīvās, gan kvantitatīvās metodes avotu izpētē. Turpmāka modeļa uzstādījumu pielāgošana un izpēte būtu nepieciešama, lai LDA metodoloģiju varētu izmantot tematu reālisma pieejā vai situācijās, kurās pētniekiem ir ierobežotas iespējas pārlūkot rakstus un mazāk informācijas par to, kā modelis veidots un kādi ir tā ierobežojumi.

Izvēloties modeļa apmācības uzstādījumus, tikai atmesti vārdi, kas lietoti vairāk nekā 50 % tekstu, tādējādi tikai izslēgta lielākā daļa palīgvārdu, vietniekvārdu, bieži lietotu apstākļa vārdu. Lietojumu scenārijos, kuros digitālas kolekcijas lietotājam būtisks tieši raksta priekšmets

13 Raksti tika skaitīti, par pamatu ņemot to tematu, kuram rakstā ir augstākā procentu vērtība.

(referents) un redzami nevis 30, bet tikai daži temata atslēgvārdi, iespējams, būtu ieteicams pielāgot uzstādījumus tādējādi, lai modeli būtu vēl mazāk vārdu bez patstāvīgas nozīmes vai tiktu iekļauti tikai lietvārdi. Savukārt lietojumu scenārijos, kuros pētnieciskie mērķi ir saistīti arī ar tekstu stilistikas un retorikas pētniecību, vārdšķiru daudzveidība paturama. Kā parādīja Oskara Kalpaka modeļa piemērs, skaitļu iekļaušana modelī ne vien nodrošina, ka saglabājas tematam nozīmīgi gadskaitļi (kā 1919. gads), bet ļauj arī identificēt materiālus, kurus pētnieks, iespējams, vēlētos izņemt no pētāmo datu kopas.

Turpinot metodes izstrādi, būtu nepieciešams turpināt pielāgot LDA tematu modeļa uzstādījumus, analizējot arī citas datu kopas, kā arī ieteicams LDA rezultātus salīdzināt ar citu algoritmu veikspēju.

- Abney, Steven, Bird, Steven (2010). The Human Language Project: building a universal corpus of the world's languages. *Proceedings of the 48th Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 88–97.
- Alabi, Jesujoba, Amponsah-Kaakyire, Kwabena, Adelani, David, et al. (2020). *Massive vs. Curated Embeddings for Low-Resourced Languages: the Case of Yorub'a and Twi*. *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, pp. 2754–2762.
- Alves, Diego, Thakkar, Gaurish, Tadić, Marko (2020). Evaluating Language Tools for Fifteen EU-oficial Under-resourced Languages. *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, pp. 1866–1873.
- Baklāne, Anda, Saulespurēns, Valdis (2022). The application of latent Dirichlet allocation for the analysis of Latvian historical newspapers: Oskars Kalpaks' case study. *Nauka. tehnoloģii, inovācijai*, No. 1(21), s. 29–37.
- Blei David M., Lafferty, John D. (2007). A correlated topic model of Science. *Annals of Applied Statistics*, Vol. 1(1), pp. 17–35.
- Blei, David M., Lafferty, John D. (2006). Dynamic topic models. *Proceedings of the 23rd international conference on Machine Learning*, pp. 113–120.
- Blei, David M., Ng, Andrew Y., Jordan, Michael I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3 (January), pp. 993–1022.
- Blei, David (2012). Topic modeling and digital humanities. *Journal of Digital Humanities*, Vol. 2, No. 1, pp. 8–12. Available: <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/> [accessed 18.06.2022.].
- Block, Sharon (2006). Doing More with Digitization: An introduction to topic modeling of early American sources. *Common-place: The Interactive Journal of Early American Life*, 6.2. Available: <http://commonplace.online/article/doing-more-with-digitization/> [accessed 18.06.2022.].
- Bollmann, Marcel (2019). A Large-Scale Comparison of Historical Text Normalization Systems. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1. Association for Computational Linguistics, pp. 3885–3898.
- Brett, Megan R. (2012). Topic Modeling: A Basic Introduction. *Journal of Digital Humanities*, Vol. 2, No. 1, pp. 1–2. Available: <http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/> [accessed 18.06.2022.].
- Chang, Jonathan, Boyd-Graber, Jordan, Gerrish, Sean, et al. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems* 22. Available: <https://proceedings.neurips.cc/paper/2009/file/f92586a25bb3145facd64ab20fd554ff-Paper.pdf> [accessed 18.06.2022.].
- Ehrmann, Maud, Romanello, Matteo, Clematide, Simon, et al. (2020). Language Resources for Historical Newspapers: The Impresso Collection. *IREC 2020 Proceedings*, pp. 958–968.
- Goldstone, Andrew, Underwood, Ted. (2012). What Can Topic Models of PMLA Teach Us About the History of Literary Scholarship? *Journal of Digital Humanities*, Vol. 2, No. 1, pp. 39–48. Available: <http://journalofdigitalhumanities.org/2-1/what-can-topic-models-of-pmla-teach-us-by-ted-underwood-and-andrew-goldstone/> [accessed 18.06.2022.].
- Hall, David, Jurafsky, Daniel, Manning, Christopher D. (2008). Studying the history of ideas using topic models. *Proceedings of the 2008 conference on empirical methods in natural language processing*, pp. 363–371.
- Hengchen, Simon (2017). *When Does it Mean? Detecting Semantic Change in Historical Texts*. Ph.D. thesis. Université libre de Bruxelles.
- Jēkabsons, Ēriks (2022). Oskars Kalpaks. *Nacionālā enciklopēdija*. Pieejams: <https://enciklopedija.lv/skirklis/26024-Oskars-Kalpaks> [skatīts 18.06.2022.].
- Krūmiņa, Līga (2012). Digitalizācija Latvijā pasaules pieredzes kontekstā. *Bibliotēku pasaule*, Vol. 57, 39.–45. lpp.
- Kurvinen, Heidi (2020). Towards Digital Histories of Women's Suffrage Movements. Fridlund, Matts, Oiva, Mila, Paju, Petri (eds.) *Digital Histories: Emergent Approaches within the New Digital History*. Helsinki University Press, pp. 149–163.
- Marjanen, Jani, Zosa, Elaine, Hengchen, Simon, et al. (2020). Topic

- Modelling Discourse Dynamics in Historical Newspapers. *Post-Proceedings of the 5th Conference Digital Humanities in the Nordic Countries (DHN 2020)*, pp. 63–77.
- McGillivray, Barbara (2021). Computational methods for semantic analysis of historical texts. Kristen Schuster, Stuart Dunn. *Routledge International Handbook of Research Methods in Digital Humanities*. London; New York: Routledge, Taylor & Francis Group, pp. 261–274.
- Nelson, Robert K. (2011). *Mining the Dispatch*. Available: <https://dsl.richmond.edu/dispatch/introduction> [accessed 18.06.2022.].
- Newman, David, Chemudugunta, Chaitanya, Smyth Padhraic, et al. (2006). Analyzing entities and topics in news articles using statistical topic models. *Intelligence and Security Informatics, IEEE International Conference on Intelligence and Security Informatics*, pp. 93–103.
- Pääkkönen, Juho, Ylikoski, Petri (2020). Humanistic interpretation and machine learning. *Synthese*, 199 (Sept.), pp. 1461–1497.
- Pētersone, Inta (1999) (red.). *Latvijas Kareivis. Latvijas Brīvības cīņas 1918–1920: enciklopēdija*. Preses Nams, 187. lpp.
- Řehůřek, Radim, Sojka, Petr (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Available: <http://is.muni.cz/publication/884893/en> [accessed 18.06.2022.].
- Rhody Lisa M. (2012). Topic Modeling and Figurative Language. *Journal of Digital Humanities*, Vol. 2, No. 1. Available: <http://journalofdigitalhumanities.org/2-1/topic-model-data-for-topic-modeling-and-figurative-language-by-lisa-m-rhody/> [accessed 18.06.2022.].
- Röder, Michael, Both, Andreas, Hinneburg, Alexander (2015). Exploring the Space of Topic Coherence Measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining WSDM '15*, pp. 399–340.
- Sievert, Carson, Shirley, Kenneth (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the workshop on interactive language learning, visualization, and interfaces*. Association for Computational Linguistics, pp. 63–70.
- Skadiņa, Inguna, Veisbergs, Andrejs, Vasiljevs, Andrejs et al. (2012). The Latvian Language in the Digital Age / Latviešu valoda digitālajā laikmetā. *META-NET White Paper Series: Latvian*. Berlin: Springer.
- Templeton, Thomas C., Brown, Travis, Battacharyya, Sayan, et al. (2011). Mining the Dispatch under Supervision: Using Casualty Counts to Guide Topics from the Richmond Daily Dispatch Corpus. *Chicago Colloquium on Digital Humanities and Computer Science*.
- Ūdre, Dace, Baltiņa, Dagnija et al. (2019). *Digital Approaches in Cultural Heritage: towards a pan-Baltic cooperation network: final report*. Riga: National Library of Latvia. Available: <https://dom.lndb.lv/data/obj/781145.html> [accessed 15.09.2022.].
- Underwood, Ted (2012). *Topic modeling just made simple enough*. Blog post. Available: <https://tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough/> [accessed 18.06.2022.].
- Viksna, Rinalds, Kirikova, Marite, and Kiopa, Daiga (2020). Exploring the Use of Topic Analysis in Latvian Legal Documents. *COURT - CAiSE for Legal Documents, Virtual Workshop*. Available: <http://eur-lex.org/Vol-2690/COURT-paper4.pdf> [accessed 18.06.2022.].
- Viola, Lorella, Verheul, Jaap (2019). Mining ethnicity: Discourse-driven topic modelling of immigrant discourses in the USA, 1898–1920. *Digital Scholarship in the Humanities*, Vol. 35(4), pp. 921–943.
- Wallach, Hanna, Mimno, David, McCallum, Andrew (2009). Rethinking LDA: Why priors matter. *Advances in Neural Information Processing Systems*, Vol. 23 (January), pp. 1973–1981.
- Znotiņš, Artūrs, Cīrule, Elita (2018). NLP-PIPE: Latvian NLP Tool Pipeline. Human Language Technologies. *The Baltic Perspective*, IOS Press, Vol. 307, pp. 183–189.
- Zariņš, Uldis (2014). Eiropas kultūras mantojums digitālajā vidē. *Latvijas intereses Eiropas Savienībā*, Nr. 2, 41.–55. lpp. Pieejams: <https://dom.lndb.lv/data/obj/61436.html> [skatīts 15.09.2022.].

The Model of Latent Dirichlet Allocation in the Topic Analysis of Latvian Soldier: Oskars Kalpaks' Case Study

Anda Baklāne, Valdis Saulespurēns

Keywords: topic modelling, digitized newspapers, digital history, topic coherence, National Library of Latvia

The paper presents a case study of the application of the LDA (latent Dirichlet allocation) model for the analysis of topics in the corpus of the historical daily newspaper of Latvian armed forces *Latvian Soldier* (1925–1940). Although topic modelling is one of the most popular techniques for analysing text in digital humanities, this methodology has not been extensively tested for texts in Latvian. The case study was conducted to explore the possibilities for implementing topic models as new functionality for exploring newspapers in the digital library of the National Library of Latvia. To imitate different use cases of topic modelling, two models were created: a model consisting of 50 topics for the whole corpus of the *Latvian Soldier*, as well as a six-topic model of the subcorpus compiled from articles that contain the name 'Kalpaks'. It was demonstrated that both models produced usable, semantically coherent topics that could aid the exploration of historical newspapers. It was concluded that the quality of the models in the current state was sufficient to follow the approach of topic instrumentalism, which views topics as incomplete representations of texts that are a useful augmentation of the investigative process. The acquired topic models seem particularly useful for combining research practices of distant and close reading. Further testing and adjustment of the parameters are needed to produce concise and unambiguous topics that could be reliably used in research situations where extensive analysis of the sources and verification is not expected.

LNB atvērtie dati pētniecībai

Mārīte Apenīte, Uldis Bojārs, Ansis Garda,
Anita Goldberga, Maira Kreislere,
Anita Rašmane, Elza Ungure

Šo pētījumu ir finansējusi Latvijas Zinātnes padome, projekts "Latvijas atmiņas institūciju dati digitālajā telpā: vienojot kultūras mantojumu", projekta Nr. lzp-2019/1-0365.

Ievads

Kultūras mantojuma institūcijas – galerijas, bibliotēkas, arhīvi un muzeji – dokumentē, kontekstualizē, saglabā un indeksē kultūru un kolektīvo atmiņu (Pessach 2008). Tādējādi šīs institūcijas nodrošina pagātnes pieredzes un radošo izpausmju pārnesi uz tagadni un nākotni (Byrne 2014). Mūsdienās atmiņas institūcijām jārežķinās ar paradigmas maiņu tajā, kā zināšanas un dati tiek saglabāti un izplatīti, publiskajā diskursā arvien lielāku uzsvāru liekot uz digitālu brīvpieklūves datu un saistīto atvērto datu¹ (*linked open data*) nozīmi, kas izpaužas tādos starptautiskos projektos kā *European Open Science Cloud* un no tā atvasinātajos projektos – *HIRMEOS*, *OPERAS* un *OpenAIRE*. Salīdzinājumā ar vēsturiskajām alternatīvām tieši atvērti, identificējami, strukturēti un saistīti mašinlasāmi dati tiek uztverti kā labāk pieejami, atkārtoti lietojami un neizsmeļami, tādēļ tiek uzskatīts, ka tiem ir lielāks potenciāls tikt izmantotiem dažādos pētījumos, uzlabot zinātnisko atklātību, kā arī veicināt sadarbību starp dažādām pusēm (Zhang u. c. 2011).

Eiropas līmenī augstas kvalitātes datu pieejamības nodrošināšanu radošiem un inovatīviem risinājumiem paredz Eiropas datu stratēģija². Līdzīgi kā Eiropas datu stratēģijā, arī Latvijas attīstības stratēģijā līdz 2030. gadam (Latvija 2030) kā viens no rīcības virzieniem noteikta Latvijas kultūras kapitāla saglabāšana un attīstība, paplašinot Latvijas digitālo kultūras telpu, “piedāvājot precīzus un objektīvus informācijas avotus, kā arī kompetentus komentārus un analīzi” (Saeima 2010: 18). Tādēļ arī LNB 2019.–2023. gada attīstības stratēģija nosaka, ka “datus veidojam ne vien ērtai un efektīvai piekļuvei, bet arī pētniecībai. Nacionālas nozīmes datus bagātināsim atkārtotai izmantošanai un tīmeklī publicēsim saistīto datu formā, lai LNB veidotie dati būtu savienojami ar citu institūciju datiem, veidojot lietotājiem draudzīgas un pilnīgas piekļuves formas” (LNB 2019), kā arī lai veicinātu starpnozaru zināšanu jēgpilnu savietošanu un atkārtotu izmantošanu.

- 1 Saistītie dati ir principu un tehnoloģiju kopums (Berners-Lee 2006; Bizer et al. 2011) strukturētu datu publicēšanai tīmeklī. Saistītie dati ir veids, kā tīmeklī publicēt mašinlasāmus datus par dažādiem objektiem (piemēram, bibliotēkas kolekcijas darbiem), to īpašībām un savstarpējām saitēm. Uz objektiem var atsaukties, izmantojot unikālus URI (vienotais resursu identifikators). Atvērtie dati ir dati, kuriem jebkurš lietotājs var brīvi piekļūt, tos brīvi izmantot un izplatīt tālāk. Atvērto saistīto datu veidošanā saistīto datu principi tiek lietoti datu publicēšanai brīvpieejā.
- 2 <https://eur-lex.europa.eu/legal-content/LV/TXT/PDF/?uri=CELEX:52020DC0066&from=EN>

LNB misija nepārtraukti attīstīt pārbaudītu, ticamu, mūsdienu standartiem atbilstošu zināšanu bāzi, izmantojot atbilstošus datu standartus un apstrādes rīkus, un tās piekļuvi pētniekiem, studentiem un plašākai sabiedrībai, izriet gan no *Latvija 2030* ietvertā rīcības virziena, gan LNB attīstības stratēģijas. Tomēr vienlaikus LNB kā atmiņas institūcija apzinās, ka tās glabātais kultūras mantojums ir tikai daļa no kopējā nacionālā kultūras mantojuma un lietotāju vajadzību labākai apmierināšanai nepieciešama sadarbība ar citām atmiņas institūcijām – kultūras mantojuma glabātājām.

Kultūras informāciju sistēmu centra (KISC) mērķis ir nodrošināt kultūras mantojuma institūcijas ar nepieciešamajiem informācijas tehnoloģiju resursiem un sistēmām katrā nozarē. Tas nozīmē, ka arhīvos, bibliotēkās, muzejos u. c. atmiņas institūcijās katrā nozarē tiek izmantotas līdzīgas vai vienas un tās pašas zināšanu organizācijas formas un datu vērtības tiek interpretētas līdzīgi. Tomēr vienlaikus tas nozīmē, ka kopumā Latvijas atmiņas institūcijas izmanto dažādus, savstarpēji ne vienmēr saderīgus datu un zināšanu organizācijas modeļus, kas apgrūtina resursu integrāciju vienotās digitālās platformās un kolekcijās. Digitālā vide rada jaunas sadarbības un semantiskās kontekstualizācijas iespējas, kas Latvijā līdz šim nav izpētītas un ieviestas atmiņu institūciju sadarbībā, kā tas, piemēram, veiksmīgi noticis Somijā u. c. Eiropas un pasaules valstīs³.

Tādēļ viens no Latvijas atmiņas institūciju, t. sk. LNB, mērķiem ir attīstīt semantiskās sadarbības potenciālu datu jomā un veicināt dažādu institūciju nevienmērīgo datu kopu sadarbību digitālajā telpā, ļaujot sasaistīt šo informāciju ar ārējiem avotiem un nodrošinot viedu, intuitīvu meklēšanas funkcionalitāti, kas var palīdzēt pētniekiem, citām ieinteresētajām pusēm (piemēram, izdevējiem, žurnālistiem utt.) un plašākai sabiedrībai (skat. Gruber 2008).

Paredzot uzdevumu uzlabot LNB veidoto datu kopu un vārdnīcu kvalitāti un radot iespējas tās padarīt par lietotājam maksimāli draudzīgiem datiem, t. sk. saistīto datu kopām un ontoloģijām, LNB 2020. gadā uzsāka projektu “Latvijas atmiņas institūciju dati digitālajā telpā: vienojot kultūras mantojumu” (projekta Nr. lzp-2019/1-0365). Tā ietvaros 2021. gadā tika izveidots LNB Atvērto datu portāls⁴, kurā publicētas četras uzticamu, pārbaudītu un strukturētu datu kopas.

Saistīto datu kolekcija “Rainis un Aspazija” satur atvērtus un strukturētus datus, kas kolekcijas pētniekus aizved uz konkrētiem faktiem par Raini un Aspaziju. Nacionālā tēzaura datu kopas, kas ir publicētas LNB ontoloģiju servīs, atkārtotai izmantošanai padara pieejamu informāciju par tematiskajiem priekšmetiem, ģeogrāfiskajiem nosaukumiem un formas/žanra terminiem. Nacionālās bibliogrāfijas datu kopa satur informāciju par izvēlētajā laika periodā (2017–2020) Latvijā publicētajām grāmatām, to autoriem u. c. saistīto informāciju. Izdevējdarbības statistikas datu kopā ir dati par izdevējdarbību Latvijā, sākot no 1991. gada. Dati publicēti neapkopotā veidā, neuzrādot ziņas par izdevumu tirāžu.

3 Daži jau publicētu atmiņas institūciju atvērto saistīto datu kopu piemēri ir Britu nacionālā bibliogrāfija, ASV Kongresa bibliotēkas saistīto datu serviss, vairākas Somijas Nacionālās bibliotēkas datu kopas (ontoloģiju serviss *Finto*, nacionālā bibliogrāfija un citas) kā arī un VIAF projekts (Hickey and Toves 2014).

4 <https://dati.lnb.lv/>

Turpmākajās nodaļās iepazīstinām ar datu kopu saturu, datu kopu veidošanas specifiku un ar to saistītajiem izaicinājumiem un nākotnes perspektīvām, kas, iespējams, būs aktuālas arī citām atmiņas institūcijām, veidojot nacionālajā un starptautiskajā digitālajā telpā sadarbīgus datus.

Saistīto datu kolekcija “Rainis un Aspazija” (RunA)

Saistīto datu kolekcija “Rainis un Aspazija” (RunA)⁵ sāka veidot 2014. gadā un iecerēta kā vienots piekļuves punkts abu dzejnieku literārajam un epistulārajam mantojumam, kā arī visdažādākajiem ar viņu radošo darbību un dzīvi saistīto rakstisko materiālu, foto, audio, kinohroniku u. c. digitalizētajiem resursiem. Šobrīd kolekcijā ir iekļauti tādi satura veidi kā Raiņa un Aspazijas darbu pirmizdevumi (tajā skaitā darbi, kuri publicēti tikai periodikā), šo darbu literārās anotācijas, dzejnieku savstarpējā sarakste (pirmslaulību periods un Šveices trimdas periods), fotogrāfijas, arhīvu dokumenti u. c. materiāli.

RunA kolekcijas satura daudzpusību, dažādību un plašo aptvērumu nebūtu iespējams nodrošināt vienas institūcijas robežās, jo tajā iekļautie materiāli fiziski glabājas dažādās kultūras mantojuma institūcijās. Kolekcija ir kopdarbības rezultāts, kas tapis, iesaistoties Latvijas Nacionālā arhīva, Latvijas Nacionālās bibliotēkas, Latvijas Universitātes Literatūras, folkloras un mākslas institūta, Memoriālo muzeju apvienības un Rakstniecības un mūzikas muzeja speciālistiem.

Digitālo objektu skaits RunA kolekcijā 2021. oktobrī ir vairāk nekā 1300 objektu visdažādākajās digitālajās formās un izmantojamības līmeņos (t. sk. 83 literārie darbi un 395 vēstules un to anotētie teksti). Kolekcija satur arī audio un video datnes, arhīva dokumentu digitalizētos attēlus, prezentāciju datnes un citus satura veidus. Tekstveida objekts “vēstule” kolekcijā iekļauts gan kā vēstules vai atklātnes lappušu attēlu grupa, gan kā anotēts mašīnlasāms teksts; oriģināldarbs – kā digitalizēta grāmata un strukturēti bibliogrāfiskie dati. Kolekcijā šobrīd nav iekļauti Raiņa un Aspazijas tulkoto darbu teksti, izņemot Johana Volfganga fon Gētes “Faustu”. Turpmāk paredzēts iekļaut arī anotētus Raiņa dienasgrāmatas tekstus.

RunA kolekciju no citām digitālajām kolekcijām un digitālo resursu repozitorijiem atšķir datu organizācijas metode – tā ir saistīto datu kolekcija, kas papildus digitalizētajiem objektiem satur identificēto entīšu (personu, organizāciju, vietu, datējumu, notikumu u. c.) tīklu, kas sasaista šos objektus, piešķir tiem nozīmi, vietu un laiku, piesaista datus no citiem publiski pieejamiem datu avotiem, ļaujot analizēt un vizualizēt Raiņa un Aspazijas darbības sociālo vidi un biogrāfiskos faktus. RunA šim nolūkam tiek izmantoti arī bibliogrāfiskie metadati.

5 <https://runa.lnb.lv/en/>

Rainis un Aspazija



Ātrā meklēšana...

Objekti



Entītes

1. attēls. Digitālo objektu un entīšu meklēšanas saskarne.

Daļa no kolekcijas tekstveida materiāliem – darbu literārās anotācijas un Raiņa un Aspazijas sarakste – ir manuāli anotēti, tekstā atzīmējot dažādu entīšu (personas, vietas, datumi u. c.) pieminējumus. Šī darbība tiek veikta īpašā teksta anotēšanas rīkā (Bojārs u. c. 2018), un anotācijas ir tās, kas veido tīklu ar saitēm starp teksta dokumentiem un tajos pieminētajām entītēm. Šobrīd, anotējot RunA kolekcijas materiālu saturiskajai apstrādei, tiek izmantotas deviņas entīšu klases (persona, institūcija, darbs, vieta, ēkas un celtnes, notikums, laiks, jēdziens, lieta), kurām tieši RunA materiālu apstrādei atbilst 78 specializētas anotāciju klases. Anotēšanas rīkā ir iespējams ne tikai atzīmēt entīšu pieminējumus, veidojot vienkāršas anotācijas un saglabājot atbilstošās entītes datus īpašā entīšu datubāzē, bet arī veidot saliktās anotācijas (vienkāršo anotāciju loģiskus “teikumus”), kā arī iezīmēt dokumenta struktūru ar strukturālajām anotācijām (Bojārs u. c. 2018).

Anotācijas veido datu kopu, kas ļauj jēgpilni saistīt gan RunA kolekcijas objektus un entītes, gan tekstā minētās entītes ar ārējos tīmekļa informācijas avotos atrodamiem datiem. RunA kolekcijā tiek publicēti atvērti, strukturēti dati par kolekcijas objektiem, entītēm un saitēm to starpā (piemēram, saite no dzejnieku darba literārās anotācijas uz vēstulē pieminētu entīti).

Balstoties uz datiem par kolekcijas objektiem, entītēm un saitēm to starpā, tiek veidoti RunA datu tīklojumi. Datu tīklojums vizuāli attēlo kolekcijas objektu un entīšu savstarpējo saistību. Kolekcijā ir divu veidu tīklojumi – objektu tīklojums un entīšu tīklojums. Objektu tīklojums ir datu vizualizācija (grafs), kurā divas virsotnes (kolekcijas objekti) ir savstarpēji saistītas, ja abos dokumentos ir pieminēta viena un tā pati entīte. Entīšu tīklojums ir vizualizācija, kurā divas virsotnes (entītes) ir savstarpēji saistītas, ja abas entītes ir pieminētas vienā un tajā pašā anotētajā dokumentā.

Anotējot kolekcijas tekstveida materiālus, tika izveidoti vairāk nekā 14 tūkstoši anotāciju, kas atsaucas uz 2600 identificētām entītēm. Publiski pieejamajā RunA kolekcijas portālā šo bagātīgo informāciju iespējams atrast un izmantot tikpat vienkārši kā digitālos objektus. Lai to izdarītu, tikai jāpārvelk virtuāls meklēšanas “slēdzītis” no “Objekti” uz “Entītes”:

ENTĪTE

Rediģēt

Rediģēt saites

Dzēst

Formāts

Normālforma

Žans Pjērs Beranžē, 1780-1857

Klase

Persona

Normālformas

- lav Žans Pjērs Beranžē, 1780-1857 primārā
- lav Пjērs Žанс Berанžē, 1780-1857
- rus Пjер-Жан де Беранже, 1780-1857 primārā
- ger Pierre-Jean de Béranger, 1780-1857 primārā
- eng Pierre-Jean de Béranger, 1780-1857 primārā
- fre Pierre-Jean de Béranger, 1780-1857 primārā

Ārējo identifikatoru saraksts

- Web ALEPH AUT https://kopkatalogs.lv/F/?func=direct&local_base=LNC10&doc_number=000075826
- Web VIAF <http://viaf.org/viaf/12319139>
- Web ISNI <http://www.isni.org/isni/0000000120956040>
- Web Nekropole <https://nekropole.info/lv/person/view?id=38733>
- Web EN Wikipedia https://en.wikipedia.org/wiki/Pierre-Jean_de_Béranger
- Web RU Wikipedia https://ru.wikipedia.org/wiki/Беранже,_Пjер-Жан_де
- Web DE Wikipedia https://de.wikipedia.org/wiki/Pierre-Jean_de_Béranger
- Linked Data Wikidata <https://www.wikidata.org/wiki/Q324998>
- Web Vikikrātuve https://commons.wikimedia.org/wiki/Category:Pierre-Jean_de_Béranger
- Web ADK
- Linked Data WorldCat identity <https://www.worldcat.org/identities/lccn-n81055604>

Piezīmes

Franču dzejnieks.

Sākuma lapa > Žans Pjērs Beranžē, 1780-1857

Turietē PDF

RDF/XML

Datu shējums

Beļāms

Žans Pjērs Beranžē, 1780-1857

Persona

Citas rakstības formas

- lav Žанс Pjјерс Berанжē, 1780-1857
- fre Pierre-Jean de Béranger, 1780-1857
- eng Pierre-Jean de Béranger, 1780-1857
- rus Пjер-Жан де Беранже, 1780-1857
- ger Pierre-Jean de Béranger, 1780-1857
- lav Pjјерс Жанс Беранжē, 1780-1857

Saites

ALEPH AUT 000075826 →

VIAF 12319139 →

ISNI 0000000120956040 →

Nekropole person/view?id=38733 →

EN Wikipedia Pierre-

Jean, de Béranger →

RU Wikipedia Беранже, Пjер-Жан_де →

DE Wikipedia Pierre-Jean, de Béranger →

Wikidata Q324998 →

Vikikrātuve Pierre-Jean de Béranger →

ADK 159114 →

WorldCat identity lccn-n81055604 →

Pieminējumi (2)

NR	Tips	Objekts	Pieminējuma forma
1	Teksts	Jānis Raiņis,"Vidpūda kalendārs"(1907-1908),Annotācija	Žанс Pjјерс Berанжē
2	Teksts	The Whirlwind Calendar,1907-1908 of J. Raiņis, Compendium	Pierre Jean de Beranger

↑ 2. attēls. Franču dzejnieka Žана Pjјера Berанжē entītes lapa anotatora entīšu datubāzē.

↓ 3. attēls. Franču dzejnieka Žана Pjјера Berанжē entītes lapa kolekcijā RunA.

Letonica 47

LNB atvērtie dati pētniecībai

2022

Piemēram, 2. attēlā ir redzami dati par franču dzejnieku Žanu Pjēru Beranžē, kas anotatora entīšu datubāzē glabājas teksta un hipersaišu formā.

RunA portālā šie dati tiek parādīti, informāciju bagātinot gan ar datiem no Vikipēdijas, gan informāciju, kuros RunA kolekcijas teksta dokumentos šī entīte pieminēta. Portālā iespējams piekļūt atvērtajiem saistītajiem datiem par šo entīti un tās datu tīklojumam.

Vairums kolekcijas satura ir brīvi pieejams atkārtotai lietošanai un ir izmantojams kā atvērtie dati. Mašīnlasāmi dati par kolekcijas objektiem, entītēm un saitēm to starpā ir publicēti atvērto saistīto datu formā (Bojārs u. c. 2020). Šie dati ir pieejami gan RunA kolekcijas vietnē, kur tie ir publicēti atbilstoši saistīto datu principiem (Berners-Lee 2006), gan apkopotā formā LNB atvērto datu portālā. No kolekcijas tīmekļa lapām objektu un entīšu datiem var piekļūt, spiežot pogas “Turtle RDF” un “RDF/XML” (vieglāk uztverams ir Turtle RDF formāts) vai pievienojot lapas adresei paplašinājumu atbilstošā formātā (.ttl vai .rdf) – piemēram, <https://runa.lnb.lv/60316.ttl>. LNB atvērto datu portālā ir publicētas divas šīs kolekcijas datu kopas, kas satur apkopotus kolekcijas datus Turtle RDF un RDF/XML formā.

Nacionālais tēzaurs

Datu un zināšanu organizācijas un pārvaldības politikas pārmaiņas notiek ne tikai bibliotēku katalogu vidē pasaulē, bet arī LNB Nacionālā tēzaura autoritatīvajos datos, jo ir būtiski nodrošināt vienotu kontrolēto terminu lietojumu ne vien bibliogrāfiskajos ierakstos bibliotēku elektroniskajos katalogos un dažādās datubāzēs, bet veicināt arī datu atkārtotu izmantojamību un savietojamību ar citām starptautiskajām kontrolētajām vārdnīcām tīmekļa vidē.

Latvijas Nacionālā tēzaura pamatu veido LNB Priekšmetu saraksts (LNBPS), kura izveidē kopš 2000. gada tiek lietots ASV Kongresa bibliotēkas Priekšmetu saraksts (*Library of Congress Subject Headings* – LCSH⁶), kas tiek adaptēts atbilstoši latviešu valodas un zinātņu nozaru terminoloģijas prasībām. LNBPS tiek veidots kā kontrolēta un strukturēta vārdnīca ar tēzauram līdzīgu struktūru, kurā jēdzieni tiek attēloti ar terminiem, starp kuriem ir hierarhiskās, ekvivalentās un asociatīvās attiecības, un tas tiek izmantots universāla satura indeksēšanai ar kontrolētiem terminiem, tādējādi nodrošinot dažādu informācijas resursu tematisko aptvērumu, piekļuvu un informācijas izguvi.

Lai nodrošinātu priekšmetu datu kvalitātes uzturēšanu, piekļuves un informācijas izguves precizitāti, piedāvājot lietotājiem izvēlēties starp plašākiem, šaurākiem, asociatīviem vai līdzīgas nozīmes terminiem, kontrolējot to sinonīmiju, kvazisinonīmiju un homonīmiju, LNB tiek veidota un uzturēta Autoritatīvā datubāze, kas ietver Nacionālā tēzaura autoritatīvos datus (tematiskie, ģeogrāfiskie nosaukumi un formas/žanra priekšmeti) un bibliotēku datu

6 <https://id.loc.gov/authorities/subjects.html>

atveidošanā ieviešamā jaunā starptautiskā standarta RDA (*Resource Description and Access*)⁷ entīšu (personas; dzimtas/ģimenes; institūcijas; pasākumu nosaukumi; darbu nosaukumi) autoritatīvos datus.

Kopš 2004. gada tiek veidoti ģeogrāfisko nosaukumu un tematisko priekšmetu autoritatīvie ieraksti, kas nodrošina ne vien šo priekšmetu pamatformu kontrolētu lietošanu, bet arī bilingvālu (latviešu un angļu) un daudzvalodu priekšmetisko piekļuvi un informācijas izguvi atsevišķu nozaru terminiem vai konkrētu objektu nosaukumiem kā tematiskajiem priekšmetiem vai ģeogrāfisko nosaukumu priekšmetiem.

2014. gadā uzsākta arī pakāpeniska formas/žanra priekšmetu veidošana, izmantojot FAST (*Faceted Application of Subject Terminology*)⁸ un LCGFT (*Library of Congress Genre/Form Terms*)⁹. Atsevišķos gadījumos, ja nav ekvivalenta priekšmeta ne FAST, ne LCGFT, tad LNB tiek izmantoti arī Čehijas Nacionālās bibliotēkas ECZENAS autoritatīvie ieraksti, vai tiek veidoti jauni lokālie priekšmeti atbilstoši nacionālās valodas terminradei, kultūras vajadzībām un zinātņu nozaru terminoloģijai. Kopš 2018. gada Latvijas vietu nosaukumiem pievieno arī koordinātas no Latvijas Vietvārdu datubāzes.

Latvijas Nacionālā tēzaura kontrolētie priekšmeti tiek izmantoti LNB daudzveidīgā krājuma satura atspoguļošanai un priekšmetiskās piekļuves nodrošināšanai elektroniskajā katalogā¹⁰, speciālo krājumu kolekcijās, Nacionālajā analītiskās datubāzē un Digitālās bibliotēkas kolekcijās¹¹. Tie nodrošina kontrolētu tematisko piekļuvi un datu izguvi arī Valsts nozīmes kopkataloga bibliotēkās un publisko bibliotēku katalogos.

Lai arī LNBPS, kas veido Nacionālā tēzaura pamatu, lietošana ir nodrošinājusi vienotu LNB un citu bibliotēku datu pārvaldības un zināšanu organizēšanas politiku jau 21 gadu, tomēr mūsdienu informācijas tehnoloģiju attīstība, dažādu informācijas resursu plašā un daudzveidīgā pieejamība interneta vidē un kultūras mantojuma digitalizācija veicina ne vien datu apjoma palielināšanos, bet maina arī lietotāju informācijas meklēšanas un lietošanas paradumus un vajadzības.

Zināšanu organizācijas pārmaiņu nepieciešamību bibliotēku vidē ietekmē arī saistītie dati, kuri dod iespēju datus publicēt tīmeklī, piešķirot to ierakstiem unikālus URI identifikatorus¹² un veidojot saites ierakstu starpā. Tomēr bibliotēku datu veidošanas formāta MARC21 iespējas neatbilst saistīto datu veidošanas principiem un nenodrošina saistīto datu veidošanu

7 <https://www.rdatoolkit.org/about>

8 <https://www.oclc.org/research/areas/data-science/fast.html>

9 <https://id.loc.gov/authorities/genreForms.html>

10 <https://lnb.lv/>

11 <https://lnb.lv/lv/digitala-biblioteka>

12 URI (vienotais resursu identifikators) ir identifikatoru veids, kas ļauj viennozīmīgi norādīt uz informācijas resursiem. HTTP URI izmantošana resursu identifikācijai ir viens no saistīto datu principiem.

un eksponēšanu lietotājiem, tāpēc autoritatīvajiem datiem jāpievieno hipersaites, lai nodrošinātu veidu, kā no MARC ieraksta piekļūt tīmekļa resursiem.

2019. gadā tika sāta Nacionālā tēzaura jaunveidoto un iepriekšveidoto priekšmetu autoritatīvo ierakstu bagātināšana ar hipersaitēm “uz autoritatīviem ārējiem resursiem (identifikatoru sistēmām un enciklopēdijām)” (LNB 2021: 17), tādējādi nodrošinot LNB autoritatīvo datu iekļaušanu “tīmeklī esošu strukturētu enciklopēdisku datu vidē” (LNB 2021: 17).

2020. gadā automatizēti tika atlasītas LNB tematisko priekšmetu angļu formu sakritības ar LCSH. Tika izskatīti trīs atlasīto ierakstu saraksti – tiešās atbilstības (ieraksti ar vienu LCSH identifikatoru), dublikāti (ieraksti ar vairākiem LCSH identifikatoriem) un kļūdu gadījumi (neatbilst nevienam LCSH identifikatoram). Pēc atlasīto datu izskatīšanas un kļūdu novēršanas LCSH URI identifikatori tika automatizēti pievienoti LNB tematisko priekšmetu autoritatīvajos ierakstos. 2021. gadā līdzīgi tika sāta arī FAST identifikatoru pievienošana LNB tematisko priekšmetu autoritatīvajiem ierakstiem.

Saistot Nacionālā tēzaura autoritatīvos ierakstus ar Kongresa bibliotēkas saistītajiem datiem (*LC Linked Data*), kas ietver saites ar *FAST Linked Data*, *Wikidata* un *General Finnish Ontology YSO*, Nacionālā tēzaura autoritatīvie ieraksti var tikt savstarpēji saistīti arī ar citām atvērto saistīto datu kopām tīmeklī.

Turpinās arī jauno Nacionālā tēzaura autoritatīvo ierakstu bagātināšana ne vien ar LCSH un FAST identifikatoriem, bet tiek pievienotas hipersaites arī ar LCNAF (*Library of Congress Name Authority File*), LCGFT, VIAF, *Wikidata*, NKMP (Nacionālā kultūras mantojuma pārvalde), Latvijas Vietvārdu datubāzes un Nacionālās enciklopēdijas identifikatoriem. Pakāpeniski redakcijas procesā ar ārējo resursu hipersaitēm tiek bagātināti arī iepriekšveidotie autoritatīvie ieraksti.

Lai ievicinātu Nacionālā tēzaura autoritatīvo datu atkalizmantošanu un sadarbību ar citu institūciju datiem un veidotu lietotājiem jaunas piekļuves formas, 2021. gadā tika arī sāts jauns daudzpakāpju process autoritatīvo datu pielāgošanai tīmekļa tehnoloģijām un publicēšanai atvērto saistīto datu formā, un LNB tika izveidots ontoloģiju serviss¹³ (Apenīte, Bojārs 2021).

Ontoloģiju kā zināšanu pārvaldības modeļa izstrāde ir salīdzinoši jauna prakse kultūras un publiskajā sektorā. Daudzviet to potenciāls vēl nav pilnībā apzināts un izmantots, taču, piemēram, ir atzīts un ieviests Somijā, kur kopš 2011. gada tiek izmantots Somijas tēzauru un ontoloģiju serviss (*Finto*). Praktiski šajā pārvaldības modelī zināšanas, ar kurām līdz šim nav bijis iespējams apmainīties digitāli, ir jāattēlo tā, lai automatizēta sistēma tās saprastu, izmantotu un nodotu šo zināšanu kopumu to lietotājiem – cilvēkiem – saprotamā veidā (Milić u. c. 2018).

13 <https://dati.lnb.lv/onto/lv/>

Skosmos Vocabularies About Feedback Help | Latviski

LNB tematisko priekšmetu ontoloģija Content language: Latvian Search

Alphabetical Hierarchy New

Filoloģija

- Gramatika, salīdzināmā un vispārīgā
- Indiešu filoloģija
- Literatūra
- Salīdzināmā literatūrzinātne
- Valoda un valodas

PREFERRED TERM	Filoloģija 
NARROWER CONCEPTS	Gramatika, salīdzināmā un vispārīgā Indiešu filoloģija Literatūra Salīdzināmā literatūrzinātne Valoda un valodas
NOTE	Šeit apkopoti vispārīga satura darbi par kādas tautas vai tautu grupas valodu, literatūru un folkloru kopumā. Darbiem par atsevišķu tautu vai tautu grupas filoloģiju piešķir tematisko priekšmetu Filoloģija ar atbilstošu etnisko vai lingvistisko apzīmējumu, piem. Latviešu filoloģija; Krievu filoloģija; Slāvu filoloģija; Baltu filoloģija; Vācu filoloģija Source: Enciklopēdiskā vārdnīca, 1991:
IDENTIFIER	LNC10-000047488 => ALEPH authority record
IN OTHER LANGUAGES	Philology English
URI	http://dati.lnb.lv/onto/subject/LNC10-000047488 
DOWNLOAD THIS CONCEPT:	RDF/XML Turtle JSON-LD Created 6/2/04, last modified 3/18/20
CLOSELY MATCHING CONCEPTS	philology (en) ALLFO - Allmän finländsk ontologi Philology (en) Library of Congress Subject Headings

4. attēls. Jēdziena “filoloģija” lapa LNB ontoloģiju servīsā, *Skosmos* vidē.

LNB ontoloģiju servīsā ietilpst trīs datu kopas: vispārīgo un individuālo jēdzienu ontoloģija (*NLL subject authority ontology*), ģeogrāfisko nosaukumu ontoloģija (*NLL geo ontology*) un formas/žanra ontoloģija (*NLL genre/form ontology*). Kopumā tajās ir 55 740 jēdzieni.

Tematisko priekšmetu ontoloģija (*NLL subject authority ontology*) ietver tematiskos priekšmetus: abstraktus jēdzienus, dažādu nozaru terminus, svētku nosaukumus, arhitektoniskos objektus ar vai bez nosaukuma: arkas, baznīcas, klosterus katedrāles u. tml. reliģiskas celtnes, muižas un pilis, pieminekļus, memoriālos ansambļus, konkrētu ēku nosaukumus u. c., kas raksturo dažādu informācijas resursu saturu.

Ģeogrāfisko nosaukumu ontoloģijā (*NLL geo ontology*) ietverti administratīvie (valstis, apgabali, novadi, rajoni, provinces, pilsētas, pagasti, ciemi) un neadministratīvie vietu nosaukumi (planētas, kontinenti, okeāni, jūras, upes, salas, kalni, meži, kultūrvēsturiski novadi, neadministratīvas pilsētu daļas, ostas, dzelzceļa stacijas, dzelzceļa līnijas, izzudušas pilsētas, arheoloģisko izrakumu vietas, kapsētas, nacionālie parki un dabas liegumi, parki, dārzi u. tml. vietu nosaukumi).

Formas/žanra ontoloģija (*NLL genre/form ontology*) ietver priekšmetus, kas raksturo dažādu informācijas resursu formu vai žanru, piem., biogrāfijas, enciklopēdijas, katalogi, terminu vārdnīcas, dzeja, proza, latviešu proza, angļu stāsti u. tml. formas vai žanra priekšmeti.

Lai izveidotu ontoloģiju servisa datu kopas, Latvijas Nacionālā tēzaura autoritatīvie ieraksti tika pārveidoti atvērto saistīto datu formā un publicēti tīmeklī, izmantojot *Skosmos*¹⁴, kas ir atvērta pirmkoda tīmekļa rīks SKOS aprakstīto kontrolēto vārdnīcu un tēzauru publicēšanai, pārlūkošanai, meklēšanai un rediģēšanai (skat. 4. attēlu).

SKOS¹⁵ (*Simple Knowledge Organization System* – Vienkārša zināšanu organizēšanas sistēma) izstrādātās specifikācijas un standarti (Miles, Bechhofer 2009) atbalsta zināšanu organizēšanas sistēmu, kā tēzauru, klasifikācijas shēmu, priekšmetu sarakstu un taksonomiju, izmantošanu semantiskajā tīmeklī un nodrošina priekšmetu autoritatīvo ierakstu pārveidi atvērto saistīto datu formā. SKOS nodrošina zināšanu organizēšanas sistēmu attēlošanas standarta veidu, informācijas kodēšanā izmantojot Resursu apraksta ietvaru (RDF), lai šo informāciju savstarpēji savietojami nodotu starp datoru lietojumprogrammām.

Sākotnēji bibliotēku datu standarta MARC21 formā pieejamo Nacionālā tēzaura autoritatīvo ierakstu pārveidošana un publicēšana atvērto saistīto datu formā ir daudzpakāpju process, kas ietver:

1. MARC21 autoritatīvo ierakstu bagātināšanu – ierakstiem tiek pievienoti unikāli identifikatori, ierakstu sistēmas numuri, lai norādītu, ka ieraksts satur saites ar citiem ierakstiem datu kopā, kā arī visiem pamatjēdzieniem tiek pievienots valodu koda apzīmējums *lv*;
2. MARC21 autoritatīvo ierakstu konvertēšanu SKOS jēdzienos, izmantojot *mc2skos* rīku¹⁶, kas MARC/XML datus pārveido SKOS RDF datu formā. Līdzīgi kā LCSH, arī Nacionālā tēzaura autoritatīvo ierakstu hipersaišu norādei ar citām ārējo resursu datu kopām tiek izmantota īpašība *skos:closeMatch*;
3. SKOS datu kopu publicēšanu atvērto saistīto datu formā, izmantojot *Skosmos*, kas nodrošina daudzvalodu lietotāju saskarni jēdzienu pārlūkošanai un meklēšanai, kā arī hierarhijas vizualizēšanai. Katram ontoloģijas jēdzienam ir savs unikāls URI identifikators, ko var izmantot, lai vienozīmīgi atsauktos uz šo resursu, un visi dati ir pieejami kā atvērtie saistītie dati.

Darbs LNB ontoloģiju servisa datu kopu bagātināšanā un uzlabošanā tiks turpināts ar šādiem procesiem: 1) datu bagātināšana ar valodu kodiem un citu ārējo resursu identifikatoriem; 2) regulāra datu atjaunināšana ar jaunajiem vai rediģētajiem Nacionālā tēzaura autoritatīvajiem ierakstiem; 3) jaunizveidoto datu publicēšana.

Nākotnē LNB ontoloģiju servisa datu kopas, kas publicētas atvērto saistīto datu formā, varētu tikt izmantotas bibliotēku informācijas saistīšanai ar Latvijas muzeju un arhīvu datiem, kultūras mantojuma objektu aprakstā izmantojot ontoloģiju servisa objektu URI

14 <http://skosmos.org/>

15 <https://www.w3.org/TR/skos-reference/>

16 <https://github.com/scriptotek/mc2skos>

identifikatorus. Tās varētu saistīt arī ar citu valstu atmiņas institūciju datiem un starptautiskajām kontrolētajām vārdnīcām tīmekļa vidē.

2022. gadā tika sākti arī jaunveidoto Nacionālā tēzaura tematisko priekšmetu autoritatīvo ierakstu bagātināšana ar Somijas Finto ontoloģijas YSO (*General Finnish Ontology*) somu un angļu valodas jēdzieniem (ja nav ekvivalenta jēdziena LCSH) un identifikatoriem, kā arī atlasītas LCSH un YSO angļu jēdzienu sakritības, lai veiktu arī iepriekšveidoto autoritatīvo datu automatizēto papildināšanu ar YSO identifikatoriem. 2022. gadā tiks sākti arī ģeogrāfiskā nosaukuma un formas/žanra priekšmetu automatizēta bagātināšana ar Kongresa bibliotēkas (LCNAF, LCSH un LCGFT) un FAST identifikatoriem. 2022. gada rudenī tiek plānots arī sākt Notikuma nosaukuma priekšmetu autoritatīvo ierakstu veidošanu un to papildināšanu ar ārējo resursu jēdzieniem un hipersaitēm.

Nacionālā tēzaura autoritatīvo datu bagātināšana ar jauniem priekšmetiem un ar dažādu ārējo resursu jēdzieniem un identifikatoriem ne vien veicina ilgtspējīgu tēzaura attīstību, bet arī šo datu sadarbību un atkārtotu izmantošanu atvērto saistīto datu formā.

Nacionālā bibliogrāfija

Nacionālā bibliogrāfija ir informācijas kopums par fiziskā un digitālā formā Latvijā publicētiem izdevumiem, kas ir nacionālā kultūras mantojuma sastāvdaļa, un ar to saistīta struktūrēta, sistematizēta informācija un autoritatīvie dati – standartizētas, strukturētas, pārbaudītas un uzticamas ziņas par personām, institūcijām, vietām, notikumiem, pasākumiem, darbiem, priekšmetiem un citiem kontrolētu klasifikatoru vārdiskiem apzīmējumiem¹⁷. LNB kā nacionālās bibliogrāfijas centrs ir atbildīga par Latvijā radīto publikāciju autoritatīvu bibliogrāfisko datu nodrošināšanu un pieejamību tīmeklī. Ar mērķi palielināt datu atvērto pieejamību un izmantošanu gan citu valstu nacionālās bibliogrāfijas centriem, gan bibliotēkām un citām kopienām, t. sk. kultūras mantojuma institūcijām un izdevējiem, LNB publicē aktuālu starptautisku standartu prasībām atbilstošus, pārbaudītus un iespējami savlaicīgus nacionālās bibliogrāfijas datus. Veidojot nacionālās bibliogrāfijas datus, LNB izvērtē katra datu elementa atbilstību dažādu lietojumu vajadzībām un identificē saistīto datu pakalpojuma sniegšanai trūkstošos komponentus ne tikai aktuālajos, bet arī vēsturiskajos datos.

LNB atvērto datu portālā publicēto nacionālās bibliogrāfijas datu kopu veido strukturēti dati par saskaņā ar Obligāto eksemplāru likumu saņemtām Latvijā publicētām grāmatām, kas ir apjoma ziņā vislielākā nacionālās bibliogrāfijas daļa un kas tiek pakāpeniski rediģēta atbilstoši RDA standartam. Papildus minimālajām bibliogrāfiskajām ziņām, kas nodrošina konkrētā izdevuma atrašanu, identificēšanu, izvēli un atlasi, ierakstā tiek iekļauts detalizēts izdevumu tapšanā iesaistīto personu un institūciju uzskaitījums, ar priekšmetiem atklāts izdevumu

17 <https://likumi.lv/ta/id/62905-par-latvijas-nacionalo-biblioteku>

saturs, kas nodrošina tematisko meklēšanu, norādīti grāmatu paralēlie un oriģinālie nosaukumi tulkojumiem, izdevēji, tipogrāfijas, kā arī ziņas par līdzautoriem – tulkotājiem, ilustratoriem, sastādītājiem, priekšvārdu un pēcvārdu autoriem, fotogrāfiem, redaktoriem u. tml. Izvēlētajā nacionālās bibliogrāfijas datu kopā ietvertie dati par 2017.–2020. gadu ir labi strukturēti, kvalitatīvi un homogēni, jo pirms datu kopas publicēšanas atvērto datu portālā tie sagatavoti atbilstoši starptautiskajam datu veidošanas standartam RDA.

Portālā publicētā datu kopa satur informāciju par aptuveni 11,3 tūkstošiem grāmatu, un apjoms tiks pakāpeniski paplašināts gan ar jaunākajiem nacionālās bibliogrāfijas grāmatu daļas datiem (respektīvi, tuvākajā laikā tiks pievienoti dati arī par 2021. gadā LNB saņemtajām un reģistrētajām Latvijā publicētajām grāmatām), gan ar nacionālās bibliogrāfijas grāmatu daļas vēsturiskajiem datiem. Nacionālās bibliogrāfijas grāmatu daļas vēsturiskās datu kopas apjoms ir vairāk nekā 250 000 bibliogrāfisko ierakstu MARC21 formātā, un tas aptver laika periodu kopš pirmās zināmās latviešu valodā publicētās grāmatas 16. gadsimtā. Lai sagatavotu bibliogrāfiskos datus publicēšanai LNB atvērto datu portālā atbilstoši RDA un saistīto datu principiem, nacionālās bibliogrāfijas grāmatu daļas vēsturiskie dati tiek rediģēti un bagātināti, izmantojot automatizētus un manuālus datu labošanas paņēmienus. Publicētā datu kopa šobrīd satur tikai bibliogrāfiskos datus bez atbilstošajiem autoritātajiem datiem, kurus nākotnē paredzēts publicēt atsevišķā datu kopā.

Ilgāk nekā 20 gadus LNB sagatavo datus MARC21 formātā, kurā ir svarīgs datu elementu izkārtojums ierakstā noteiktā secībā, izmantojot īpaši kodētus laukus, apakšlaukus un indikatorus, ko veido tikai speciālistiem saprotams ciparu un zīmju kopums un ko spēj “nolasīt” slēgtas bibliotēku sistēmas. Saistīto datu standartos, t. sk. RDA, ir būtisks katrs atsevišķais informācijas apgalvojums, kuru var izmantot atkārtoti pilnīgi citā kontekstā neatkarīgi no pilna grāmatas ieraksta MARC formātā. Bibliotēku datu veidošana saskaņā ar RDA saistīto datu formā nākamās paaudzes bibliotēku datu veidošanas sistēmās balstītos uz entīšu savstarpējo attiecību un hierarhijas modeli, izmantojot autoritātos datus un RDA atsauces (*RDA Reference*) datus (dažādas entītes un to attiecību apzīmētāji).

Nemot vērā pasaules pieredzi un Starptautiskās bibliotēku asociāciju un institūciju federācijas (*International Federation of Library Associations and Institutions*) IFLA Bibliogrāfijas pastāvīgās komitejas sagatavotās vadlīnijas “Nacionālo bibliogrāfiju kopējā prakse elektroniskajā laikmetā” (*Common Practices for National Bibliographies in the Electronic Age*)¹⁸, LNB datu portālā nacionālās bibliogrāfijas dati ir publicēti (1) sākotnējā formā, bibliotēku datu standartā MARC21¹⁹ un (2) pārveidoti saistīto datu formā atbilstoši BIBFRAME datu modelim un noformēti NTriples un RDF/XML datu formātos²⁰. BIBFRAME

18 https://www.ifla.org/wp-content/uploads/2019/05/assets/bibliography/common_practices_for_national_bibliographies_2021-01.pdf

19 https://dati.lnb.lv/files/nat_bibliography-2017_2020-marc.zip

20 https://dati.lnb.lv/files/nat_bibliography-2017_2020-bibframe-ntriples.zip un https://dati.lnb.lv/files/nat_bibliography-2017_2020-bibframe-rdfxml.zip

datu modelis ir paredzēts bibliogrāfiskās informācijas izteikšanai saistīto datu formā²¹. MARC21 datu pārveidošana BIBFRAME formā veikta, izmantojot *marc2bibframe2* atvērtā pirmkoda rīku²².

Publicējot nacionālo bibliogrāfiju tīmeklī saistīto datu formā, tiek veicināta šīs informācijas tālāka izmantošana un nodrošināta iespēja šos datus saistīt ar citām datu kopām, t. sk. ar datiem dažādos digitalizācijas projektos, piemēram, *Europeana* vai LNB digitalizēto grāmatu portālā. Tos varētu izmantot arī citos ar kultūras mantojumu un izdevējdarbību saistītos risinājumos, t. sk. nododot datus mantisko tiesību kolektīvā pārvaldījuma organizācijai LATREPRO. LATREPRO nodrošina, lai autori un izdevēji saņemtu taisnīgu atlīdzību par viņu darbu reprogrāfisko reproducēšanu Latvijā – saskaņā ar Autortiesību likuma 35. pantu atlīdzību par darbu kopēšanu var saņemt jebkurš autors un izdevējs, kura radītais darbs ir piegādāts LNB kā obligātais eksemplārs²³. Šobrīd datus par iepriekšējā gadā publicētajām grāmatām LNB sniedz MARC/XML formātā. Attīstoties izdevēju organizāciju platformām un rīkiem, izdevējus un autorus pārstāvošā organizācija tos varētu saņemt modernākā, saistīto datu formā, līdz ar to iegūstot efektīvākas datu izmantošanas iespējas.

Datu kopa pētniecībai par izdevējdarbību Latvijā

LNB ir viena no statistikas iestādēm, kas iesaistīta Oficiālās statistikas programmas²⁴ izpildē. Programmā LNB atbildībā ir arī statistikas nodrošināšana par izdevējdarbību Latvijā. Izdevējdarbības statistika tiek apkopota, pamatojoties uz Obligāto eksemplāru likumu²⁵, Ministru kabineta noteikumiem Nr. 281²⁶, Statistikas likuma 4. pantu²⁷ un balstoties uz LNB krājumā saņemtajiem izdevumu obligātajiem eksemplāriem (turpmāk – OE)²⁸. Statistika par izdevējdarbību Latvijā tiek veidota un publicēta saskaņā ar Statistikas likumu²⁹, Eiropas Statistikas prakses kodeksu³⁰ un Eiropas Parlamenta un Padomes Regulu (EK)

21 <https://www.loc.gov/bibframe/docs/bibframe2-model.html>

22 <https://github.com/lcnetdev/marc2bibframe2>

23 <https://likumi.lv/ta/id/5138-autortiesibu-likums>

24 <https://www.csb.gov.lv/lv/dokumenti/oficialas-statistikas-sistema/ofici%C4%81las-statistikas-programma>

25 <https://likumi.lv/doc.php?id=136682>

26 <https://likumi.lv/ta/id/291156>

27 <https://likumi.lv/ta/id/274749-statistikas-likums>

28 <https://lnb.lv/lv/izdevejiem/obligatie-eksemplari>

29 <https://likumi.lv/ta/id/274749-statistikas-likums>

30 <https://ec.europa.eu/eurostat/documents/4031688/9332152/KS-02-18-142-LV-N.pdf/82ba14e8-a49e-4215-a9ad-8f3f0b8b1b64?t=1540214583000>

Nr. 223/2009³¹. Atšķirībā no Oficiālās statistikas portālā un ikgadējā krājumā “Latvijas izdevējdarbības statistika” (iepriekš – “Latvijas prese”) publicētajiem apkopotajiem datiem, kas raksturo izdevējdarbības nozari, datu kopā pētniecībai par izdevējdarbību Latvijā LNB ievāktie un XLSX formātā sagatavotie dati³² ir pieejami neapkopotā veidā un primāri paredzēti pētniekiem, kuri datu kopu var izmantot, veidojot konkrētiem pētnieciskajiem mērķiem atbilstošus apkopojumus.

Datu kopa ar neapkopotajiem izdevējdarbības statistikas datiem pirmo reizi tika publicēta 2019. gadā, ietverot datus par grāmatām un brošūrām no 1995. gada un citiem izdevumu veidiem no 2017. gada. Projektā datu kopa tika papildināta ar datiem par grāmatām un brošūrām 1991.–1994. gadā, digitalizējot iepriekš tikai kartīšu formātā pieejamos statistikas datus izdevumu nosaukumu līmenī. Dati datu kopā tiek atjaunoti un publicēti reizi gadā līdz 31. maijam. Datu kopa satur bibliogrāfiskus metadatus, bet nesniedz informāciju, piemēram, par pārdošanas rādītājiem un kanāliem, izdevumu patēriņu, izdevniecību darbinieku skaitu, apgrozījumu u. tml., kas, lai gan starptautiskajā praksē tradicionāli tiek izmantoti izdevējdarbības nozares raksturošanai, šobrīd LNB kā statistikas iestādei nav pieejami.

Datu kopa ietver datus par visiem Latvijā reģistrētu juridisku personu un Latvijā dzīvojošu fizisku personu kalendārā gada laikā izdotajiem (sagatavotajiem un publiskotajiem) iespaiddarbiem, elektroniskiem izdevumiem (kasešu, diskešu, disku u. tml. formātā), tiešsaistes izdevumiem un nekonvencionālajiem jeb “pelēkajiem” izdevumiem (izdevumiem, kas satur vispārpieejamu informāciju un tiek izplatīti bez maksas, piemēram, konferenču materiāli, valsts pasūtītu pētījumu rezultātu ziņojumi, deponēti zinātniskie darbi u. tml.), kuru OE saņēmusi LNB neatkarīgi no tā, vai izdevumiem ir vai nav piešķirts ISBN, ISMN vai ISSN. Izdevumu veidi, uz kuriem attiecas Obligāto eksemplāru likums un kurus izdevējiem ir pienākums piegādāt LNB, ir definēti Obligāto eksemplāru likuma pielikumā³³. Datu kopa satur datus par grāmatām un brošūrām, Masu informācijas līdzekļu reģistrā (turpmāk – MIL) ierakstītiem žurnāliem un laikrakstiem, citiem periodiskajiem izdevumiem, kuri atbilst laikraksta vai žurnāla pazīmēm, bet nav reģistrēti MIL, attēlizdevumiem, kartogrāfiskajiem materiāliem, nošu izdevumiem, promocijas darbu kopsavilkumiem, sīkiespieddarbiem un e-grāmatām.

Datu kopā tiek ietverti arī dati par grāmatām, kuru OE piegādāti ar viena kalendārā gada novēlošanas – pēc tam, kad pabeigta iepriekšējā pārskata perioda datu apkopošana. Par citiem novēloti piegādātiem izdevumu veidiem dati netiek apkopoti. Šobrīd datu kopā ietvertie dati par e-grāmatām nav uzskatāmi par pilnīgiem un reprezentatīviem, jo normatīvajos aktos, kas regulē OE piegādi, pastāv nepilnības, kā dēļ daļa izdevēju, t. sk. lielākie e-grāmatu izdevēji, LNB nepiegādā e-grāmatu OE, līdz ar to statistikā nav atspoguļota liela daļa e-grāmatu.

31 <https://eur-lex.europa.eu/legal-content/LV/TXT/PDF/?uri=CELEX:02009R0223-20150608&from=LV>

32 <https://lnb.lv/lv/izdevējdarbības-statistika/datu-masivs-petniecibai-par-izdevējdarbību-latvija>

33 <https://likumi.lv/doc.php?id=136682>

Oficiālās statistikas programmā veidotā izdevējdarbības statistika ietver arī ziņas par tirāžu, kas iegūtas, balstoties uz izdevēju iesniegtajiem pārskatiem par izdevējdarbību³⁴, kuros tiek sniegta ziņa par kalendārā gada izdoto iespiesto izdevumu tirāžu, kā arī jebkura izdevuma papildtirāžu. Šīs izdevēju sniegtās informācijas konfidencialitāti aizsargā Statistikas likuma 7. panta otrās daļas 8. punkts, kas statistikas iestādei nosaka pienākumu nodrošināt statistisko konfidencialitāti, 17. pants, kas nosaka datu apstrādes kārtību un to aizsardzības prasības, kā arī 19. panta pirmā daļa, kas regulē privātpersonu vai valsts institūciju identificējošu statistikas datu izplatīšanu. Tādēļ, lai nodrošinātu statistisko konfidencialitāti un neizpaustu ierobežotas pieejamības informāciju, datu kopā nav iekļauti un publiski nav pieejami dati par iespiesto izdevumu tirāžu izdevumu nosaukumu līmenī.

LNB ilgstoši strādā, lai nodrošinātu izdevējdarbības statistikas datu kvalitāti, kā arī saskaņotību un salīdzināmību gan Latvijas, gan starptautiskā mērogā. Tomēr gan datu konvertēšanas rezultātā, laika gaitā attīstoties informācijas tehnoloģijām un mainoties statistikas sagatavošanā lietotajai programmatūrai, gan ieviešot korekcijas datu atveides metodoloģijā, datu masīvā ietvertajos datos iespējams paralēls dažādu pieraksta formu lietojums, kam pētniekiem atsevišķos gadījumos jāpievērš uzmanība, kā arī nebūtiskas kļūdas, kas novērojamas, salīdzinot datus krājumos “Latvijas prese” un “Latvijas izdevējdarbības statistika” un Latvijas izdevējdarbības pētniecības datu kopā. Lai uzlabotu statistikas metodoloģiju, ieviestu metožu efektivitāti un diskutētu par statistikas veidotājiem un lietotājiem aktuālajiem jautājumiem, LNB sadarbojas ar zinātniekiem, pētniekiem un statistikas veidotājiem un lietotājiem no Latvijas un ārzemēm, piemēram Ziemeļvalstu tīklošanās grupu (*Nordic Networking Group*) un Eiropas Izdevējdarbības studiju asociāciju (*European Publishing Studies Association EuroPub*). Lai uzlabotu nacionālo datu atbilstību starptautiskiem standartiem, kas nodrošina datu salīdzināmību starptautiskā līmenī, šobrīd norit darbs pie Latvijas izdevējdarbības statistikas sagatavošanas metodoloģijas saskaņošanas ar ISO 9707 “Informācija un dokumentācija – Statistika par grāmatu, laikrakstu, periodisko izdevumu un elektronisko izdevumu izdošanu un izplatīšanu”.

Neapkopotu datu publicēšana atvērto datu kopas formātā valsts pārvaldē ir starptautiski atzīta labā prakse, jo nodrošina, ka dati pētniekiem un citiem interesentiem ir brīvi pieejami, atkārtoti izmantojami un lietojami atbilstoši individuālajām vajadzībām (Milić u. c. 2018). Tomēr, līdzīgi kā citu valstu pārvaldes iestāžu gadījumā (turpat: 55–56, 58–60), arī LNB vēl atrodas semantisko tehnoloģiju apgūšanas sākumposmā un tikai apzina iespējas padarīt izdevējdarbības datu lietotāju pieredzi “gudrāku” un bagātīgāku, atklājot arvien plašākas datu lietojuma un informācijas izguves iespējas, kā arī pilnveidojot statistikas izstrādi un padarot efektīvāku datu vākšanu, sagatavošanu un izplatīšanu.

Šobrīd projektā LNB ekspertu darba grupa strādā pie rekomendāciju sagatavošanas, lai uzlabotu izdevējdarbības statistikas datu kvalitāti, tos bagātinātu ar datiem no citām LNB datu kopām, kā arī potenciāli pilnveidotu efektivitāti iekšējai darba plūsmai statistikas datu

34 <https://lnb.lv/lv/izdevejiem/izdevejdarbibas-statistika#Kurpieejamaparskataveidlapa>

sagatavošanā. Izdevējdarbības statistikas dati tiek veidoti LNB portālā izdevējiem^{35,36}, kurā mūsdienīgā veidā nodrošināta iespēja iesniegt pārskatu par izdevējdarbību, informāciju par izdevumu tirzām, kā arī saņemt starptautiskos standartnumurus (ISBN, ISMN, ISSN) un UDK. Līdzās statistikai, balstoties uz datiem, kas saņemti, izdevējiem pieprasot kādu no starptautiskajiem standartnumuriem vai UDK, LNB portālā izdevējiem top minimālie bibliogrāfiskie apraksti, kas tiek eksportēti uz Valsts nozīmes bibliotēku kopkatalogu, vienlaikus izdevumu ierakstos LNB portālā izdevējiem fiksējot kopkataloga bibliogrāfisko ierakstu sistēmas numuru jeb identifikatoru. Līdz ar to darba grupa secinājusi, ka statistikas datus iespējams padarīt kvalitatīvākus, t. sk. tos bagātinot ar autoritatīvajiem un citiem kontrolētiem datiem. Darba grupa turpina darbu, izstrādājot rekomendācijas datu bagātināšanai un apzinot gan datu, gan tehnoloģisko risinājumu aspektus.

Nobeigums

Rakstā apskatītas četras Latvijas Nacionālās bibliotēkas datu kopas, kuras ir publicētas projektā, – saistīto datu kolekcija “Rainis un Aspazija”, Nacionālā tēzaura atvērtie dati ontoloģijas formā, Nacionālās bibliogrāfijas atvērtie dati un Latvijas izdevējdarbības statistikas dati. Publicējot minēto informāciju atvērto un saistīto datu formā, LNB radusi iespēju šo datu tālākai izmantošanai, t. sk. digitālajās humanitārajās zinātnēs. LNB plāno paplašināt LNB atvērto datu portālu ar vēl divām datu kopām, iekļaujot tajā aptuveni 5000 kultūras mantojuma jomā izmantotu terminu no RDA standarta, kā arī LNB autoritatīvo datu kopu.

Projektā “Latvijas atmiņas institūciju dati digitālajā telpā: vienojot kultūras mantojumu” izvirzītais uzdevums uzlabot LNB veidoto datu kopu un vārdnīcu kvalitāti un pārveidot tās par atkārtoti izmantojamām atvērto datu kopām, t. sk. saistīto datu kopām un ontoloģijām, ir izpildīts līdz ar četru atvērto datu kopu publicēšanu. Tomēr šim un vēl iepļānotajām atvērto datu kopām vēl ir “jāiztur” praktiskā atkārtotā lietojuma pārbaude, t. i., izmantošana pētnieku vidē un citos praktiskos risinājumos (līdzīgi kā minētā kultūras mantojuma platforma). Ņemot vērā, ka Atvērto datu portāls tika atklāts īsi pirms šī raksta tapšanas, portālā jau publicēto un nākotnē publicējamo datu kopu funkcionalitāte un izmantošanas iespējas ir tālākas izpētes objekts. Tomēr nesen LNB ontoloģiju serviss ir publicēts Latvijas atvērto datu portālā³⁷. Tāpat ontoloģija un atbilstošu entīšu informācija no pārējām LNB portāla datu kopām tiek integrēta Latvijas vienotajā kultūras mantojuma platformā (ERAF projekts “Kultūras mantojuma satura digitalizācija” (2. kārtā)). Šīs norises apstiprina, ka pārdomātam, pieredzē un reālu vajadzību izpētē balstītam semantisko jēdzienu tīklam ir potenciāls tikt veiksmīgi pārnestam tehnoloģiskā risinājumā, ko iespējams lietot starpinstitucionālā līmenī integrētas datu pārvaldības un zināšanu pārneses nodrošināšanai.

35 <https://izdevejiem.lnb.lv/>

36 <https://lnb.lv/lv/izdevejiem/izdeveju-portals>

37 <https://data.gov.lv/dati/lv/dataset/zanru-vietu-un-prieksmetu-tezauru-dati>

Paplašinot atvērto un atvērto saistīto datu apjomu, LNB pilda vairākus uzdevumus digitālās pētniecības atbalstam, t. i., piesaista pārbaudītiem datiem kontekstu no dažādiem avotiem, saista starpnozaru informāciju semantiskā līmenī un paplašina zināšanu bāzi. Tīmeklī publicēti atvērtie saistītie dati veicina demokrātiskāku piekļuvi zināšanām, tai pašā laikā nodrošinot sasaisti ar pirmavotiem un veicinot piekļuvi tiem. Projektā radītais atvērto datu portāls ir arī solis publisko datu kopīgošanas virzienā. Tā veidotāji apzinās, ka datu savietojamībai, kvalitātei un noteiktai struktūrai ir būtiska nozīme datu plašākā izmantošanā, t. sk. pētniecībā un izglītībā, no kā labumu iespējams gūt visai sabiedrībai.

- Apenīte, Mārīte, Bojārs, Uldis (2021). National Library of Latvia Subject Headings as Linked Open Data. Verborgh, Ruben u. c. (ed.). *The Semantic Web: ESWC 2021 Satellite Events*. ESWC 2021. Print ISBN 978-3-030-80417-6; pp 33–37. Online ISBN 978-3-030-80418-3. Lecture Notes in Computer Science, Vol 12739. Springer, Cham.
- Berners-Lee, Tim (2006). *Linked data*. Available: <http://www.w3.org/DesignIssues/LinkedData.html> [accessed: 27.10.2021.].
- Bizer, Christian, Heath, Tom, Berners-Lee, Tim (2011). Linked data: The story so far. *Semantic services, interoperability and web applications: emerging concepts*. IGI global, pp. 205–227.
- Bojārs, Uldis, Rašmane, Anita, Žogla, Artūrs, Bāliņa, Signe, Salna, Edgars (2018). Semantic Annotation Tool for Cultural Heritage Content. *Baltic Journal of Modern Computing*, No. 6(4), pp. 449–463.
- Bojārs, Uldis, Rašmane, Anita, Goldberga, Anita (2020). Linked data for digital humanities scholars and researchers: “Rainis and Aspazija” (RunA) collection. *5th Conference Digital Humanities in the Nordic Countries, DHN 2020*. CEUR Workshop Proceedings, No. 2865, pp. 185–194.
- Byrne, Alex (2014). Memory institutions shaping the past, present and future. *The Australian library journal*, No. 64(4), pp. 1–11.
- Gruber, Tom (2008). Ontology. Liu, Ling, Özsu, Tamer M. (eds.). *Encyclopedia of Database Systems*. Springer-Verlag.
- Hickey, Thomas B., Töves, Jenny A. (2014). Managing ambiguity in VIAF. *D-lib Magazine*, No. 20(7/8).
- Sacima of the Republic of Latvia (2010). *Sustainable Development Strategy of Latvia until 2030*. Available: https://pkc.gov.lv/sites/default/files/inline-files/LIAS_2030_en_1.pdf [accessed 27.10.2021.].
- Latvijas Nacionālā bibliotēka (2019). *Latvijas Nacionālās bibliotēkas attīstības stratēģija 2019.–2023. gadam*. Pieejams: https://www.lnb.lv/sites/default/files/media/par-lnb/institucija/LNB_strategija_LV.pdf [skatīts 27.10.2021.].
- Latvijas Nacionālā bibliotēka (2021). *Rekomendācijas Atsauču datu semantiskās sadarbības veicināšanai starp Latvijas kultūras mantojuma institūcijām*. DOI: <https://doi.org/10.52197/IBIV6839>. Pieejams: <https://dom.lndb.lv/data/obj/file/30193606.pdf> [skatīts 27.10.2021.].
- Miles, Alistair, Bechhofer, Sean (2009). *SKOS simple knowledge organization system reference*. W3C recommendation. Available: <https://www.w3.org/TR/skos-reference/> [accessed 27.10.2021.].
- Milić, Petar, Veljković, Nataša, Stojmenov, Leonid (2017). Semantic Technologies in e-government: Toward Openness and Transparency. Bolívar, Manuel Pedro Rodríguez (ed.). *Smart Technologies for Smart Governments: Transparency, Efficiency and Organizational Issues*. Cham: Springer, pp. 55–66.
- Pessach, Guy (2008). *Networked memory institutions: Social remembering, privatization and its discontents*. Cardozo Arts & Ent.
- Zhang, Junsong, Zhao, Wu, Xie, Gang, Chen, Hong (2011). Ontology based Knowledge Management System and Application. *Procedia Engineering*, No. 15, pp. 1021–1029.

Open Research Data of the National Library of Latvia

Mārīte Apenīte, Uldis Bojārs, Ansis Garda,
Anita Goldberga, Maira Kreislere,
Anita Rašmane, Elza Ungure

Keywords: linked data, open data, NLL open data portal, NLL ontology service, national thesaurus, national bibliography, official publishing statistics

This paper describes open and linked open datasets of the National Library of Latvia (NLL) that are available to researchers. This includes data from the linked digital collection *Rainis and Aspazija*, the National Thesaurus and National Bibliography datasets and the dataset of Latvia's publishing statistics. These datasets are published on NLL's open data portal <https://dati.lnb.lv/> and are available for research and reuse.

The datasets have been published as part of the project *Latvian memory institution data in the digital space: connecting cultural heritage* (No. lzp-2019/1-0365), which is funded by the Fundamental and Applied Research Programme of the Latvian Council of Science.

The paper describes the methods used for the dataset analysis, processing and transformation as well as the necessary data improvements. It also looks at the possibilities to use these datasets for cooperation with related fields (e.g. archives and museums) and at the plans for improvements and additions to the datasets and the open data portal.

Variantums kā konvertācijas izaicinājums: Georga Mancela tekstu atveide mūsdienu rakstībā

Everita Andronova, Anna Frīdenberga,
Renāte Siliņa-Piņķe, Elga Skrūzmane,
Anta Trumpa, Pēteris Vanags

Raksts tapis Valsts pētījumu programmas projektā "Humanitāro zinātņu digitālie resursi:
integrācija un attīstība" (VPP-IZM-DH-2020/1-0001).

Ievads

Mūsdienu latviešu valodas rakstība ir labi izveidota. Parasti problēmas rada tikai grafēmas <o> izmantošana trīs skaņu apzīmēšanai (/uo/, /o/ un /o:/), kā arī fonēmu /e/, /æ/ un /e:/, /æ:/ nešķīrums rakstu valodā. Savukārt pirmo rakstu pieminēkļu uztveri mūsdienu lasītājam apgrūtinā tieši to rakstība.

Lai agrākos latviešu rakstu avotus padarītu pieejamākus lietotājiem, sniedzot iespēju ne tikai lasīt un iepazīties ar faksimiliem, bet arī tos pētīt, tika izveidots Latviešu valodas seno tekstu korpus "Senie". Tas tika atklāts 2002. gadā, kļūstot par pirmo publiski pieejamo latviešu valodas speciālo korpusu, kas turpmākos gadus nodrošināja ērtu agrāko latviešu rakstu pieminēkļu pieejamību vienuviet (Milčonoka 2003b; Andronova 2007; Andronova 2020). Korpusa platformai nemainoties, nākamajos gados korpus tika papildināts ar jauniem rakstu avotiem. Šobrīd tas kalpo par pamatu "Latviešu valodas vēsturiskajai vārdnīcai" un sniedz datus valodas un citiem humanitāro zinātņu pētniekiem.

Latviešu valodas seno tekstu korpusā šobrīd jau ir apkopoti 74 latviešu 16.–18. gadsimta drukātie un rokraksta avoti (SENIE). Pirmie rakstu pieminēkļi ir tapuši reformācijas un kontrreformācijas kontekstā. Tā kā šī perioda latviešu tekstu autori bija galvenokārt vācieši un vācu valodai līdz pat 20. gadsimta sākumam bija raksturīgs gotiskais raksts (fraktūra), tad arī šie teksti ir gotiskajā rakstā. Veidojot korpusu, digitalizējot senos tekstus un pārvēršot tos elektroniskā formātā, 16.–18. gadsimta avoti tika transliterēti, proti, mūsdienu latviešu alfabēta burtiem atbilstošie gotiskā raksta burti tika aizstāti ar latīņu burtiem (antīkvu), savukārt mūsdienu latviešu alfabētā neeksistējoši, bet senajos tekstos lietoti burti tika aizstāti, izmantojot vienburta un kombinētus simbolus¹. 2017. gadā teksti paralēli tika konvertēti arī *Unicode* sistēmā², tomēr korpusa meklēšanas rīks ir piesaistīts iepriekš transliterētajiem tekstiem. Tā kā latviešu vecajā ortogrāfijā, kas pastāvēja no pirmajiem rakstu pieminēkļiem līdz 20. gadsimta pirmajai pusei, ir lietoti t. s. "čupu burti" (vairāku grafēmu kombinācijas), tad lietotājam, meklējot kādu vārdu korpusā, šobrīd ir ne tikai jāzina, kādi gotisko burtu aizstājēji simboli ir korpusā lietoti (tie visi ir atšifrēti meklēšanas rīkā), bet arī jāorientējas vecās ortogrāfijas pamatprincipos, jāzina, kādas grafēmas vai grafēmu kopas jāievada meklētājā, lai atrastu meklējamo: vārdu vai vārda daļu.

1 Transliterācijā izmantotie simboli pieejami <http://senie.korpus.lv/notations.htm>.

2 Korpusā izmantoto simbolu un to *Unicode* atbilstmju tabula pieejama <http://senie.korpus.lv/unicode/tabula.pdf>.

Lai korpusu varētu izmantot plašāks lietotāju loks, viens no korpusa modernizācijas soļiem paredz izstrādāt meklēšanas rīku vaicājumu veikšanai mūsdienu rakstībā. Korpusa modernizāciju veic LU Latviešu valodas institūta un LU Matemātikas un informātikas institūta pētnieku grupa projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (VPP-IZM-DH-2020/1-0001). Lai izveidotu šādu meklēšanas rīku, ir nepieciešams visus korpusa tekstus pārveidot, tuvinot tos mūsdienu rakstībai, bet vienlaikus saglabājot agrāko gadsimtu valodas īpatnības. Citu pētnieku pieredze liecina, ka teksta automatizētā pārveidē var izmantot jau pastāvošus valodas vārdu sarakstus, ģenerēt dažādas iespējamās grafēmu kombinācijas vai arī rakstot likumus. Darba grupa izvēlējusies iepriekš sagatavot katra teksta konvertācijas likumu kopumu, kas apkopoti konvertācijas tabulās³.

Lai gan izvirzītais uzdevums ir praktisks, darbs pie katras tabulas ir neliels zinātnisks pētījums, kas atklāj katra avota ortogrāfijas sistēmu. Projekta beigās konvertācijas vajadzībām sagatavotais konvertācijas tabulu kopums būs bagātīgs materiāls, kas, no vienas puses, apstiprinās jau Annas Bergmanes un Ainas Blinkenas monogrāfijā “Latviešu rakstības attīstība” (Bergmane, Blinkena 1986) sniegtos secinājumus ar kvantitatīviem datiem, ko bez korpusa būtu grūti izdarīt, no otras puses, jaunais materiāls var kļūt par pamatu plašākam pētījumam par latviešu tekstu ortogrāfijas veidošanos 16.–18. gadsimtā.

Jebkuras valodas ortogrāfijas veidošanās un attīstība ir unikāls process, kuru ir ietekmējuši dažādi faktori, piemēram, rakstības veidotāju dzimtā valoda, valodas prasmes līmenis, citas valodas tradīcijas attiecīgo skaņu atveidē, kas kopā arī rada bagātīgu variantu klāstu. Varianti ir raksturīga iezīme latviešu senajos rakstos. Par variantiem 17. un 18. gadsimta normu avotos rakstījusi Ina Druviete, konstatējot, ka šajos tekstos plaši pārstāvēti dažāda senuma un lokalizācijas fonētiskie varianti, morfoloģiskie varianti, īpaši dzimtes un celma formu varianti, tāpat arī afiksu varianti (Druviete 1993: 5–6). Anna Frīdenberga rakstījusi par variantumu Georga Manceļa darbos, aplūkojot vārddarināšanas variantus, kā arī dialektālos variantus viņa tekstos (Frīdenberga 2016), savukārt Everita Andronova skatījusi lietvārdu morfoloģiskos variantus tieši Manceļa sprediķu grāmatā (Milčonoka 1999) un uzsverusi, ka valodas īpatnības līdz šim netiek pētītas vienotā sistēmā un nav tikušas pārbaudītas visu tekstu materiālā (Milčonoka 2003a).

Šī raksta mērķis ir iepazīstināt ar pirmajiem rezultātiem viena autora – Georga Manceļa (*Georg Mancelius*, arī *Manzel*, 1593–1654) – tekstu konvertācijas procesā, īpašu uzmanību pievēršot seno tekstu daudzveidībai – variantiem, kas novērojami ne vien leksikas un morfoloģijas līmenī, bet arī ortogrāfijā. Šādā veidā salīdzinot viena autora darbus, kas tapuši vairāku gadu desmitu gaitā, redzama ortogrāfijas sistēmas attīstība. Konvertācijas likumu tabulās apkopotais materiāls dod plašas iespējas aprakstīt šo procesu, balstoties uz ļoti precīzu un pilnīgu materiālu.

Avoti. Manceļa darbību, un līdz ar to arī izdoto darbu rakstības iezīmes, var iedalīt vairākos periodos. Pie pirmā perioda pieder 1631. gadā izdotie darbi – Mancelis pārlabo un papildina pirmo latviešu luterāņu rokasgrāmatu, ko izdod ar nosaukumu “Lettisch Vade mecum”. Tajā

3 Sīkāk par projektu, par latviešu valodas seno tekstu transkripcijas, transliterācijas un konvertācijas un ar šo tēmu saistītajām terminoloģijas problēmām sk. Andronova u. c. 2022.

bez perikopju krājuma, Kristus ciešanu stāsta un stāsta par Jeruzalemes izpostīšanu ar atsevišķām titullapām ietilpst arī dziesmu grāmata “Lettische geistliche Lieder vnd Psalmen”, katehisms “Der kleine Catechismus”, kā arī Zīraka gudrības grāmata no Vecās Derības – “Das Haus=, Zucht= vnd Lehrbuch Jesu Syrachs”, ko tulkojis Mancelis pats. Pie otrā perioda pieskaitāmi 1637.–1638. gada darbi. Ar 1637. gadu datētas Manceļa tulkotās Zālamana pamācības – “Die Sprüche Salomonis”, kas izdotas kopā ar viņa sastādīto vārdnīcu “Lettus” 1638. gadā. Vienlaikus iznāk vārdnīcas otrā daļa “Phraseologia Lettica”, kurā ietilpst arī sarunu piemēri – t. s. “Desmit sarunas”. Pie trešā perioda pieskaitāms Manceļa nozīmīgākais darbs, ko viņš rakstījis daudzus gadus. Tā ir sprediķu grāmata trijās daļās “Lang=gewünschte Lettische Postill”, kas iznāk 1654. gadā un ir lielā mērā oriģināls teksts. Raksta piemēriem pamatā izmantoti septiņi no šiem darbiem, kas pārstāv autora dažādus rakstības posmus, – “Lettisch Vade mecum” (Manc1631_LVM), “Lettische geistliche Lieder vnd Psalmen” (Manc1631_LGL), “Das Haus=, Zucht= vnd Lehrbuch Jesu Syrachs” (Manc1631_Syr), “Die Sprüche Salomonis” (Manc1637_Sal), “Lettus” (Manc1638_L), “Desmit sarunas” (Manc1638_Run), “Lang=gewünschte Lettische Postill II” (Manc1654_LP2)⁴.

Vispirms ir aplūkoti agrākie Manceļa ortogrāfijas pētījumi. Raksta pamatdaļu veido dažu konvertācijas aspektā problemātisku viņa darbos izmantotu grafēmu lietojuma hronoloģisks apskats. Nobeigumā, balstoties uz atsevišķu grafēmu hronoloģisku izpēti, sniegts īss Manceļa rakstības attīstības iezīmju raksturojums.

Pētījums veikts, izmantojot deskriptīvi analītisko un filoloģisko metodi.

Līdzšinējie pētījumi par Manceļa ortogrāfiju

Manceli dēvē par latviešu rakstu valodas tradīcijas iedibinātāju (Augstkalns 2009 [1931]: 178; Birkerts 1927: 83), savukārt 1631. gadu, kad iznāk Manceļa “Lettisch Vade mecum”, uzskata par laiku, no kura skaitāma tagadējā latviešu rakstu valoda (Augstkalns 2009 [1930]: 96). Mancelis pirmais ķērās pie latviešu rakstu valodas reorganizācijas. Viņa pārلابotajā luterāņu baznīcas rokasgrāmatā “Lettisch Vade mecum” bija radīta jauna grafētikas un ortogrāfijas sistēma, kas bija sistemātiskāka par līdzšinējo un kas precīzāk un pilnīgāk atspoguļoja latviešu valodas fonoloģisko sistēmu (Vanags 2013: 219). Viņa iedibinātie ortogrāfijas principi latviešu valodā pastāvēja vairāk nekā 200 gadus.

Latviešu valodas pētnieki dažādos laika posmos pievērsušies Manceļa devumam, vairāk gan apcerot viņa nozīmi latviešu rakstu valodas attīstībā kopumā, arī viņa darbu leksikas, morfoloģijas, sintakses iezīmes, bet tieši par Manceļa ortogrāfiju rakstījuši Pēteris Šmits, Alvils Augstkalns, Arturs Ozols, Jānis Zēvers.

4 Visus rakstā izmantoto Seno tekstu korpusa avotu atšifrējumus sk. <http://senie.korpuss.lv/toc.jsp>.

Jāatzīmē, ka Mancelis pats dažos savu darbu priekšvārdos (visvairāk “Lettisch Vade mecum”, arī “Lettus”) plaši runā par savu valodu – izskaidro un motivē savus rakstības paņēmienus, apraksta dažu latviešu skaņu izrunu, kritizē agrāko rakstu kļūdas. Piemēram, “Lettisch Vade mecum” (1631) priekšvārdā viņš raksta, ka rokasgrāmatā šķīris platā /æ/ un šaurā /e/ lietojumu, pirmo apzīmējot ar <ä>, kā, piemēram, vārdos *fwāhtz* un *fwetze*. Viņš min – ja kādā vārdā rakstīts <ee> [t. i., mūsd. <ie>], tas jāizrunā tā, kā vācu vārdā *Seele* (piemēram, *Meeffa*, *Weeffis*, *Meers*); skaidro, ka ar <h> aiz <a, ä, e> viņš apzīmē garo patskani; /g/ Mancelis apzīmē ar <gh> un stāsta, kā tas būtu izrunājams – diezgan stipri piesitot mēli pie aukslējām, tomēr izrunai jābūt gana maigai, kā izrunājot /k/, dažos vārdos pirms patskaņiem izruna esot starp /g/ un /k/, piemēram, *ghabbals*, *Ghohß*, *Ghramata* utt. “Lettus” (1638) priekšvārdā autors raksta, ka agrāko *Kunx*, *Prabtiex*, *Sirtz*, *Tahzt* vietā būtu jāraksta *Kungs*, *Prabtiegs*, *Sirrds*, *Tahds* (ko arī vēlākajos darbos, piemēram, sprediķu grāmatā, pats ievēro).

Valodas pētnieki, aprakstot Manceļa ortogrāfiju, visbiežāk to darījuši, atsaucoties uz viņa “Lettisch Vade mecum” priekšvārdā dotajiem skaidrojumiem.

Manceļa rakstības īpatnības aprakstījis Šmits rakstā “Osobennosti jazyka latyshskogo pisatelja Mancelija” (1895). Viņš piemin, piemēram, patskaņu īsuma apzīmējumu ar sekojošā līdzskaņa dubultojumu, garo patskaņu apzīmējumu ar sekojošu burtu <h> vai to, ka sekojošais līdzskanis netiek dubultots, utt.

“Lettisch Vade mecum” pirmajam izdevumam un tā autora priekšvārdā sniegtajām norādēm par rakstību pievērsies arī Zēvers rakstā “Juŗa Manceļa “Vade mecum” un daŗi lŗdz ņim nepazŗstami apsveikumi” (1927). Manceļa ortogrāfijas iezŗmes plaŗāk aprakstŗjis Augstkalns rakstā “Veclatvieŗu rakstu apskats” (1930), pieskŗries ņai tŗmai arī darbā “Mŗsu valoda, viŗas vēsture un pŗtŗtŗji” (1934). Augstkalns, piemŗram, atzŗmŗ, ka Mancelis, atŗķŗrŗbā no agrāko rakstu autoriem, pareizi izŗķŗr patskaŗus: izŗķŗr divskani /ie/, atŗķŗr plato /æ/ un ņauro /e/, pareizi izŗķŗr patskaŗus galotnŗs, apzŗmŗ patskaŗu garumu (gan ar burtu <h>, gan citādi). Viŗŗ no-vŗro, ka Mancelis pareizi izŗķŗr arī lŗdzskaŗus – tiek atzŗmŗti mŗkstŗnātie lŗdzskaŗi, apzŗmŗti svelpeŗi, ņŗŗeŗņi un afrikātas, atŗķŗrot balsŗgos no nebalsŗgiem (Augstkalns 2009 [1930]). Protams, ne vienmŗr ir ievŗrota konsekvence rakstŗbā.

Manceļa rakstību detalizŗti aprakstŗjis Ozols grāmatā “Veclatvieŗu rakstu valoda” (Ozols 1965: 162–170), plaŗi citŗjot arī iepriekŗŗjo autoru – Augstkalna, Zŗvera, arī Luda Bŗrŗiŗa, ņmita – darbus.

Manceļa ortogrāfijas jautājumiem pieskaras arī Daina Zemzare monogrāfijā “Latvieŗu vārdnŗcas (lŗdz 1900. gadam)”, rakstot par vārdnŗcu “Lettus” (Zemzare 1961: 15–16), kā arī tos diezgan plaŗi apskatŗjis Rŗdofs Grabis “Pārskatā par 17. gadsimta latvieŗu valodas gramatikām”. Grabis uzskata – lai arī Mancelis nav sarakstŗjis tieŗi gramatiku, viŗa darbs ir pieminams ņajā pārskatā, jo tajā izveidota zināma sistŗma tā laika latvieŗu ortogrāfijā (Grabis 2006 [1955]: 87–94). Bergmanes un Blinkenas grāmatā “Latvieŗu rakstŗbas attŗstŗba” sniegts neliels Manceļa rakstŗbas raksturojums kopumā (Bergmane, Blinkena 1986: 20), kā arī katras atseviŗķas fonŗmas apskatā, kur tas bijis nepiecieŗams, pieminŗtas arī Manceļa rakstŗbas īpatnības.

Manceļa ortogrāfijas apraksti šajos pētījumos ir diezgan konspektīvi, neparādās viņa rakstības analīze vai salīdzinājums atsevišķos darbos vai dažādos laika posmos.

Atsevišķu grafēmu un grafēmu kopu lietojums Manceļa darbos – konvertācijas izaicinājums

Šajā apakšnodaļā ir aplūkots divu problemātisku, daudznozīmīgu grafēmu (<x>, <y>) un vienas problemātiskas grafēmu kopas (<tz>) lietojums Manceļa darbos, mēģinot parādīt, kā tas laika gaitā mainījies, kā arī meklējot katram gadījumam vislabāk atbilstošos konvertācijas risinājumus. Projekta vajadzībām veidojot konvertācijas tabulas, pirmkārt, tiek meklēti dažādi paņēmieni, kā pēc iespējas veiksmīgāk pārveidot tekstu mūsdienu rakstībā. Informācija par katras grafēmas lietojumu un tā izmaiņām gadu desmitu gaitā ir ortogrāfijas pētniecībai ļoti noderīgs šī darba papildu produkts.

Katras apakšnodaļas sākumā ir ievietots fragments no darba grupas izveidotās grafēmu tabulas, kurā, balstoties uz Manceļa avotu konvertācijas tabulām, atspoguļotas Manceļa darbos lietotās grafēmas, to kombinācijas un atbildes mūsdienu ortogrāfijā. Tabulās iekļauti dati no septiņiem Manceļa darbiem, kas pārstāv dažādus autora rakstības posmus, ērtības labad izmantojot saīsinātus apzīmējumus: Manc1631_LGL (LGL), Manc1631_Syr (Syr), Manc1631_LVM (LVM), Manc1637_Sal (Sal), Manc1638_L (L), Manc1638_Run (Run), Manc1654_LP2 (LP2).

Grafēmas <x> lietojums

Grafēma	Atbildes mūsdienu ortogrāfijā	Avoti	Piemēri
<x>	<ks>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>attmaxa, nahxi, Laix</i>
	<gs>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>aisleexi, czeenix, Kunx</i>

1. tabula. Grafēmas <x> lietojums Manceļa darbos.

Latīņu alfabēta grafēma <x>, kas lietota gan latīņu, gan vācu valodas tekstos, apzīmējot skaņu kopu [ks], latviešu rakstos parādās jau kopš pirmajiem pierakstiem. Tā atrodama visos agrīnajos avotos kopš 1585. gada katoļu katehisma. Grafēma lietota gan vārda vidū, gan galā, piemēram, *twyxt, apraxte, lyxma, czyxtey, maxfat, speex, Cilwhex, Köx, notix, efaax*. Tā lietota arī gadījumos, kad mūsdienu ortogrāfijā tiek rakstīts <gs> pēc morfoloģiskā vai etimoloģiskā principa, piemēram, *Droux, Kunx, greetczyx; auxxe, auxxas*. Retākos gadījumos <x> lietots, arī apzīmējot skaņu kopu [gz] vārdā *Swaixne, zwaixnes*. Arhaiskājos avotos, t. i.,

galvenokārt CC1585 un UP1587, līdztekus <x> atrodams arī rakstījums <kf>, <ks>, <gs>, piemēram, *rakſtes, atmackſatz, näckſem; cilweks, youcks, greceņeks, lycks; greciks, KVnks, nab-baks; fargs, Drougs, czeniġs*, ūneegs. Tādējādi redzams, ka šī perioda rakstos no vienas puses valda fonētiska rakstība, kas izpaužas grafēmas <x> un kopas <ks> lietojumā, bet reizēm skaidri parādās arī morfoloģiskais princips, ja tiek rakstīts <gs>. Rakstījuma <ks> gadījumā vārda galā abu principu lietojums sakrīt.

Mancelis, reformējot latviešu valodas rakstību, tai pamatā vispirms liek skaidru fonētisko principu. Tas izpaužas arī grafēmas <x> lietojumā. 1631. gada izdevumos gandrīz visos gadījumos skaņu kopa [ks] apzīmēta ar <x> neatkarīgi no vārda morfoloģiskā sastāva. <x> lietots vārda vidū, piemēram, *attmaxa, nahxi, liexma, raxtyja, Tuxneſſy; aiſleexi, peeluxi, auxxe, ee=dexees, ſprahxi*, tāpat arī vārda galā, piemēram, *ſpāhx, blehdneex, attnahx, lab-bahx, Laix; blehdiex, czeeniex, barx, Kunx, aiſleex*. Rakstījums <ks> vārda galā atrodams ļoti reti – sešas reizes apzīmējot morfoloģisku /k-s/: *Czillwāks* (Manc1631_Cat), *Czillwābks* (Manc1631_LGL), *ſiecks, teicks, Wyrobks* (2x) (Manc1631_Syr). Manc1631_Syr <ks> divreiz lietots arī morfoloģiska /g-s/ vietā verba nākotnes formās *aiſleeks, iſſbehks*. Savukārt grafēmu kopa <gs> vārda galā, reprezentējot morfoloģisko /g-s/, atrodama tikai trīsreiz: *patitzigs* (Manc1631_LVM), *petitzięgs, Raggs* (Manc1631_LGL). Rakstījums ar grafēmu kopu <kf> vai <kf> vārda vidū 1631. gada izdevumos sastopams tikai dažas reizes – *kaukſſecta, lubkſſecta* (Manc1631_LVM), *dfirkſtheles, fickſtu, nbefadeckſſy* (Manc1631_Syr).

Nākamajā periodā, ko pārstāv avoti Manc1637_Sal, kā arī Manc1638_L, Manc1638_PhL un Manc1638_Run, iezīmējas pāreja uz plašāku morfoloģiskā principa izmantojumu Mancela ortogrāfijā. Tomēr šajos darbos vēl redzamas izteiktas svārstības. Īpaši tas redzams Manc1637_Sal tekstā. Tajā <x> vārda galā vēl lietots gan morfoloģiska /k-s/, gan /g-s/ vietā, piemēram, *Ghrāx, jaux, Eenaidneex, Kahyeneex, ghuddrahx, iſſniex; ſchebliex, laimiex, barrx, dahrx, Draux, Kunx*. Vārda vidū <x> parasti apzīmē morfoloģisku /k-s/, piemēram, *aiſbaxta, maxa, Riextes*; /g-s/ tā rakstīts tikai formā *luhxi*. Citādu rakstījumu vārda vidū nav. Savukārt vārda galā jau biežāk rakstīts arī <gs>, piemēram, *dahrgs, draugs, kungs, ſaprohtęgs*.

1638. gada vārdnīcas “Lettus” daļās jau dominē morfoloģiskais rakstījums. Tā alfabētiskajā vārdnīcā Manc1638_L <x> gala zilbē morfoloģiskā /g-s/ apzīmēšanai lietots tikai 18 reižu, piemēram, *ſchebliex, =kunx, =pallax*, taču grafēmu kopa <gs> šai nozīmē lietota jau 168 reizes, piemēram, *ſchaubięgs, barrgs, Dahrgs, Kungs*. Morfoloģiskā /k-s/ apzīmēšanai joprojām paturēta grafēma <x>, piemēram, *ſlinx, Bitteneex, jaunahx, Pullx*. <ks> lietots vien sešas reizes, piemēram, *ſlincks, Bullwehrks, Muhks*. Vārda vidū dominē grafēma <x>, kas gan vienmēr lietota fonētiskas kopas [ks] pierakstam, piemēram, *kunxt, Bixes, kluxteht*. Vārdnīcas teksta specifikas dēļ verbu nākotnes formu tajā tikpat kā nav. [ks] ar <kf> vārda vidū Manc1638_L atrodams tikai dažas reizes: *Sickſpahre, tſchackſteht*. Līdzīgs grafēmas <x> lietojums un skaņu kopas [ks] pieraksts redzams arī vārdnīcas tematiskajā daļā Manc1638_PhL, kā arī pielikumā Manc1638_Run.

Mancela darbu trešo etapu veido 1654. gadā izdotās trīs sprediķu grāmatas daļas. Šai avotā morfoloģiskais rakstības princips ir lietots konsekventāk nekā iepriekšējā perioda darbos. Tā

grafēma <x> vārda galā lietota lielākoties morfoloģiskās kopas /k-s/ apzīmēšanai, piemēram, *jaux*, *Muhx*, *auxtahx*, *Deedelneex*, *attlix*, *iβniex*. Kopas /g-s/ apzīmēšanai tā izmantota tikai verbu nākotnes trešās personas formās, piemēram, *apfax*, *behx*, *dex*, *luhx*, *kleex*. Tikai vienreiz tā rakstīts lietvārds *šneex* (Manc1654_LP1). Parasti šādas formas rakstītas ar <gs>, piemēram, *škaudiegs*, *zeenīgs*, *Wallgs*, *Waigs*, *Kunggs*, *dabrgs*. Dažreiz šādi rakstītas arī verbu nākotnes formas, piemēram, *aisdeggs*, *auggs*, *iβbehgs*, *luhgs*, *diegs*. Manc1654_LP daļās dažas reizes ar <gs> kļūdaini rakstītas arī adjektīvu pārākās pakāpes formas: *baggatahgs*, *wahjahgs*, *mielahgs*, *labbags*, no kurām neviena nav minēta kļūdu labojumu sarakstos sējumu beigās. Vārda vidū Manc1654_LP dominē grafēma <x>. Ar to apzīmētas gan kopa [ks] dažādos vārdos, gan /gs/ tikai verbu formās, piemēram, *iβniext*, *maxa*, *Ballxne*, *liexma*, *tuhxtohfchi*; *luhxi*, *ruhxt*, *dexi*, *usdiext*, *beixees*, *aisleexi*. Taču paralēli retumis lietota arī grafēmu kopa <gf> vai <gʃ>, piemēram, *steigʃees*, *apfirrgʃit*, *iβ=behgʃi*. Savukārt grafēmu kopa <kf> vai <kfʃ> atrodama ļoti reti, piemēram, *kauckʃeeta*, *pallickʃi*, *fiekʃi*, *kluckʃieht*.

Konvertējot Manceļa tekstus mūsdienu rakstībā, pārbaudījums likumu izstrādes procesā ir maksimāli atšķirt, kuros gadījumos <x> aizstājams ar <gs> un kuros ar <ks>. Manc1654_LP trīs daļu grafēmas <x> konvertācijas likumu izveide ir samērā vienkārša. Pirms vispārēja grafēmiska likuma <x> → <ks> ir jāievieto individuāli leksiski likumi <šneex> → <šneegs>, <apfax> → <apfags>, <dex> → <deggs>, <luhx> → <luhgs>, <ruhxt> → <ruhgst> u. c., kas atbilstoši mūsdienu rakstībai konvertētu gan nomenu vienskaitļa nominatīva, gan visas iespējamās verbu tagadnes vai nākotnes formas.

Manceļa pirmā un otrā perioda tekstu konvertācijas likumu izstrāde ir sarežģītāka, jo grafēma <x> daudz biežāk lietota arī mūsdienu <gs> vietā. Tomēr šais teksts ir iespējams bez individuāliem leksiskiem likumiem formulēt arī vienu morfēmisku likumu, proti, adjektīviem ar izskaņu -īgs. Tā kā šo vārdu vienskaitļa nominatīva izskaņas lielākoties rakstītas <iex>, var formulēt likumu <iex#> → <iegs#>, kā arī formulēt papildu likumus, ja aiz vārda seko punkts vai semikols. Analogu formulu var uzrakstīt arī retajiem gadījumiem, kad izskaņa rakstīta <yx>, proti – <yx#> → <īgs#>. Problēma tomēr rodas ar Manceļa pirmā perioda rakstiem, jo tajos šie adjektīvi ir rakstīti arī ar izskaņu <ix>. Likumu <ix#> → <īgs#> šeit nevar izmantot, jo tas aptvertu arī tādus gadījumus kā, piemēram, *notix* 'notīks', *pallix* 'palīks', *Laix* 'laīks'. Gan adjektīvu formas ar gala <ix>, gan visas pārējās leksēmas, kurās ar <x> vārda galā vai vidū rakstīta skaņu kopa [ks], ko mūsdienu ortogrāfijā raksta ar <gs>, var konvertēt, tikai izveidojot tām leksiskus likumus.

Atsevišķa un šķietami neatrisināma ir problēma ar Manceļa grafisko kopu <auxt>, kas var apzīmēt gan 'augsts', gan 'auksts', piemēram, *Wiffseems Zillwäko=Bährneems buhß mirrt* / *jeb taß gir a u x t s* / *jeb fämms* / *jeb leels* / *jeb maß* / *jeb baggahtz* / *jeb nabbags* / *jeb Wätz* / *jeb jauns* / *weenam tha noteek* / *ka ohtram*. (Manc1654_LP1: 155); bet *Tad gir flapfch zellfch* / *tad attkal par leeku Saule speefch* / *tad gir a u x t s* / *tad gir atkal karřts* / *Laix*. (Manc1654_LP1: 519). Tā kā semantiskā atšķirība neatspoguļojas rakstībā, atšķirīga konvertācija ar grafēmiskiem, morfēmiskiem vai leksiskiem likumiem nav iespējama. Pašlaik vienīgais risinājums ir upurēt retos adjektīva *auxts* 'auksts' lietojumus par labu daudz biežāk sastopamajam *auxts* 'augsts', lietojot leksisko konvertācijas formulu <auxt> → <augst>.

Grafēmas <y> lietojums

Grafēma	Grafēmu kopas	Atbilstmes mūsdienu ortogrāfijā	Avoti	Piemēri
<y>		<i>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>gyr, by</i>
		<ī>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>Lyxmibu, Ryfchanu, Vbdeny</i>
		<j>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>krabyahß, fayāmbtz, cy</i>
		<ij>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>ry, by, byafchana</i>
	<ay>	<ai>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>bayl, tayßne, tay</i>
	<ey>	<ei>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>Eyma, Dreymans, ney</i>
	<uy>	<ui>	LGL, LVM, Syr	<i>Wuy, Puyßfcheems</i>
	<ye>	<ī>	LGL, LVM, LP2	<i>Ryekle, myeß, byestahß</i>
	<yh>	<ī>	LGL, LVM, Syr	<i>darryhß, Wybtoleems, yhßiānu</i>
	<ny>	<ņ>	LGL, LVM, Syr, Sal	<i>winyo, Skaddrinyeems, Septinyas</i>
	<ry>	<ŗ>	LGL, LVM, Syr, Sal	<i>baryoyahß, Karya, Gbaryums</i>

2. tabula. Grafēmas <y> lietojums Manceļa darbos.

Grafēma <y> Manceļa tekstos ir daudznozīmīga, un tās lietojums it īpaši Manceļa agrīno darbu ortogrāfijā nereti ir pozicionāls. Hronoloģiski aplūkojot grafēmas <y> lietojumu septiņos avotos, kas reprezentē visus trīs Manceļa rakstības posmus, atklājas, ka ne visi lietojumi ir noturīgi, daļa no tiem gadu gaitā ir izzudusi, ortogrāfijai kļūstot konsekventākam un grafēmām – mazāk daudznozīmīgām. Tomēr salīdzinoši liela grafēmas <y> daudznozīmība saglabājas visos Manceļa rakstības periodos.

Visos septiņos avotos pirms patskaņiem vārda vidū vai vārda beigās <y> apzīmē fonēmu /j/, piemēram, *Wahyibu, krabyahß* (Manc1631_LGL), *fayāmbtz, cy* (Manc1637_Sal), *Kriyis, Riyeis* (Manc1638_L), *Eya, apfimey* (Manc1654_LP2). Abos senākajos avotos – Manc1631_LVM un Manc1631_LGL – šajā pozīcijā ir arī izņēmumi, kad <y> savienojumā ar grafēmu <e> apzīmē fonēmu /i:/, piemēram, *Ryekle, byestahß, myeß, Ryefchana*. Ja otrā Manceļa rakstības posma avotos vairs šāds izņēmums nav sastopams, tad trešajā posmā atkal var šādu reliktu

atrast, tiesa, tikai vienu reizi: *byeftahß* (Manc1654_LP2). Tikai LVM <y> dažos gadījumos apzīmē /j/ arī vārda sākumā: *yums, yuhs*. Savukārt trešā perioda rakstos <y> kā /j/ apzīmējums lietots daudz retāk.

Pozīcijā starp skaneni /r/, /n/ un patskani vai divskani grafēma <y> apzīmē līdzskaņa mikstinājumu tikai četros senākajos no aplūkotajiem avotiem, piemēram, *winyo, barryohteeß* (Manc1631_LGL), *Karya* (Manc1631_LVM), *Skaddrinyeems, baryoyahß* (Manc1637_Sal). Vēlāko gadu Manceļa avotos grafēma <y> kā līdzskaņa mikstinātāja vairs netiek izmantota.

Pozicionālā pieeja redzama arī fonēmas /i:/ atspoguļošanā ar grafēmu <y>: senākajā Manceļa rakstības posmā un avotā Manc1637_Sal grafēma <y> apzīmē fonēmu /i:/ pirms līdzskaņiem, kā arī vārda beigās, parasti lokatīva galotnēs, piemēram, *Lyxmibu, byfteeßtees, fchinny* (Manc1631_LGL), *byfteeß, Vdeny* (Mancelis1637_Sal). Avotā Manc1638_L vārda vidū vairs ar šo grafēmu fonēma /i:/ netiek apzīmēta, tikai vārda beigās: *tanny, Atzy*, savukārt rakstības trešajā posmā atkal tiek apzīmēta fonēma /i:/ gan vārda vidū, gan beigās: *Tyrumu, Ryfchanu, Vhdeny* (Manc1654_LP2).

Senākajos avotos fonēma /i:/ ir apzīmēta arī ar grafēmu kopu <yh>, piemēram: *darryhß* (Manc1631_LVM), turklāt garīgajās dziesmās tā ir sastopama arī vārda sākumā, piemēram, *yhfītānu* (Manc1631_LGL). Nevienā no vēlākajiem Manceļa avotiem grafēmu kopa <yh> vispār nav sastopama un grafēma <y> vārda sākumā nav iespējama.

Īpaši pirmā perioda avotos ir gadījumi, kad ar grafēmu <y> apzīmēta fonēma /i/, piemēram, *Bährny, Baggaty* (Manc1631_Syr), *gyr* (Manc1637_Sal). Visos analizētajos avotos grafēma <y> lietota prievārdā *py* ‘pie’. Jānis Endzelīns savā “Latviešu valodas gramatikā” minējis, ka latviešu senajos tekstos ir lietots prievārds *pi*, kas senajā rakstībā pierakstīts kā *py* (Endzelīns 1951: 678–680). Tā kā Manceļa avotos citos gadījumos grafēma <y> fonēmu /i/ tikpat kā neapzīmē, tad tomēr varētu apsvērt iespēju, ka arī šajā vārdā Mancelis ar grafēmu <y> ir apzīmējis garu patskani⁵.

Visos Manceļa darbos grafēma <y> iekļauta arī divskanī /ai/, piemēram: *tay* (Manc1631_LGL), *tayßne* (Manc1637_Sal), *way* (Manc1638_L, Manc1654_LP2) un divskanī /ei/, piemēram, *ney* (Manc1631_LGL, Manc1637_Sal), *Dreymans* (Manc1638_L), *Eyma* (Manc1654_LP2). Tomēr divskanī /ui/ šī grafēma ir lietota tikai pirmā posma avotos – Manc1631_LVM, Manc1631_LGL un Manc1637_Syr un arī tajos ļoti reti: *Wuy, Puyffcheems*. Otrā un trešā Manceļa rakstības posma avotos divskanis /ui/ ir apzīmēts tikai ar grafēmu kopu <ui>, piemēram, *dabbuit, Muifschneex*. Šo periodu rakstos <y> arī divskaņos /ai/, /ei/ lietots salīdzinoši ļoti reti.

Visos aplūkotajos avotos grafēma <y> dažreiz apzīmē skaņu kopu /ij/, piemēram, *by* (Manc1631_LGL), *ry* (Mancelis1637_Sal, Manc1654_LP2), *byafschana* (Manc1638_L).

5 Tiesa, konvertējot Manceļa tekstus mūsdienu rakstībā, darba grupa, ievērojot latviešu valodniecības tradīciju, *py* konvertē par *pi*.

No šī apskata var secināt, ka visdaudzveidīgāk grafēma <y> izmantota visos trīs vissenākajos Manceļa avotos – Manc1631_LVM, Manc1631_LGL un Manc1631_Syr. Grafēmas <y> izmantojums šajos avotos daudzos gadījumos sakrīt un norāda uz ļoti līdzīgu ortogrāfiju. Ar katru nākamā avotu grafēmas daudznozīmība samazinās. Lai gan abos 1637. un 1638. gada avotos – Manc1637_Sal un Manc1638_L – rakstība ir līdzīga, tomēr, tāpat kā grafēmas <x> gadījumā, ir dažas pazīmes, kas avotu Manc1637_Sal saista ar pirmā rakstības perioda avotiem, piemēram, fonēmu /ŋ/ un /t/ apzīmēšana ar grafēmu kopām <ny> un <ry>. Interesanti, ka arī pašā jaunākajā avotā – Manc1654_LP2 – ir dažas parādības, kas ir senākos avotos, bet nav 1637. un 1638. gada avotos, piemēram, fonēmas /i:/ apzīmēšana ar grafēmu kopu <ye>.

Grafēma <y> mūsdienu latviešu ortogrāfijā netiek izmantota, tādēļ, konvertējot senos tekstus, tā noteikti pret kaut ko tiek nomainīta. Tā kā šī ir daudznozīmīga grafēma ar nereti pozicionālu lietojumu, tad to, protams, nav iespējams visā tekstā nomainīt, izmantojot vienu vai divus konvertācijas likumus. Konvertācijas ziņā šī grafēma ir problemātiska visu trīs Manceļa ortogrāfijas periodu avotos. Lai to veiksmīgi konvertētu, agrīnā perioda avotiem ir nepieciešami vairāki desmiti konvertācijas likumu. Piemēram, avotā Manc1631_LGL grafēmas <y> un ar to saistīto kombināciju pārveidei ir nepieciešami 40 likumi. Tā kā šīs grafēmas daudznozīmība ar katru avotu mazinās, tad arī tās pārveidei nepieciešamo konvertācijas likumu skaits ar katru avotu kļūst arvien mazāks (piemēram, Manc1638_L – 12 likumu, Manc1654_LP2 – vairs tikai 8 likumi), tomēr kopumā pieeja šīs grafēmas konvertācijai ir līdzīga visu Manceļa avotu konvertācijas tabulās.

Sākotnēji konvertē, izmantojot leksiskos likumus, piemēram, visi <py#> → <pi#>, pēc tam visi <by#> → <bi#>. Kā jau minēts, grafēmas <y> lietojums nereti ir pozicionāls, tādēļ leksiskie pārveides likumi šeit nepieciešami samērā maz, daudz ko iespējams konvertēt, izmantojot morfēmiskos un grafēmiskos pārveides likumus. Piemēram, tā kā parasti <y> pirms patskaņiem apzīmē fonēmu /j/, tad šajā gadījumā iespējams pārvērst visas šīs kombinācijas: <yu> → <ju>, <ye> → <je>. Pirms tam gan jāpārlicinās, vai nav tādu gadījumu, kuros pozīcijā pirms, piemēram, <yu>, ir grafēmas <n> vai <r>, jo tādā gadījumā vispirms jāraksta likums <ryu> → <ŕu> vai <nyu> → <ņu>. Tāpat pirms <yu> nomainās pret <ju> jāpārlicinās, vai avotā nav Manceļa agrīnajos avotos reizēm sastopamā grafēmu kopa <ye> ar fonoloģisko vērtību /i:/. Kad visi šie izņēmumi ir konvertēti, jākonvertē visi gadījumi, kad pēc <ye> vai <yu> seko grafēma <h>, kas norāda uz patskaņa garumu: <yeh> → <jē>, <yuh> → <jū>. Atlikušos <ye> un <yu> pēc tam iespējams konvertēt par <je> un <ju>.

Tā kā pozīcijā pēc patskaņa grafēma <y> var būt gan divskaņa komponents, gan arī apzīmēt fonēmu /j/, tad vispirms, izmantojot leksiskos likumus, jākonvertē vārdi vai vārdu daļas, kuros <y> ir divskaņa komponents, piemēram, <tay> → <tai>, <way> → <vai>, <bayl> → <bail>, <ney> → <nei> un tikai pēc tam visus <uy> → <uj>, <ey> → <ej>, <ay> → <aj>. Visbeidzot var visus atlikušos <y> pārvērst par <i>, pirms tam gan pārlicinoties, vai agrīnā perioda darbos nav kāda grafēmu kopa <yh> jāpārvērs par <i>.

Grafēmu kopas <tz> lietojums

Grafēmu kopa	Atbilstmes mūsdienu ortogrāfijā	Avoti	Piemēri
<tz>	<c>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>lectz, Atzcems, Titzības</i>
	<cs>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>wätz, Atz</i>
	<ts>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>autz, runnahtz, brunnohtz</i>
	<ds>	LGL, LVM, Syr, Sal, L	<i>Wahrtz, kabtz, Sirttz</i>
<ttz>	<c>	Syr	<i>Swetzz</i>
	<ts>	LGL, LVM, Syr, Sal, L, Run, LP2	<i>czittz, pattz, deßmittz</i>
<dtz>	<ds>	LGL, LVM, Syr, Sal, L	<i>Ghadtz, Passfirdtz, Radtz</i>
	<c>	LGL, LVM, Syr	<i>braudtz, pedtz</i>
	<ts>	LGL	<i>dohdtz</i>

3. tabula. Grafēmu kopas <tz> lietojums Mancelļa darbos.

Grafēmu kopa <tz> ir problemātisks gadījums, mēģinot konvertēt Mancelļa tekstus mūsdienu rakstībā. Ar grafēmu kopu <tz> tajos visur tiek apzīmēta fonēma /c/, Mancelis šajā gadījumā izmanto fonētisko rakstības principu. Arī agrākajos latviešu rakstos fonēma /c/ lielākoties apzīmēta ar dažādām burtkopām, tajā skaitā <tz>, piemēram, *tapetz* (CC1585), *praatz* (EvEp1587).

Tādējādi grafēmu kopa <tz> resp. /c/ Mancelim ir tur, kur mūsdienu latviešu valodā, rakstot pēc morfoloģiskā principa, būtu <ts>, <ds>, arī <cs> vārda beigās, kā arī, protams, <c>. Grafēmu kopa <tz> lietota visu periodu Mancelļa darbos, piemēram, *lietz* (Manc1631_LGL), *Ghohtz* (Manc1631_LVM), *Swetze* (Manc1637_Sal), *autz* (Manc1638_L), *Ammatz* (Manc1654_LP2).

Otrā un trešā perioda tekstos Mancelis daļēji pāriet uz morfoloģisko rakstību. Piemēram, viņš raksta <ts> vārda beigās – *Pirrts* (Manc1638_L), *darriets* (Manc1654_LP2), lai gan šajā pozīcijā bieži arī joprojām lietota grafēmu kopa <tz>, piemēram, *attrautz*, *mielojohtz* (Manc1654_LP2). Vārda beigās mēdz būt arī <ds>, atsevišķos gadījumos jau agrīnajos darbos, piemēram, *Badds* (Manc1631_LVM), bet vēlāk, it īpaši sākot ar Manc1638_L, – arvien

biežāk, piemēram, *Ghalds*, *Sirrds* (Manc1638_L). Jāpiezīmē gan, ka otrā perioda avotos nereti lietoti abi rakstības varianti – gan ar <tz>, gan <ds>, piemēram, *Sirrds*, *Sirtz* (Manc1637_Sal), *Wabrds*, *Wabrtz* (Manc1638_L). Savukārt trešajā periodā Mancelis pilnībā pāriet uz <ds> rakstību šajā pozīcijā, piemēram, *Ghohds*, *Wabrds*, *Tahds* (Manc1654_LP2). Līdz ar to trešajā periodā grafēmu kopai <tz> atbilst <ts>, <cs> vai <c> mūsdienu rakstībā.

Fonēmu /c/ resp. <c> grafēmu kopa <tz> var apzīmēt gan vārda vidū, piemēram, *Atzeems* (Manc1631_LVM), *nitzinahtz* (Manc1631_LGL), gan arī vārda beigās, piemēram, *pehtz* (Manc1631_Syr), *lectz* (Manc1631_LGL), *nahtz* (Manc1638_L). Divās leksēmās <tz> vietā mūsdienu ortogrāfijā būtu jāraksta <cs>, piemēram, atsevišķi rakstītos vārdos vienskaitļa nominatīvā ir *wätz* ‘vecs’ (Manc1638_Run), *Atz* ‘acs’ (Manc1654_LP2), bet atvasinājumos vai locījumos *wetz-* ‘vec-’ (*wetziga*, Manc1631_LVM), *Atz-* ‘ac-’ (*Atzehms*, Manc1654_LP2).

Visbiežāk grafēmu kopa <tz> lietota vārda beigās, kur tā mūsdienu rakstībā var nozīmēt <ts>, piemēram, *fwähtz*, *Sälltz*, *patz* (Manc1631_LVM), *kulltz*, *nofauktz* (Manc1637_Sal), bieži vien šī grafēmu kopa sastopama izskaņās -īts, -ēts, -āts, -ots, piemēram, *fwähtietz* (Manc1631_LVM), *noladähtz* (Manc1654_LP2), *runnahtz* (Manc1631_LGL), *brunnohtz* (Manc1637_Sal). Tomēr it īpaši pirmā perioda darbiem, piemēram, Manc1631_LGL, Manc1631_LVM, kā arī Manc1637_Sal, kas lielā mērā vēl saglabājuši agrāko rakstu iezīmes, pārveides procesu sarežģī tas, ka <tz> vārda beigās bieži vien atbilst arī <ds> mūsdienu rakstībā vārdos, kur vārda celms beidzas ar /d/, piemēram, *kahtz* (Manc1637_Sal), *Wabrtz* (Manc1631_LVM), kā arī *Sirtz* (Manc1638_L). Visu periodu darbos vārds var beigties arī ar <c>, piemēram, *brautz* (Manc1631_LGL), *lätz* (Manc1638_L), *nahtz* (Manc1654_LP2). Ir arī tādi gadījumi, kad <tz> vienā vārdā atbilst dažādiem burtiem vai burtkopām mūsdienu rakstībā, piemēram, *fatzitz* (Manc1637_Sal), *nitzinahtz* (Manc1631_LGL), kas vēl jo vairāk sagādā problēmas, tekstu konvertējot.

Nereti fonēmai /c/ atbilst arī grafēmu kopa <dtz> vai <ttz>, piemēram, *Pafffirtz* (Manc1631_Syr), *deßmittz* (Manc1631_LVM).

Tā kā īpaši pirmā perioda un arī otrā perioda avotiem Mancelis, izmantojot fonētisko principu, visos šajos gadījumos lieto <tz>, nešķirot <ts>, <ds>, <cs> vai <c> vārda beigās kā mūsdienu rakstībā, lai konvertētu tekstu, vispirms nākas rakstīt pārveides likumus atsevišķām leksēmām, piemēram, <Sirrdtz> → <sirds>, <Nacktz> → <nakts>, <deßmittz> → <desmits> (Manc1631_LVM), <Ghohtz> → <gods>, <teicktz> → <teikts>, <Sälltz> → <zelts> (Manc1637_Sal). Šīs leksēmas nav iespējams atrast automātiski arī daudzo rakstības variantu dēļ, piemēram, *Sirrdtz*, *Sirdtz*, *Sirtz*, *Sirdts* (Manc1631_LVM), bieži vien viens pārveides likums atbilst vienam vārdlietojumam. Šajos avotos ir tikai daži vārdi ar <tz>, kuriem tekstos ir lielāks vārdlietojumu skaits, līdz ar to tos var konvertēt visus reizē, piemēram, <pattz> → <pats> (Manc1631_LVM), <kahtz> → <kāds> (Manc1637_Sal). Automātiskā meklēšana, kas nedaudz atvieglo darbu, iespējama vienīgi divdabjiem ar izskaņām -ohtz, -ietz, -abtz, -ähtz (-ots, -īts, -āts, -ēts). Tā kā to Manceļa tekstos ir salīdzinoši daudz, iespējams izveidot konvertācijas likumus katram no šiem divdabju tipiem un, iepriekš izrakstot izņēmumus, pārveidot tos visus reizē, piemēram, <ohtz#> → <ots#>, <ietz#> → <īts#>, <ahtz#> → <āts#>, <ähtz#> → <ēts#>. Šos var pieskaitīt pie morfēmiskiem konvertācijas likumiem.

Pēc tam, kad atsevišķi izrakstītas visas leksēmas, kas beidzas ar <ts#> un <ds#>, arī <cs#>, kā arī pārveidoti iepriekš minētie divdabju tipi, iespējams realizēt grafēmisku likumu, ka visos atlikušajos gadījumos <tz> → <c>. Šādu gadījumu Manceļa tekstos arī ir gana daudz – tāpat ar vienu likumu tiek pārveidota grafēmu kopa <tz> vārdos *Titzibas, patteitzigu* (Manc1631_LVM), *Nitzinatayi, Atzis* (Manc1637_Sal) u. c.

Trešā perioda darbiem grafēmu kopas <tz> pārveidei iespējama nedaudz cita pieeja, jo nav jāņem vērā visi gadījumi, kad vārds beidzas ar <ds#>. Līdz ar to vispirms var pārveidot vārdus, kur vārda beigās <tz#> → <c#>. Pie tiem pieder daži vārdi, kuriem ir ļoti liels vārdlietojumu skaits, tāpēc tie pārveidojami ar vienu likumu, piemēram, <pehtz#> → <pēc#>, <teitz#> → <teic#> (Manc1654_LP2), bet ir arī retāki gadījumi, kuriem arī tiek izmantoti leksiskie konvertācijas likumi, piemēram, <brautz#> → <brauc#> (Manc1654_LP2) u. c. Tad ar leksisko pārveidi tiek nomainīts arī <tz#> → <cs#>, proti, vārdā <wätz#> → <vecs#> (Manc1654_LP2). Līdz ar to ir iespējams atlikušos <tz> vārda beigās nomainīt ar <ts>, proti <tz#> → <ts#>, piemēram, vārdos *apracktz, peedohtz, Prabtz* u. c. (Manc1654_LP2). Pēc tam, tāpat kā agrākajos avotos, visos atlikušajos gadījumos ir iespējams veikt pārveidi <tz> → <c>. Piemēram, vārdos *Allus=Mutzu, aitzenaja, Swetze, titzi* (Manc1654_LP2).

Dažas Manceļa rakstības attīstības iezīmes

Atsevišķu Manceļa lietoto grafēmu analīze ļauj izdarīt dažus pieņēmumus par to, kā mainījusies autora pieeja latviešu ortogrāfijai. Kopējā tendence liecina, ka no sākotnējās daudzveidības un lietojuma nekonsekvences Mancelis pakāpeniski ir virzījies uz konsekventāku ortogrāfiju. Pirmā perioda darbiem īpaši raksturīgi divi principi.

Pirmais princips ir fonētiskais, kas redzams, piemēram, grafēmu <x> un <tz> lietojumā. Tās lietotas visos vai vismaz vairumā gadījumu, kad runā tām atbilst skaņu kopa [ks] resp. afrikāta [c]. Nav ņemta vērā attiecīgo skaņu cilme vai rašanās runas plūsmā. Tādējādi grafēma <x> atrodama tais pozīcijās, kur šodien raksta gan <ks>, gan <gs>. Savukārt <tz> lietots gadījumos, kad mūsdienu rakstībā var būt <c>, <cs>, <ts> vai <ds>.

Otrs Manceļa lietots princips jau no pirmajiem avotiem ir gramatiskais, kas raksturīgs dažu galotnes zilbju rakstībā. Tā grafēma <y> visai konsekventi lietota vārda galā, apzīmējot lokatīva galotni -ī, retāk arī -īs. Negala zilbēs tā konkurē ar <ie>, <ye>, <yh>, taču ne lokatīvā. Vēl zīmīgāks piemērs ir šai rakstā neaplūkotā konsekventā vienskaitļa lokatīva galotnes -ā apzīmēšana ar <à>. Gramatiskais ortogrāfijas princips lietots arī dažu homoformu šķiršanai. Šis ir tā sauktais “loģiskais princips” (Lāme 1933), kad ar ortogrāfijas palīdzību šķirts, piemēram, adverbs *labbe* ‘labi’ no adjektīva vīriešu dzimtes daudzskaitļa nominatīva formas *labbi* ‘labi’.

Sākot ar otro periodu, Manceļa rakstībā parādās arī morfoloģiskais rakstības princips, kas nosaka maksimālu morfēmu līdzību rakstībā neatkarīgi no formas fonētiskās realizācijas. Šo

principu var redzēt grafēmas <x> aizstāšanā ar <gs> nominālo vārdšķiru vienskaitļa nominatīva formā, ja vārda celms beidzas ar /g/. Līdzīgi morfoloģiskais princips izmantots, arī rakstot <ds> iepriekšējā <tz> vietā tais vienskaitļa nominatīvos, kam vārda celms beidzas ar /d/. Tomēr konsekvents šī principa izmantojumā Mancelis tā arī neklūst, jo vēl trešajā periodā vienskaitļa nominatīva galotnē parasti raksta <tz>, nevis <ts> vai <cs>, ja vārda celms beidzas ar /t/ resp. /c/.

Nekonsekvence savu ortogrāfijas principu realizācijā, kā arī variants apzīmējumos arī ir tas, kas apgrūtina, bet reizēm padara arī neiespējamu Manceļa tekstu precīzu konvertāciju mūsdienu latviešu ortogrāfijā, tomēr, kā redzams no sniegtā materiāla, sāka teksta analīze un pārdomāti postulēti leksiskie, morfēmiskie un grafēmiskie konvertācijas likumi un to izmantošanas secība ļauj radīt tekstu (sk. paraugus 1. un 2. pielikumā), kas būs atpazīstams meklēšanas rīkos, kuros izmantojama mūsdienu ortogrāfija⁶.

Secinājumi

1. Konvertācijas likumu izstrāde Manceļa darbu pārveidei mūsdienu ortogrāfijā dod plašas iespējas izprast un aprakstīt Manceļa ortogrāfijas principus un to attīstību, balstoties uz ļoti precīzu un pilnīgu materiālu.
2. Grafēmu <x>, <y> un grafēmu kopas <tz> analīze un konvertācijas likumu izstrāde atklāj, kā mainās Manceļa ortogrāfijas principi. Sākotnēji tiek izmantoti fonētiskais un gramatiskais, vēlāk tos papildina arī morfoloģiskais princips.
3. Analīze apliecina, ka Manceļa ortogrāfijas attīstības periodi būtībā atbilst trim viņa darbu hronoloģiskajiem periodiem. Pirmo periodu (1631) raksturo fonētiskais un gramatiskais, trešo periodu (1654) – fonētiskais, gramatiskais un morfoloģiskais, bet otrais periods (1637–1638) uzskatāms par pārejas periodu, jo tajā noteiktos gadījumos redzamas svārstības starp fonētisko un morfoloģisko ortogrāfijas principu.
4. Ņemot vērā Manceļa ortogrāfijas principu maiņu un variantumu, analizēto grafēmu un grafēmu kopu pārveidei mūsdienu rakstībā var tikt izveidoti triju tipu pārveides likumi šādā secībā: 1) grafēmiskie, ar kuriem pārveido atsevišķas grafēmas vai to kopas atkarīgi vai neatkarīgi no to pozīcijas vārdā; 2) morfēmiskie, ar kuriem pārveido atsevišķas gramatiskas morfēmas (priedēkļus, piedēkļus, galotnes, izskaņas); 3) leksiskie, ar kuriem pārveido atsevišķas leksēmas vai to celmus.
5. Konvertācijas likumus lietojot teksta pārveidei, tie piemērojami pretējā secībā: 1) leksiskie; 2) morfēmiskie; 3) grafēmiskie.

6 Tekstu automātisko konvertāciju veica LU MII pētniece Lauma Pretkalniņa.

- Andronova, Everita (2007). The Corpus of Early Written Latvian: current state and future tasks. *Proceedings of Corpus Linguistics Conference 2007*. Birmingham, UK. Available: <https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/245Paper.pdf> [accessed 04.11.2021.].
- Andronova, Everita (2020). Short texts in the Corpus of early written Latvian (www.korpuss.lv/senie). *Proceedings of the 5th Conference on Digital Humanities in the Nordic Countries (DHN)*, 2020, pp. 173–183. Available: <http://ceur-ws.org/Vol-2612/short1.pdf> [accessed 04.11.2021.].
- Andronova, Everita, Fridenberg, Anna, Pretkalniņa, Lauma, Siliņa-Piņķe, Renāte, Skrūzmane, Elga, Trumpa, Anta, Vanags, Pēteris (2022). Latviešu valodas senāko rakstu pieminēkļu konvertācija mūsdienu rakstībā: iepriekšējā pieredze un automatizācijas mēģinājumi. *Aktuālas problēmas literatūras un kultūras pētniecībā*, 27. Liepāja: LiePA, 346.–358. lpp.
- Augstkalns, Alvis (2009 [1930]). Veclatviešu rakstu apskats. *Darbu izlase*. Rīga: LU Latviešu valodas institūts, 89.–129. lpp.
- Augstkalns, Alvis (2009 [1931]). Valodas jautājumi (II). *Darbu izlase*. Rīga: LU Latviešu valodas institūts, 178.–180. lpp.
- Augstkalns, Alvis (2009 [1934]). Mūsu valoda, viņas vēsture un pētnīši. *Darbu izlase*. Rīga: LU Latviešu valodas institūts, 238.–254. lpp.
- Bergmane, Anna, Blinkena, Aina (1986). *Latviešu rakstības attīstība*. Rīga: Zinātne.
- Birkerts, Antons (1927). Garīgās kultūras celmlauži. Jura Manceļa dzīve pēc dažiem mazievērotiem materiāliem. Manceļa nozīme. *Latviešu inteliģence savās cīņās un gaitās*. Rīga: A. Raņķa grāmatu tirgotavas apgāde, 68.–94. lpp.
- Druviete, Ina (1993). Variantības atspoguļojums 17. un 18. gs. normu avotos. *Baltu filoloģija*, Nr. III, 4.–9. lpp.
- Endzelīns, Jānis (1951). *Latviešu valodas gramatika*. Rīga: Latvijas Valsts izdevniecība.
- Fridenberg, Anna (2016). Variantums Georga Manceļa tekstos. *Baltu filoloģija*, Nr. XXV(2), 5.–19. lpp.
- Grabis, Rūdolfs (2006 [1955]). Pārskats par 17. gadsimta latviešu valodas gramatikām. *Darbu izlase*. Rīga: LU Latviešu valodas institūts, 82.–149. lpp.
- Lāme, Emīlija (1933). Loģiskais princips Glūka rakstībā. *Filoloģijas materiāli: profesoram J. Endzelīnam sešdesmitajā dzimšanas dienā veltīts rakstu krājums*. Rīga: Ramave, 104.–110. lpp.
- Milčonoka, Everita (1999). Daiktavardziņu morfoloģijai variantai Mancelio “Lettische Postill” (1654). *Lietuvių kalbotyros klausimai XL* (1998), psl. 46–54.
- Milčonoka, Everita (2003a). Manceļa valodas izpētes problēma. *Daugavpils Universitātes Zinātnisko rakstu krājums XIII. Valoda dažādu kultūru kontekstā*. Daugavpils: DU “Saule”, 76.–82. lpp.
- Milčonoka, Everita (2003b). Latviešu valodas 17. gadsimta teksti internetā. *Baltu filoloģija*, Nr. XII(1), 139.–150. lpp.
- Ozols, Arturs (1965). *Veclatviešu rakstu valoda*. Rīga: Liesma.
- SENIE. *Latviešu valodas seno tekstu korpuss*. Pieejams: <http://senie.korpuss.lv/toc.jsp> [skatīts 26.10.2021.].
- Šmits, Pēteris (1895). Osobennosti jazyka latyšshskogo pisatelja Mancelija. *Zhivaja starina: periodicheskoje izdanie Otdelenija etnografii Russkogo geograficheskogo obshchestva*. T. 5. S.-Peterburg: Tipografija S. N. Hudekova, str. 162–170.
- Vanags, Pēteris (2013). Latviešu rakstu valodas izveide un attīstība 16.–18. gadsimtā. *Latviešu valoda*. Prof. A. Veisberga redakcijā. Rīga: LU Akadēmiskais apgāds, 201.–240. lpp.
- Zemzare, Daina (1961). *Latviešu vārdnīcas (līdz 1900. gadam)*. Rīga: Latvijas PSR Zinātņu akadēmijas izdevniecība.
- Zēvers, J[ānis] (1927). Juŗa Manceļa “Vade mecum” un daži līdz šim nepazīstami apsvēkumi. *Izglītības Ministrijas Mēnešraksts*, Nr. 5/6, 441.–446. lpp.

1. pielikums. Fragments no Manceļa darba
“Lettische geistliche Lieder vnd Psalmen” (Manc1631_LGL: 250–251).

Teksts <i>Unicode</i> formātā	Saskaņā ar konvertācijas likumiem pārveidots teksts ⁷
1. Weens Behrnings tick flawähns Gir mums schodeen pedfimmis :/ No weenas Jumprawas tick krahšnas / Par preeku mums nabbageems ļaudeems: Ja nhe buhtu tas Behrnings mums peedfimmis / Tad buhtam mehs wifli pafudduffchi/ Ta Schälafitiba pedārr mums wiffeems / Eya tu falldais Jefu CHrift / Ka tu Czillwähx peedfimmis effi / Paffargi mums no tahs Elles. 2. Tās Laix gir no leelu Preeku baggahtz / Par Ghodu Deewa Wahrđam :/ Ka CHriftus no tahs Däbbäffa Wallftibas Wirffon Semmes gir nahzis : Tā gir leela Sāmmofchana Kattru Deews no Debbeffes mums parahda / Weens Kalps gir wings tappis / Bes wiffeems Ghrākeems mums lieds / Czaur to mehs tohpam muhfchige baggati / Nāš muhflo Ghrāko Naštu. 3. Labb tam / kas to iħtane titz / Ar wiflu Sirdi palaiſchahs :/ Tām gir ta muhfchiga dſlewofchana pattees / Labb tam/ katters wirffon to czerre / Ka CHriftus ghann gir darriyis Preekſchan mums: Tapehtz gir tas iſghayis No Deewu fawu muhfchigu Tāhwu. O brienums par brienumu / Christus nāš muhfſu Nofegummu / Vnd kluffena muhfſu Eenaidibu. 4. Nu teitz tam wiſſa Chriftita draudſiba: Par tadu leelu Dawanu :/ Vnd luhds wiņja Schälafitibu / Ka tas mums joprojam paffarga Par willtigu Mahzibu / vnd launu Eeraddumu / Eekſchan kattru mehs fenn äffām ſtahweyufchi / Tas gribbātu mums to peedoht: Deews Tāhws / Dāhls / vnd fwātais Gharrs / Mehs luhđām no tōw wiffe wairs / Laid mums eekſchen Meeru dſlewoht.	1. viens bērniņš tik slāvēns jir mums šodien pedzimis :/ no vienas jumprawas tik krāsna / par prieku mums nabagiems ļaudiems: ja ne būtu tas bērniņš mums piedzimis / tad būtam mēs visi pazuduši/ ta žēlastība peder mums visiems / eja tu saldaiss jēzu krist / ka tu cilvēks piedzimis esi / pasargi mums no tās elles. 2. tas laiks jir no lielu prieku bagāts / par godu dieva vārdam :/ ka kristus no tās debesa valstibas virson zemes jir nācis : ta jir liela zemošana katru dievs no debeses mums parāda / viens kalps jir viņš tapis / bes visiems grēkiem mums līdz / caur to mēs topam mūžīge bagati / nes mūsu grēko nastu. 3. lab tam / kas to īstane tic / ar visu sirdi palaižās :/ tam jir ta mūžīga dzīvošana paties / lab tam/ katers virson to cere / ka kristus gan jir darijis priekšan mums: tapēc jir tas izgajis no dievu savu mūžīgu tēvu. o brīnums par brīnumu / kristus nes mūsu noziegumu / und klusena mūsu ienaidību. 4. nu teic tam visa kristita draudzība: par tadu lielu dāvanu :/ und lūdz viņa žēlastību / ka tas mums joprojam pasarga par viltīgu mācību / und launu ieradumu / iekšan katru mēs sen esem stāvējuši / tas gribētu mums to piedot: dievs tēvs / dēls / und svētais gars / mēs lūdzam no tev vise vairs / laid mums iekšēn mieru dzīvot.

7 Konvertējot nav saglabāti lielie burti, jo meklēšanas rīkā lielo un mazo burtu šķīrums nebūs svarīgs.

2. pielikums. Fragments no Manceļa darba
 “Lang=gewünschte Lettische Postill III” (Manc1654_LP3: 8).

Teksts <i>Unicode</i> formātā	Saskaņā ar konvertācijas likumiem pārveidots teksts
Scheitan dīfirdam mehß / ka Deewa Wahrds mums pawehļ taß Affinainas Mohkas JEfu Chrifti muhßcham nhe aismirft / bett allafch tahß peeminneht / no tahm runnaht unnd dfeedaht / und winjam no širrds dibben par tahdu leelu Mielāftibu und Labdarrifchanu pateikt / ka wings mums no Ghrākeem / Nawes / Wällu und Elles atpēstijis / und Debbēffies eeweddis gir. Nahzeeta fchurr / Bahrni / miļi Draughi / klauffaitees / ko labb Deews py jums und py juhßas Dwehßfeles darrijiß gir. Schim Briefcham ghribbam mehß aran fcheem preekchalaffiteem Wahrdeem mahzitees / Kaß ar JEfu Chrifto Dahrfa notizis gir. O Chrište Deewa Dähls paliedfi / zaur tawahm ruhktahm Mohkahm / ka mehß töw allafch padohdamees / unnd nhekahdu Blehdibu darram / Tawu Nahwu unnd winjas Wainu mums par labb apdohmajam / par to / jeb mehß nabbaghi und wahji / töw Pateizibu fatziet ghribbam / Amen!	šeitan dzirdam mēs / ka dieva vārds mums pavēļ tas asināinas mokas jezu kristi mūžam ne aizmirst / bet allāž tās pieminēt / no tām runāt und dziedāt / und viņam no širds diben par tādu lielu mīlestību und labdarišanu pateikt / ka viņš mums no grēkiem / naves / vellu und elles atpestījis / und debesīs ievēdis jir. nācieta šur / bārni / miļi draugi / klausaities / ko lab dievs pi jums und pi jūsas dvēseles darijis jir. šim brižam gribam mēs aran šiem priekšlasītiem vārdiem mācīties / kas ar jezu kristo dārzā noticis jir. o kriste dieva dēls palīdzi / caur tavām rūgtām mokām / ka mēs tev allāž padodamies / und nekādu blēdību daram / tavu nāvu und viņas vainu mums par lab apdomajam / par to / jeb mēs nabagi und vāji / tev pateicību sacīt gribam / amen!

Variation as a Challenge for Conversion: Presenting Texts by Georg Mancelius in Modern Spelling

Everita Andronova, Anna Frīdenberga,
Renāte Siliņa-Piņķe, Elga Skrūzmane,
Anta Trumpa, Pēteris Vanags

Keywords: history of the Latvian language, spelling, phoneme,
grapheme, transcription, transliteration

Keywords: history of the Latvian language, spelling, phoneme, grapheme, transcription, transliteration The spelling of the early written sources forms a major obstacle for modern users to explore them. The Corpus of early written Latvian (www.korpuss.lv/senie) developed in 2002 covers 74 sources from the 16th–18th cc. One of the steps towards modernising the corpus is the development of a mechanism that can handle searches in modern spelling. The corpus modernisation is performed by researchers from the Institute of the Latvian language, UL and the Institute of Mathematics and Computer Science, UL. The aim of the article is to present the development of conversion rules for the sources by one particular author, G. Mancelius, which show a great variety of spelling. The data of the study are seven printings which represent different stages of the author's writings – 'Lettisch Vade mecum' (Manc1631_LVM), 'Lettische geistliche Lieder vnd Psalmen' (Manc1631_LGL), 'Das Haus=, Zucht= vnd Lehrbuch Jesu Syrachs' (Manc1631_Syr), 'Die Sprüche Salomonis' (Manc1637_Sal), the dictionary 'Lettus' (Manc1638_L), 'Ten conversations' (Manc1638_Run) and 'Lang=gewünschte Lettische Postill II' (Manc1654_LP2). The introduction to the article provides an overview of previous research on G. Mancelius' orthography, which is usually quite brief and lacks any analysis or comparison of individual sources or different phases in Mancelius' works. The study was conducted by means of the descriptive analytical and the philological method. A case study of two problematic ambiguous graphemes (<x>, <y>) and one problematic set of graphemes (<tz>) is presented, showing changes over time and seeking out the best conversion solutions for each case. The conversion was implemented according to 3 groups of laws: 1) graphemic, used for individual graphemes or sets of graphemes, either with or without regard for their position in the word; 2) morphemic, used for certain grammatical morphemes (prefixes, suffixes, flexions); 3) lexical, which convert individual lexemes or their stems. The conversion shall be performed in the following order: 1) lexical; 2) morphemic; 3) graphemic laws. On the basis of the chronological study of individual graphemes, a brief description of the development of G. Mancelius' orthography is presented in the final part.

Latgaliešu valodas korpuss citu Eiropas mazāk lietoto valodu kontekstā: korpusa raksturojums, lietojums un potenciālā iespējošana

Sanita Martena, Anna Briška,
Nikole Naua (*Nicole Nau*)

Raksts tapis Valsts pētījuma programmas “Humanitāro zinātņu digitālie resursi” projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001).

Ievads

2012. gadā tika izveidots pirmais mūsdienu latgaliešu valodas korpuss – MuLa (<http://www.korpuss.lv/id/MuLa2012>), kurā tika iekļauti latgaliešu rakstu valodā publicētie teksti. Korpusa apjoms ir 1 milj. vārdlietojumu, un korpusa tekstu lielāko daļu veido literārie un informatīvie teksti, kas tapuši laika posmā no 1988. līdz 2012. gadam. Kopš 2020. gada ir atjaunots darbs ar MuLa, atlasot korpusam tekstus, kas publicēti pēc 2012. gada, plānojot tajā vairāk iekļaut zinātnisku, populārzinātnisku un periodikas tekstu fragmentus.

Valodas korpusi sākotnēji tika izstrādāti valodnieku vajadzībām; arī mūsdienās korpusi galvenokārt tiek izmantoti lingvistiskiem pētījumiem, taču citi iemesli, kāpēc valodas korpuss ir nepieciešams, var būt pat svarīgāki. Korpuss ir izejas punkts valodas tehnoloģijām un rīkiem (piemēram, tādiem kā pareizrakstības pārbaudītāji), korpusa dati ir pamats mūsdienu vārdnīcu sagatavošanai, un korpusi tiek izmantoti, izstrādājot valodu mācību materiālus. Mazāk lietotajās valodās, tādās kā latgaliešu, korpusam ir svarīga loma valodas dokumentēšanā vēl jo vairāk tad, ja tā saturs ir speciāli sagatavots, atlasīts un tajā ir ietverti mazāk pieejami avoti un valodas kopienai nozīmīgi teksti. Tātad korpuss kalpo ne tikai pētniekiem, bet arī valodas runātāju kopienai.

Nemot vērā iepriekš minēto, raksts sniedz ieskatu mūsdienu latgaliešu valodas korpusa MuLa pilnveides un attīstīšanas procesā, īpašu uzmanību pievēršot korpusa pašreizējai un potenciālajai mērķauditorijai un salīdzinot korpusu ar citiem Eiropas mazāk lietoto valodu korpusiem.

Veidojot rakstu, tika izvirzīti trīs galvenie pētījuma jautājumi:

- kas ir raksturīgs Eiropas mazāk lietoto valodu korpusiem,
- kāds būs šobrīd papildinātais un pilnveidotais mūsdienu latgaliešu valodas korpuss, salīdzinot ar 2012. gadā plašākai sabiedrībai publicēto korpusu;
- kas ir līdzšinējais MuLa korpusa lietotājs, kādiem nolūkiem korpuss ir izmantots, un kā korpuss varētu tikt lietots nākotnē.

Iepriekš minētie pētījuma jautājumi ir noteikuši raksta struktūras izveidi. Raksta pamatā ir trīs galvenās nodaļas: pirmajā nodaļā ir dots ieskatš citu Eiropas mazāk lietoto valodu korpusu piedāvājumā, lai noteiktu situāciju ar MuLa korpusa izstrādi plašākā Eiropas kontekstā. Otrajā nodaļā ir piedāvāts īss MuLa 2012. gada versijas raksturojums un salīdzinājums ar šobrīd izstrādājamo jauno versiju. Trešās

nodaļas pamatā ir 2021. gada pavasarī veiktās aptaujas rezultāti; aptauja tika veikta ar mērķi izzināt latgaliešu valodas korpusa lietotāju profilu un viņu redzējumu korpusa uzlabošanā un lietošanā. Raksta noslēgumā ir dots pētījuma kopsavilkums un secinājumi par MuLa korpusa funkcionalitāti un izmantošanas iespējām nākotnē.

1. Mazāk lietoto valodu korpusi

Elektroniskie valodas korpusi, kuros ir liels rakstīto tekstu fragmentu skaits, kas reprezentē kādu valodu vai valodas paveidu, tiek izstrādāti jau kopš 20. gs. 60. gadiem. Paralēli ir attīstījusies arī korpuslingvistika – valodniecības nozare, kas pēta korpusu sastādīšanu un izmantošanu valodas struktūru analīzei (Sampson, McCarthy 2007; McEnery, Hardie 2012: 71). Sākotnēji pētījumi tika koncentrēti uz angļu valodu un citām valodām, kam ir liels runātāju un arī pētnieku skaits, bet kopš 90. gadu beigām tiek uzsvērtā korpusu nozīme tieši mazāk lietoto valodu dokumentēšanā un pētīšanā, kas vienlaikus nozīmē arī šo valodu attīstību (McEnery, Ostler 2002; Ostler 2009; Cox 2011; Knight et al. 2021). Tāpat arvien vairāk paplašinās un kļūst daudzveidīgāks korpusu izmantošanas lauks.

21. gs. sākumā UNESCO darba grupa, pievēršot uzmanību daudzu pasaules valodu apdraudētībai, izvirzīja 9 galvenos kritērijus valodu vitalitātes jeb dzīvotspējas pakāpes novērtēšanai. Pēc šiem kritērijiem ir iespējams noteikt valodas pašreizējo funkcionalitāti, kā arī prognozēt tās nākotni (UNESCO 2017). Starp 9 kritērijiem divi gan tiešā, gan netiešā veidā ir saistīti ar korpusu lomu valodas funkciju nodrošināšanā un paplašināšanā. Pirmais no šiem kritērijiem ir valodu dokumentēšana (*Type and Quality of Documentation*), uzsverot steidzamību valodu dokumentēšanā, lai noteiktas valodas runātāju kopiena varētu formulēt specifiskus uzdevumus valodas attīstīšanai un lai identificētais valodas materiāls (teksti un audiovizuālais materiāls) būtu pietiekami apjomīgs un daudzveidīgs, izstrādājot gramatikas izdevumus, vārdnīcas u. tml. Otrais kritērijs ir – jaunu jomu un mediju attīstīšana (*Response to New Domains and Media*), ar to saprotot valodas lietojumu gan nu jau tādos tradicionālajos medijos kā televīzija, radio, gan arī valodas daudzveidīgu lietojumu interneta vidē, tīmeklī pieejamos medijos par visdažādāko saturu un ar daudzveidīgu valodas variantu lietojumu. Valodu korpusi šajā ziņā ir neatsverams tehnoloģiju laikmeta resurss gan valodu dokumentēšanai un izpētei, gan paši var kļūt par vērtīgu krātuvi dažādu lingvistisko avotu, mācību līdzekļu un citu materiālu izstrādei.

2000. gadā tika publicēts raksts ar programmatisku nosaukumu “Korpuslingvistikas jaunā dienas kārtība: strādāt ar visām pasaules valodām” (*A new agenda for corpus linguistics: Working with all of the world’s languages*) (McEnery, Ostler 2000). Raksta autori sniedza īsu kopsavilkumu par korpusu sagatavošanu pasaules valodās 20. gadsimtā. 20. gadsimta beigās daudzām valodām joprojām vēl nebija korpusu, un autori uzskatīja, ka šī situācija ir jāmaina, jo korpusiem būs arvien svarīgāka loma valodas attīstīšanā. Rakstā tika minētas dažādas valodu kategorijas, kuru valodu korpusu izstrāde kavējas, un aplūkotas atsevišķas problēmas katrā no kategorijām (McEnery, Ostler 2000: 405–410). Tika norādīts, ka kategorijas

raksturo šādas pazīmes: runātāju skaits, valodas statuss (piemēram, valsts valoda, atzīta minoritāšu vai reģionālā valoda, dialekts), runātāju statuss (pamatiedzīvotāji/pirmiedzīvotāji vai imigranti); atsevišķa kategorija ir zīmju valodas.

Aplūkojot situāciju Eiropā, var konstatēt, ka 21. gadsimta pirmajā desmitgadē korpusi ir izstrādāti valodām, kam Eiropas mērogā ir samērā maz runātāju, piemēram, islandiešu valodai (korpusa izstrāde 2006.–2010. g., tagadējais apjoms 25 miljoni vārdlietojumu; avots: *Tagged Icelandic Corpus*) vai latviešu valodai (Levāne-Petrova 2012). Savukārt otrajā desmitgadē korpusi tiek veidoti tādām Eiropas valodām, kuras tiek iekļautas citās valodu kategorijās (Mc Enery un Ostler 2000). Tās ir oficiāli atzītas reģionālās vai minoritāšu valodas, vai arī valodas un dialekti, kam attiecīgajā valstī nav noteikta statusa. Runātāju skaita ziņā šīs valodas stipri atšķiras. Piemēram, katalāņu un galisiešu valodai, kurām Spānijā attiecīgajos reģionos ir oficiālas valodas statuss, ir vairāki miljoni runātāju, savukārt ziemeļsāmu valodā, kam Norvēģijā ir īpašs pirmiedzīvotāju valodas statuss, runā mazāk par 25 000 cilvēku, bet lejas sorbu valodā (minoritātes valoda Vācijā) ir tikai 7000 runātāju. Pie valodām, kuru runātāju skaits ir starp 100 000 un 1 miljonu, pieder basku valoda Spānijā, velsiešu valoda Lielbritānijā, rietumfrīzu valoda Nīderlandē (visām trim ir oficiāls statuss savā reģionā), kā arī silēziešu valoda Polijā un astūriešu valoda Spānijā (nevienu no abām valstīm valodām nav piešķirts kāds oficiāls valodas statuss).

1. tabulā ir dota pamatinformācija par dažu iepriekš minēto mazāk lietoto valodu tekstu korpusiem. Informācijas avoti ir šo korpusu tīmekļvietnes (sk. avotu sarakstu) un raksti: Bartels 2012 (par DTK), Kulik 2018 (par KŠM), Viejo u. c. 2008 (par ESLEMA).

Valoda (valsts)	Statuss	Korpusa nosaukums, pieejams no ... (gads)	Korpusa apjoms (vārdlietojumi)
Basku (Spānija)	Oficiālā valoda reģionā	ETC; 2013. g.	355 miljoni
Astūriešu (Spānija)	Nav statusa	ESLEMA; 2008. g.	(nav pabeigts)
Lejas sorbu (Vācija)	Minoritātes valoda	DTK; 2010. g.	23 miljoni
Velsiešu (Lielbritānija)	Oficiālā valoda reģionā	CorCenCC; 2020. g.	11 miljoni
Ziemeļsāmu (Norvēģija)	Pirmiedzīvotāju valoda	SIKOR; 2015. g.	8,9 miljoni
Silēziešu (Polija)	Nav statusa	KŠM; 2018. g.	2 miljoni
Latgaliešu (Latvija)	Valsts valodas paveids	MuLa; 2012. g.	1 miljons

1. tabula. Piemēri ar Eiropas mazāk lietotajām valodām, to statusu, korpusu izveides gadu un apjomu.

Iegūstot informāciju par Eiropas mazāk lietoto valodu korpusiem, raksta autore ir secinājusi, ka runātāju skaits nekorelē ar korpusa pieejamību vai tā apjomu. Svarīgāks ir tas, vai valodai ir finansiāls un institucionāls atbalsts. Šajā ziņā valodas ar noteiktu statusu ir labākā situācijā, jo tām bieži ir savi institūti ar pastāvīgu finansējumu, savukārt citām mazāk

lietotajām valodām un dialektiem šāda atbalsta var trūkt vai tas tiek piešķirts neregulāri (sk. Viejo u. c. 2008 par astūriešu valodu). Tikpat nozīmīgas ir arī atsevišķu pētnieku vai valodas aktivistu iniciatīvas. Tā, piemēram, kašubu valodai, kam Polijā ir reģionālās valodas statuss un valsts atbalsts, tekstu korpusi tikai tagad top (Pomierska, Stanulewicz 2019), bet oficiāli neatzītajai silēziešu valodai korpusi ir tapis, pateicoties viena cilvēka entuziasmam (Kulik 2018).

Tālāk rakstā ir aplūkotas trīs pazīmes, pēc kurām var raksturot jebkuru tekstu korpusu, tajā pašā laikā īpaši nozīmīgas tās ir tieši mazāk lietoto valodu korpusu raksturošanā: (i) ko korpusi ietver (kādu žanru, kāda laikposma, kāda stila tekstus); (ii) kādam nolūkam tas ir domāts, kādi lietojumi tikuši paredzēti, plānojot korpusu; (iii) kā tas ticis sastādīts (vairāk manuāli vai vairāk automatiski). Visi minētie jautājumi ir savstarpēji cieši saistīti.

Mazāk lietotās rakstu valodas bieži netiek lietotas visās sfērās, kurās funkcionē Eiropas valstu oficiālās jeb valsts valodas. Līdz ar to tāda korpusa sastādīšana, kas būtu gan līdzsvarots, gan reprezentatīvs, var būt īpaši problemātiska, tomēr korpusu veidotāji meklē un atrod tam dažādus risinājumus. Tāpat arī dažām Eiropas reģionālajām vai mazākumtautību valodām ir bagāta vēsture, tomēr mūsdienās tās samērā maz tiek lietotas rakstu formā. Ja galvenais korpusa mērķis ir valodas dokumentēšana un avotu, tostarp vēsturisko, bāzes veidošana, tad korpusa līdzsvarotība ir mazāk svarīga. Korpusā tiek iekļauts pēc iespējas vairāk veselu, pilna apjoma tekstu, nevis tikai to fragmenti, kā tas mēdz būt nacionālajos korposos. Piemēram, lejas sorbu korpusā (DTK, sk. 1. tabulu) plānots ietvert visus lejas sorbu valodā publicētos (drukātos) tekstus; pašlaik tajā ir iekļauti gandrīz visi teksti, kas publicēti periodā no 1848. līdz 1937. gadam, un ir iesākta jaunāku tekstu iekļaušana (Bartels 2012: 13; Bartels 2020). Savukārt Skotijas gēlu valodas korpus *Corpas na Gàidhlig* ietver 317 izvēlētos tekstus no 18. gadsimta līdz mūsdienām. Šādi korpusi var būt īpaši noderīgi kā datubāze vārdnīcu izstrādei, kā arī valodas attīstības pētniekiem; mazāk noderīgi tie ir mūsdienu valodas apguvējiem vai pētniekiem un tikai nosacīti ir lietojami valodas tehnoloģijas rīku izstrādei.

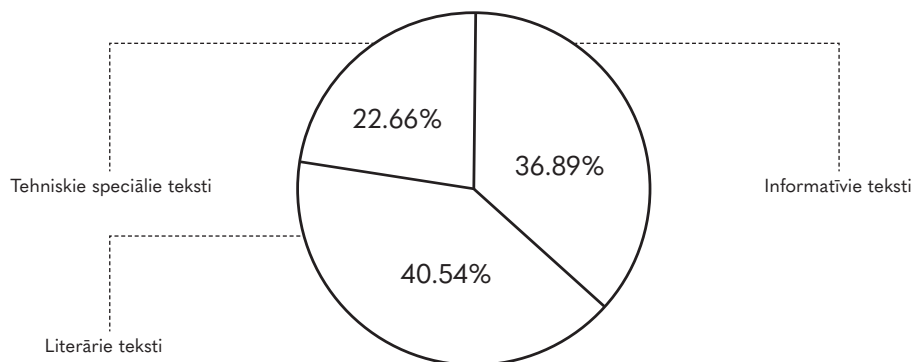
Jaunais velsiešu valodas korpus *CorCenCC* ir tapis ar mērķi atspoguļot mūsdienu velsiešu valodu niansētā un plašā lietojumā, lai korpusi kalpotu dažādām tā lietotāju grupām (pētniekiem, pašai valodas lietotāju kopienai, valodas apguvējiem un citiem velsiešu valodas interesentiem). Atkarībā no korpusa lietotāju grupas ir uzsvērti arī tā izmantošanas daudzveidīgie mērķi: valodnieciskiem pētījumiem, savas valodas izlokšņu izziņošanai, valodas mācībām, vārdnīcu un digitālo rīku izstrādei (Knight et al. 2020: 3). Šie mērķi ir noteikuši iekļaujamo tekstu izvēli. Korpusa veidotāji ir centušies sameklēt tādu paraugus, kas būtu tipiski mūsdienu velsiešu valodas lietojumam autentiskās valodas lietojuma jomās, ņemot vērā, ka tā ir bilingvāla sabiedrība, kurā velsiešu valoda funkcionē paralēli angļu valodai, bet ne visās jomās (Knight et al. 2020: 6). Korpusā tika iekļauti dažādu žanru drukātie teksti, kā arī teksti no elektroniskajiem medijiem (tīmekļvietnes, emuāri, e-pasta vēstules, izziņas); trešais korpusa satura komponents ir runas paraugu atšifrējumi (Knight u. c. 2020: 7–9). Minētā vīzija par korpusa *CorCenCC* plašo lietojumu ir redzama arī korpusa prezentācijā internetā, kas ar lietotājiem draudzīgo dizainu uzrunā gan profesionāļus, gan ikvienu valodas interesentu. Pieejami ir arī rīki valodas mācībām, bet perspektīvā ir tēzaura izveide, kas ļaus atrast noteikta semantiskā lauka vārdus un teicienus (Knight et al. 2020).

Velsiešu valodas korpuss CorCenCC noteikti kalpos par paraugu un iedvesmas avotu mazāk lietoto (tomēr dzīvo un dzīvotspējīgo) valodu korpusu izveidei nākotnē. Autori par savu pieredzi un secinājumiem ir arī publicējuši grāmatu (Knight et al. 2021). Tomēr ir skaidrs (to atzīst arī paši autori), ka tikai nedaudzas Eiropas mazāk lietotās valodas var cerēt uz tādiem finansiāliem un cilvēku resursiem, kādi vajadzīgi šādam lielam projektam. Resursu trūkuma dēļ arvien biežāk korpusi tiek sastādīti (pus)automātiski, izmantojot programmatūras, kas sameklē tekstus internetā. Šī metode izmantota arī ar vairāk lietotām un labi dokumentētām valodām, lai iegūtu lielāku daudzumu datu. Piemēram, latviešu valodas korpuss lvTenTen14, kas ir pieejams platformā *sketchengine.eu*, šobrīd ir vislielākais latviešu valodas korpuss, tam ir 530 367 474 vārdlietojumi (par TenTen sērijas korpusiem sk. Jakubíček u. c. 2013). Mazu resursu valodām (*low resource languages*) automātiskā korpusa izveide dažreiz ir vienīgais veids, kā tikt pie korpusa (Scannel 2007; Jauhiainen et al. 2020). Taču tieši šīm valodām tāda metode var radīt tehniskas problēmas (piemēram, ar automātisko valodas atpazīšanu) un būtiskus trūkumus: internetā brīvi pieejamiem tekstiem ne vienmēr ir laba kvalitāte, un tie nepietiekami atspoguļo rakstu valodas žanru un stilu daudzveidību (Jauhiainen et al. 2020; Millour, Fort 2020). Automātiski no interneta resursiem sastādītie korpusi var būt piemēroti valodas tehnoloģijas rīku izstrādei (Bernhard u. c. 2021), taču valodas attīstīšanai un apguvei, tāpat arī valodas dokumentēšanai un vārdnīcu izstrādei labāk der manuāli sagatavotie korpusi, kuros ir rūpīgi atlasīti un pārbaudīti teksti vai to fragmenti.

2. Latgaliešu valodas tekstu korpuss

Pirmo latgaliešu valodas tekstu korpusu tika sākts veidot pirms aptuveni 10 gadiem (2011–2012), kad Rēzeknes Augstskolā ar Latvijas–Lietuvas pārrobežu sadarbības programmas atbalstu tika realizēts projekts “HiPiLatLit” (“Humanitārās izglītības pētniecības infrastruktūras izveide Austrumlatvijā, Lietuvā”). Projekta laikā, sadarbojoties Rēzeknes Augstskolai, Vītauta Dižā Universitātei Kauņā un Latvijas Universitātes Matemātikas un informātikas institūtam, kā viens no rezultātiem tika izveidots Mūsdienu latgaliešu tekstu korpuss – MuLa. Korpusa apjoms bija 1 milj. vārdlietojumu, un tajā tika iekļauti Latvijā latgaliešu rakstu valodā publicēti teksti laika posmā no Atmodas (1988) līdz 2012. gadam. Tas tika izveidots kā speciālais valodas korpuss, iekļaujot tekstus noteiktās proporcijās atbilstoši latgaliešu rakstu valodas lietojuma specifikai.

Speciālā latgaliešu valodas korpusa tekstu veidu iedalījums tika aizgūts no čehu nacionālā korpusa izstrādātājiem (Čermák 2011). Informatīvo tekstu grupā tika iekļauti raksti laikrakstos, avīzēs, žurnālos, interneta portālos, blogos, kā arī ziņas un ceļojumu apraksti. Tehniskie un specifiskie teksti tika iedalīti zinātniskajos rakstos, reliģiskajos tekstos un mācību līdzekļos. Pie daīlliteratūras tekstu kopas tika pievienota latgaliski publicētā proza un dzeja. Sākotnēji tika plānots, ka tehniskie un specifiskie teksti veidos 35 % no visa korpusa apjoma, bet darba procesā tika konstatēts, ka reāli pieejamo tekstu proporcija atšķiras no plānotās, un tie bija tikai ~22 %



1. attēls. MuLa korpusa vārdlietojumi (%) dažādos teksta veidos (Briška 2013: 32).

(sk. 1. attēlu). Savukārt literāro tekstu daļa no plānotajiem 30 % tika palielināta līdz ~40 %, bet informatīvie teksti – nepilni 37 % vārdlietojumu plānoto 35 % vietā (Briška 2013: 13, 32).

Pēc 10 gadu pārtraukuma, Valsts pētījumu programmas (VPP) projekta “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (2020–2022) laikā, latgaliešu valodas tekstu korpusu MuLa tiek papildināts ar tekstiem, kas ir publicēti pēc 2012. gada. Tāpat paralēli tiek veidots pilnīgi jauns – latgaliešu mutvārdu runas korpus, kurā transkribēti Latgales iedzīvotāju un Sibīrijas latgaliešu runas ieraksti (vairāk sk. šajā izdevumā: Juško-Štekele, Kļavinska).

VPP projektā MuLa korpusa izstrādei tika izvirzīti šādi mērķi: precizēt iepriekšējo korpusa versiju, novēršot atsevišķas nepilnības un dažu tekstu dubultošanas; papildināt MuLa ar jauniem tekstiem, kas veidotu kopumā 1 milj. vārdlietojumu (palielinot korpusa apjomu līdz 2 milj. vārdlietojumu); saglabāt tekstu līdzsvarotību, apzinot visus iespējamus latgaliski publicētos dažādu žanru tekstus drukātā versijā un elektroniski pieejamos materiālus digitālajā vidē; korpusam atlasīt pēc iespējas rediģētus un publicētus (arī e-vidē pieejamus) tekstus.

Mūsdienu latgaliešu rakstu valodā, īpaši neformālās valodas lietojuma situācijās (sociālo tīklu vietnēs, privātā komunikācijā u. c.) un arī uzrakstos publiskajā telpā ir vērojama lingvistiskā varietāte. Tā kā valodas lietotāji latgaliešu valodu pārsvarā ir apguvuši tās mutvārdu formā, sazinoties ģimenē, un rakstīšanas prasmi nav mācījušies (Martena, Marten, Šuplinska 2022), tad teksti tiek rakstīti nevis ievērojot ortogrāfijas noteikumus, bet pēc izrunas, intuīcijas, atbilstoši vietējai izloksnei u. tml. Domājot par korpusa ilgtermiņa izmantošanas iespējām, tika atlasīti un korpusā iekļauti tikai rediģēti un publicēti teksti.

Kā izskatās papildinātā MuLa korpusa versija, salīdzinot ar korpusu, kas tapa pirms desmit gadiem? Pēc kādiem kritērijiem un kuri teksti ir atlasīti 2022. gada versijai?

Šobrīd, salīdzinot ar situāciju pirms 10 gadiem, latgaliešu valodā vairāk ir pieejami zinātniskie un populārzinātniskie teksti, kā arī masu mediji (drukātie un digitālajā vidē). Piemēram, regulāri kopš 2012. gada sešas reizes gadā iznāk žurnāls “A12”, portālā “LaKuGa” (“Latgalīšu Kultūrys Gazeta”) kopš 2013. gada katru mēnesi tiek publicēti vidēji 45 raksti (vidējais rādītājs laika posmā no 2013. gada janvāra līdz 2021. gada oktobrim). Līdz ar to 2022. gada korpusa versijā pieaug zinātnisko un populārzinātnisko, kā arī periodikas tekstu īpatsvars.

Korpusam no periodikas ir atlasīti teksti latgaliešu valodā no tādiem laikrakstiem un žurnāliem kā “Latgales Laiks”, “A12”, no digitālajā vidē pieejamā latgaliešu kultūras ziņu portāla *lakuga.lv* un no Latvijas sabiedrisko mediju portāla *lsm.lv*.

No zinātniskajiem un populārzinātniskajiem tekstiem šobrīd visvairāk tekstu ir sagatavots no Latgales lingvoteritoriālās vārdnīcas (Šuplinska 2012), kura tika izdota 2012. gadā. Atlasot vārdnīcas tekstus, tika ievērota šķirķļu autoru un viņu dzimumu proporcionalitāte, tematika un joma (ģeogrāfija, vēsture, folklorā, valoda un literatūra u. c.), kā arī pašu šķirķļu satura daudzveidība (pilsētas, kultūras reālijas, notikumi, izdevumi, personības, dievības u. tml.). Starp šīs žanru grupas tekstiem korpusā ir iekļauti arī zinātniskie raksti, kas latgaliski publicēti tādos izdevumos kā “Via Latgalica” un “Latgalistikys kongresu materiali”. Tāpat korpusā ir iekļauts arī populārzinātnisks izdevums “Latgalīša CV”, “Latgalistikys kongresa materiali” (tēzes), LR Satversme, kas ir pieejama arī latgaliešu valodā u. c.

Mūsdienās ir pieaudzis arī daiļliteratūras klāsts latgaliešu valodā, un no tekstu pieejamības viedokļa tie ir iegūstami elektroniski, sazinoties gan ar pašiem autoriem, gan izdevniecībām (pirms 10 gadiem daudz laika tika veltīts tekstu ieskenēšanai no grāmatām, žurnāliem, laikrakstiem un jo īpaši – tekstu pārbaudīšanai). Līdz ar to atjaunotajā MuLa versijā ir redzama lielāka daiļliteratūras tekstu daudzveidība, jo ir iekļauti arī jaunu literātu teksti.

Starp korpusam atlasītajiem daiļliteratūras tekstiem ir gan teksti, kas rakstīti oriģinālvalodā (latgaliski), gan arī tie, kas pārveidoti latgaliski no latviešu literārās valodas vai tulkoti no citām valodām, piemēram, Kārļa Skalbes pasaka “Kaķīša dzirnavas”, Antuāna de Sent-Ekziperi stāsts “Mazais princis”, Lūisa Kerola darbs “Alises piedzīvojumi Brīnumzemē”. Ir ievērots arī žanriskais princips, atlasot korpusam gan dzejoļu krājumus, gan prozas tekstus, gan arī atsevišķus dramaturģijas darbus (publicēto lugu starp visiem daiļliteratūras žanriem latgaliešu valodā ir vismazāk).

Viena gada laikā kopš jaunās VPP sākuma (pamatā – 2021. gadā) papildinātajai MuLa versijai kopumā ir iegūts aptuveni 565 000 vārdlietojumu. Starp tiem ir dažādu žanru mūsdienu teksti (publicēti pēc 2012. gada) ar metadatiem: periodika ~385 000 vārdlietojumu, daiļliteratūra ~106 000 vārdlietojumu, zinātniskie un populārzinātniskie teksti ~74 000 vārdlietojumu.

3. MuLa korpusa lietotāju profils un redzējums par korpusa izmantošanu

2021. gada pavasarī šī raksta autore veica anketēšanu par latgaliešu valodas tekstu korpusa “MuLa” lietošanu. Mērķgrupas, kuras īpaši tika uzrunātas, bija baltu valodu pētnieki Latvijā un ārzemēs, studenti, RTA tīmekļvietnes apmeklētāji, projekta “HipiLatLit” laikā 2012. gadā rīkoto korpusa lietošanas mācību dalībnieki (studenti, pētnieki, skolotāji), latviešu un latgaliešu valodas skolotāji, portāla “LaKuGa” veidotāji un lasītāji. Anketu varēja aizpildīt arī ikviens cits interesents. Anketēšanas mērķis bija noskaidrot, cik lielā mērā un kādos nolūkos korpusu tiek lietots, kā arī apzināt, ar kādām grūtībām saskaras korpusa lietotāji, lai turpmāk korpusu varētu pilnveidot un paplašināt tā lietojuma iespējas.

Anketa tika publicēta kā *Google* veidlapa, un no 2020. gada 17. marta līdz 20. aprīlim tika saņemtas pavisam 214 respondentu atbildes. Tika apzināti gan tie cilvēki, kuri korpusu lieto (grupa “Zina un lieto”), gan tie, kuri zina par korpusu, bet dažādu iemeslu dēļ nelieto (grupa “Zina un nelieto”), gan tie, kuriem korpusu ir kaut kas jauns (grupa “Nezina un nelieto”). Zīmīgi, ka procentuāli visvairāk savā grupā (86 % gadījumu) latgaliešu valodu privātajā sfērā lieto tie, kuri par korpusu nav dzirdējuši, turpretī MuLa korpusa lietotāji tikai 53 % gadījumu lieto latgaliešu valodu privātajā sfērā, toties šajā grupā par to ir lielāka zinātniskā un profesionālā interese. Savukārt tie, kuriem ir pedagoģiska interese par latgaliešu valodu, procentuāli visvairāk ir tajā respondentu grupā, kuri par MuLa korpusu zina, ir dzirdējuši, bet to nelieto.

Pirmā grupa “Zina un lieto”

Starp 214 respondentiem ir tikai 15 cilvēki, kuri korpusu lieto, no tiem deviņas sievietes vecumā no 32 līdz 52 gadiem un seši vīrieši vecumā no 32 līdz 69 gadiem.

Astoņiem respondentiem no 15 ir zinātniska interese par latgaliešu valodu, astoņi lieto valodu arī privātajā jomā, seši lieto latgaliešu valodu savā darbā, bet tikai viens respondents šajā grupā ir ar pedagoģisku interesi par latgaliešu valodu. Korpusa lietotāji to izmanto, lai meklētu piemērus, kas parāda, kā var lietot kādu vārdu, lai meklētu vārdu kontekstā, lai labāk saprastu tā nozīmi vai lai pārbaudītu pareizrakstību. Četri respondenti ir lietojuši korpusu, pētot konkrētu valodas parādību zinātniskam darbam.

Lai arī kopumā korpusa saturs un izmantošanas iespējas tika vērtēti kā labi (11 respondentu), tomēr starp būtiskākajiem korpusa trūkumiem tika minēti šādi: korpusu ir par maz (7 respondenti), korpusa platformā ir pārāk maz informācijas par to, kā lietot korpusu (6 respondenti), nevar meklēt pēc gramatiskajām kategorijām (7 respondenti), nevar atrast visas viena vārda formas uzreiz (5 respondenti). Komentāros pavīdēja arī vēlme meklēt korpusā informāciju pēc noteikta teksta žanra, *lai salīdzinātu kādas vienības lietošanu dažādos žanros*.

Korpusa lietotāji anketā sniedza arī plašus komentārus par to, kas būtu uzlabojams. Apkopojot teikto, var secināt, ka ar korpusa platformu respondenti nav apmierināti divu

iemeslu dēļ: pirmkārt, platforma nav saprotama lietotājam, jo tā ir tikai angļu valodā (*tā nav latviskota (latgaliskota)*), otrkārt, nav plašāka apraksta ne par pašu korpusu, ne par tā lietošanu (*varbūt būtu nepieciešama saskarnes atvieglota (light) versija neprofesionāliem lietotājiem; citreiz īsti nesaprotu, kā mēklēt, kā tur darboties, pietrūkst pacietības, un metu mieru. Lai gan gribētu prast ar to strādāt labāk*).

Otrā grupa “Zina un nelieto”

Viena piektā daļa jeb 43 aptaujas dalībnieki ir tie, kuri par MuLa korpusu ir dzirdējuši, taču to nelieto. Galvenie informācijas kanāli, kur respondenti ir uzzinājuši par korpusu, ir skola vai augstskola (37 %), sociālie mediji un draugi, kolēģi (30 %), platforma *korpus.lv* (30 %). Pieci respondenti ir pat piedalījušies korpusu lietošanas mācībuursos, tomēr MuLa korpusu nelieto. Kā galvenie iemesli, kāpēc respondenti nelieto korpusu, ir minēti: nav vajadzības (84 %), nezina, ko īsti ar to darīt (19 %), kā arī – kā ar to rīkoties (12 %). Uz jautājumu, vai ir lietoti citi valodu korpusi, 70 % respondentu ir atbildējuši apstiprinoši (populārākie ir latviešu valodas, krievu valodas un angļu valodas korpusi).

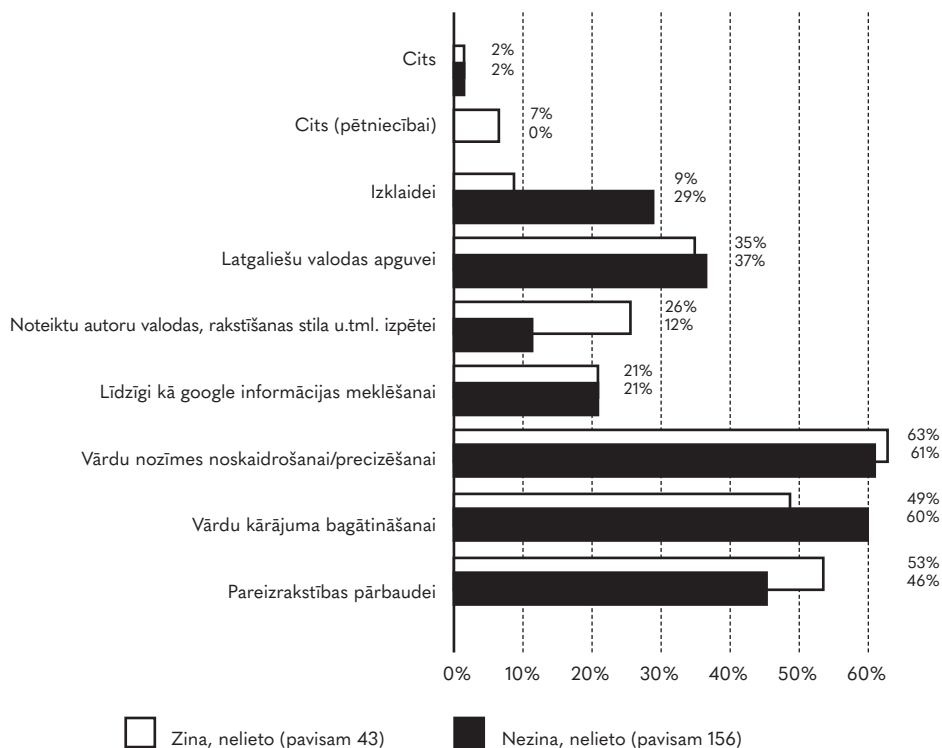
Šīs grupas respondentu komentāros (kopā 10 komentāru) tika minēts gan potenciālais latgaliešu valodas korpusa lietotājs (*korpus vajadzīgs latgaliešu valodas pētniekiem, arī lietotājiem*), gan izteiktas idejas, kas būtu darāms, lai korpusu varētu izmantot vairāk, un šie komentāri ir diezgan līdzīgi iepriekšējās – korpusa lietotāju – grupas viedokļiem. Tomēr šajā grupā vēlmes ir izteiktas precīzāk (*varētu kādu vebināru vai attālinātu semināru izveidot, popularizējot un izskaidrojot, kā praktiski to varētu lietot*), kā arī – attiecinot uz sabiedrības informēšanu kopumā (*plašāk informēt sabiedrību par korpusa izveidi un mērķi*).

Trešā grupa “Nezina un nelieto”

Trešā grupa – respondenti, kuri nav dzirdējuši par latgaliešu valodas korpusu, – bija vislielākā, pavisam 156 respondenti. Šai grupai pamatinformācija par MuLa korpusu, iespējams, tika iegūta, aizpildot anketu (tajā bija dota arī saite uz pašu korpusu, kuru anketas aizpildīšanas laikā varēja aplūkot).

Šajā grupā, salīdzinot ar abām iepriekšējām grupām, procentuāli visvairāk bija atbilžu, kurās bija norādīts, ka korpusu varētu noderēt arī izklaidei (sk. 2. attēlu). Tas ļauj izteikt pieņēmumu, ka potenciālais korpusa lietotājs sagaida arī kaut ko atraktīvu. Tie, kuri jau ir kaut ko zinājuši par korpusiem un lietojuši citu valodu korpusus, salīdzinoši vairāk ir norādījuši, ka latgaliešu tekstu korpusu varētu noderēt noteiktu autoru valodas, rakstīšanas stila izpētei, nekā tie, kuri neko nav dzirdējuši par korpusiem, attiecīgi 26 % un 12 % katrā grupā (sk. 2. attēlu). Tomēr populārākās atbildes uz jautājumu, kur latgaliešu tekstu korpusu varētu noderēt, bija trīs: vārdu nozīmes precizēšanai, vārdu krājuma bagātināšanai, pareizrakstības pārbaudei (sk. turpat). Visi trīs aspekti ir ļoti nozīmīgi jebkuras valodas apguvē, līdz ar to šis viedoklis būtu jāņem vērā, piemēram, izstrādājot uzdevumus latgaliešu valodas apguvei. Tos varētu

"Kā Jums šķiet, kā latgaliešu tekstu korpuss varētu Jums noderēt?"
(Atbildes procentuāli grupas ietvaros)



2. attēls. Otrās un trešās grupas respondentu redzējums
par latgaliešu tekstu korpusa izmantošanas iespējām.

lietot gan skolēni un skolotāji, gan arī pieaugušie valodas apguvēji, tādā veidā ļaujot prognozēt, ka potenciālais korpusa lietotājs nākotnē varētu būt arī valodas apguvējs.

Apkopojot komentāru sadaļā rakstīto šajā grupā, kurā bija visvairāk respondentu (156 no 214), var minēt trīs būtiskākos – un tieši izglītojošos – aspektus. Pirmais – anketas aizpildīšana kā sevis informēšana par MuLa korpusu un pateicība par šo iespēju (*Paldies par MuLa vietnes informēšanu! Brīnišķīga iespēja, ar kuru noteikti iepazīšos. Pirmo reizi ieraudzīju, ņemšu vērā – paldies! Brīnišķīgs korpuss, paldies par darbu!*). Otrais – ierosinājumi korpusa popularizēšanai un cilvēku izglītošanai (*Noderētu plašāks info par šo lietotni. Lai iespējami vairāk parādās par to informācijas. Tā ir ļoti laba iespēja cilvēkiem uzlabot savas latgaliešu valodas zināšanas u. c.*). Trešais aspekts ir cieši saistīts ar raksta pirmajā

nodaļā minētajiem UNESCO kritērijiem valodas dzīvotspējas nodrošināšanai, starp kuriem valodas dokumentēšana, reaģēšana uz jaunajiem medijiem un izmantošana dažādās tehnoloģijās ir īpaši nozīmīgi mazāk lietoto valodu kontekstā. Arī respondenti, komentējot MuLa korpusu, ir norādījuši uz tā nozīmi pašas valodas popularizēšanā, attīstīšanā un apguvē (*Ļoti jauki, ka Latgaliešu valoda tiek izzināta. Vai ir plānots popularizēt latgaliešu valodu? Tas noteikti palīdzētu saglabāt valodas dzīvību tālākajā nākotnē. Vēlētos, lai latgaliešu valodu neaizmirstu un mācītu arī interesentiem. Paļdis par latgališu volūdys cīneišonu, ceļšonu gaismā, jaunu asneņu deidzēšonu! Priks.*).

Nobeigums un secinājumi

Mazāk lietoto valodu korpusi ir nozīmīgi gan pašas valodas dokumentēšanā un attīstīšanā, gan ir vērtīgs resurss lingvistisko un lingvodidaktisko avotu izstrādē, vienlaicīgi arī popularizējot mazāk lietoto valodu lietojumu. MuLa korpus nedz apjoma, nedz izmantošanas iespēju ziņā nav salīdzināms ar tādiem reģionālo valodu korpusiem kā basku, velsiešu vai ziemeļsāmu valodas korpusiem. Tajā pašā laikā tas ir rūpīgi izveidots latgaliešu rakstu valodas reprezentatīvo tekstu korpus.

Apjoma ziņā MuLa korpus (2022. gada rudenī ir plānoti 2 milj. vārdlietojumu) ir salīdzināms, piemēram, ar silēziešu valodas korpusu Polijā. MuLa ir ticis un tiek veidots manuāli, atlasot dažādu žanru tekstus un rūpējoties par žanru un tekstu daudzveidību, kā arī par valodas kvalitāti. Korpusa lielākais trūkums ir – tas nav lemmatizēts un nav morfoloģiski marķēts, t. i., kāda vārda formas nevar atrast ar vienu vaicājumu, bet jāmeklē katra forma atsevišķi.

MuLa korpus jau gandrīz 10 gadu ir pieejams jebkurai interesentam. Tomēr aptaujas dati rāda, ka līdz šim tas ir ticis lietots ļoti maz (galvenokārt – zinātniskiem un profesionāliem nolūkiem), taču valodas runātāju kopienā par to ir interese. Iepazīstinot ar korpusu un tā iespējām skolotāju semināros (2022. gada februārī un jūnijā), Latgales kongresā (2022. gada aprīlī) u. c. pasākumos, tiek uzdoti jautājumi un ierosinājumi, kas liecina par nepieciešamību popularizēt korpusu vairāk un mērķtiecīgāk. Līdz šim korpusa lietotāji to ir izmantojuši vārda nozīmes noskaidrošanai un savas izpratnes padziļināšanai, aplūkojot vārda lietojuma kontekstu, kā arī pareizrakstības pārbaudei. Veicot lingvistiskos pētījumus, korpus tiek izmantots latgaliešu valodas leksikas vai gramatikas izzināšanai, taču šādu pētnieku un pētījumu ir ļoti maz.

Respondentu atbildes ļauj ieskicēt arī MuLa korpusa trūkumus, kuru novēršana potenciāli varētu paplašināt korpusa lietotāju skaitu. Lietotājam draudzīgāks korpus, kā rāda aptaujas dati, būtu, ja tiktu veikti gan saturiski, gan tehniski uzlabojumi. Satura ziņā korpus būtu jāpaplašina kvantitatīvi (tas šobrīd jau tiek darīts), tāpat būtu jāuzlabo, jāvariē meklēšanas iespējas korpusā (piemēram, pēc vārda pamatformas (lemmas), pēc morfoloģiskām pazīmēm, pēc tekstu žanriem). Gan satura, gan tehnisko uzlabojumu ziņā jādodomā par korpusa vietnes aprakstu un lietošanas pamācībām latviešu un/vai latgaliešu valodā. Jāpiedāvā vienkāršas, bet

precīzas instrukcijas, piemēram, video formātā, kā rīkoties ar korpusu, kā arī jāsniedz pamata informācija par pašu korpusu.

Aptaujas dati atklāj tieši reģionālo valodu un līdz ar to arī korpusu izmantošanas specifiku: reģionālo valodu funkcionalitāte, salīdzinot ar nacionālajām valodām, ir šaurāka, tās vairāk tiek lietotas neformālos kontekstos, diezgan nestabila ir arī šo valodu vieta oficiālajās izglītības sistēmās. Aptaujas dalībnieku otrajā grupā (“Zina un nelieto”) no 43 respondentiem 70 % ir atzinuši, ka nezina, ko tieši varētu darīt ar latgaliešu valodas korpusu, lai gan citu valodu korpusus (piemēram, latviešu, angļu) viņi lieto. Starp respondentiem ir vērojama arī pedagoģiskā interese mācīt latgaliešu valodu, tomēr korpusa valodas apguves procesā gandrīz netiek izmantots.

Aptaujas dati un līdzšinējā saziņa ar dažādām mērķauditorijām (studentiem, skolēniem, skolotājiem, pētniekiem) ļauj izdarīt secinājumus, ka korpusa funkcionalitātes paplašināšana ir plānojama trīs virzienos: valodu izglītības, mediju satura bagātināšanas un pētniecības virzienā. Šī mērķa sasniegšanai ir nepieciešama korpusa popularizēšana un sabiedrības informēšana kopumā, bet vēl nozīmīgāka ir iesaistīšanās praktiskā darbā un dialogā ar latgaliešu valodas skolotājiem, ar studentiem – nākamajiem latviešu valodas skolotājiem un Latgales mediju darbiniekiem. Mērķauditorijai, kas nav valodnieki, ir nepieciešami konkrēti, praktiski piemēri, kas atklāj korpusu kā vērtīgu resursu valodas un satura iepazīšanai un apguvei. Viņiem jāpalīdz pamanīt tās iespējas, ko piedāvā korpus, piemēram, iepazīt daudzveidīgu valodas lietojumu autentiskās situācijās (intervijās, emuāros u. tml.). Tāpat arī humanitāro un sociālo zinātņu (mediju) pētniekiem korpus var kļūt par žanru, satura un valodas stila ziņā bagātu tekstu krātuvi individuālo pētījumu veikšanai. Savukārt valodniekiem darbā ar korpusu vēl aizvien aktuāls un neatrisināts ir lemmatizācijas un morfoloģiskās marķēšanas jautājums, kas īpaši nozīmīgi būtu, piemēram, vārdnīcu izstrādē.

Avoti (tekstā minētie valodu korpusi)

CorCenCC (velsiešu valodas korpus). Knight, D., Morris, S., Fitzpatrick, T., Rayson, P., Spasić, I., Thomas, E-M., Lovell, A., Morris, J., Evas, J., Stonelake, M., Arman, L., Davies, J., Ezeani, I., Neale, S., Needs, J., Piao, S., Rees, M., Watkins, G., Williams, L., Muralidaran, V., Tovey-Walsh, B., Anthony, L., Cobb, T., Deuchar, M., Donnelly, K., McCarthy, M. and Scannell, K. (2020). *CorCenCC: Corpus Cenedlaethol Cymraeg Cyfoes – the National Corpus of Contemporary Welsh*. Cardiff University. Pieejams: <http://doi.org/10.17035/d.2020.0119878310> [skatīts 02.07.2022.].

Corpas na Gàidhlig (skotu gēlu valodas korpus). *Corpas na Gàidhlig, Digital Archive of Scottish Gaelic (DASG)*. University of Glasgow. Pieejams: <https://dasg.ac.uk/corpus/> [skatīts 05.07.2022.].

DTK (lejas sorbu valodas korpus). Serbski institut (n.d.). *Dolnoserbski tekstowy korpus*. Pieejams: <https://www.niedersorbisch.de/korpus/> [skatīts 05.07.2022.].

ESLEMA (astūriešu valodas korpuss). *Eslema Corpus de la llingua asturiana*. Pieejams: <https://eslema.it.uniovi.es/corpus/busqueda.html> [skatīts 01.07.2022.].

ETC (basku valodas korpuss). Euskara Institutua / Basque Language Institute (2013-2021). *Eguno Testuen Corpua (ETC)/ Corpus of Contemporary Basque (ETC)*. Pieejams: <https://www.ehu.eus/etc/> [skatīts 05.07.2022.].

KŚM (silēziešu valodas korpuss). Kulik, Grzegorz (2018). *Korpus Ślōnskij Mowy*. Pieejams: <https://korpus.silling.org/query?corpname=silesian> [skatīts 05.07.2022.].

MuLa (mūsdienu latgaliešu tekstu korpuss). LU Matemātikas un informātikas institūts, Rēzeknes Tehnoloģiju akadēmija (2011–2013). Pieejams: <http://www.korpuss.lv/id/MuLa> [skatīts 01.07.2022.].

SIKOR (ziemeļsāmu valodas korpuss). Giellatekno - Saami Language Technology, UiT The Arctic University of Norway and The Divvun group at UiT The Arctic University of Norway, 2015, *SIKOR North Saami free corpus*, Common Language Resources and Technology Infrastructure Norway (CLARINO) Bergen Repository, Pieejams: <http://hdl.handle.net/11509/100> [skatīts 01.07.2022.].

Tagged Icelandic Corpus (islandiešu valodas korpuss). Stofnun Árna Magnússonar í íslenskum fræðum / Árni Magnússon Institute for Icelandic Studies (2006-). *MÍM Mörkuð Íslensk Málheild / The Tagged Icelandic Corpus*. Pieejams: <https://clarin.is/en/resources/mim/> [skatīts 05.07.2022.].

- Bartels, Hauke (2012). Massnahmen zur Dokumentation des Niedersorbischen. *Slavia Occidentalis*, No. 69, pp. 7–22.
- Bartels, Hauke (2020). Das niedersorbische Globalkorpus als Ziel einer ganzheitlichen Konzeption zum Aufbau von Textkorpora. *LĚTOPIŠ. Zeitschrift für sorbische Sprache, Geschichte und Kultur*, 2020, No. 2, pp. 3–44
- Bernhard, Delphine; Ligozat, Anne-Laure; Bras, Myriam; Martin, Fanny; Vergez-Couret, Marianne; Erhart, Pascale; Sibille, Jean; Todorascu, Amalia; Boula de Mareuil, Philippe; Huck, Dominique (2021). Collecting and annotating corpora for three under-resourced languages of France: Methodological issues. *Language Documentation & Conservation*, No. 15, pp. 316–357 Available: <https://scholarspace.manoa.hawaii.edu/handle/10125/74645> [accessed 13.11.2021.].
- Briška, Anna (2013). *Mūsdienu latgaliešu rakstu valodas korpus: izveide un izmantošanas iespējas*. Maģistra darbs. Npublicēts. Rēzekne: Rēzeknes Augstskola.
- Cox, Christopher (2011). Corpus linguistics and language documentation: challenges for collaboration. Newman, John, Baayen, Harald, Rice, Sally (eds.). *Corpus-based studies in language use, language learning, and language documentation*. Leiden: Brill, pp. 239–264.
- Čermák, František (2011). The Case of The Czech National Corpus: Its Design and Development. Gozdz-Roszkowski, S. (ed.). *Explorations across Languages and Corpora*. Frankfurt: P. Lang, pp. 29–44. Available: [https://nanopdf.com/download/1-general-re-](https://nanopdf.com/download/1-general-re-marks-esky-narodni-korpus_pdf)
- marks-esky-narodni-korpus_pdf [accessed 13.11.2021.].
- Jakubíček, Miloš, Kilgariff, Adam, Kovář, Vojtěch, Rychlý, Pavel, Suchomel, Vít (2013). The TenTen Corpus Family. *7th International Corpus Linguistics Conference CL*, pp. 125–127. Available: https://www.sketchengine.eu/wp-content/uploads/The_TenTen_Corpus_2013.pdf [accessed 13.11.2021.].
- Jauhiainen, Heidi, Jauhiainen, Tommi, Lindén, Krister (2020). Building web corpora for minority languages. *Proceedings of the 12th Web as Corpus Conference (LREC 2020)*, Marseille, 11–16 May 2020, pp. 23–32. Available: <http://www.lrec-conf.org/proceedings/lrec2020/workshops/WAC-II/index.html> [accessed 13.11.2021.].
- Knight, Dawn, Morris, Steve, Fitzpatrick, Tess (2021). *Corpus Design and Construction in Minoritised Language Contexts: The National Corpus of Contemporary Welsh*. London: Palgrave.
- Knight, Dawn, Morris, Steve, Fitzpatrick, Tess, Rayson, Paul, Spasić, Irena, Thomas, Enlli Môn (2020). *The National Corpus of Contemporary Welsh: Project Report* | Y Corpws Cenedlaethol Cymraeg Cyfoes: Adroddiad y Prosiect. arXiv:2010.05542, October 2020.
- Kulik, Grzegorz (2018). Korpus Ślōnskij Mōwy – prymiera. *Wachtyrz* 19.12.2018. Available: <https://wachtyrz.eu/korpus-slonskij-mowy-prymiera/> [accessed 12.11.2021.].
- Levāne-Petrova, Kristīne (2012). Līdzsvarots mūsdienu latviešu valodas tekstu korpus un tā tekstu atlases kritēriji. *Baltistica* VIII Priedas 2012, 89.–98. lpp.
- Martena, Sanita, Marten, Heiko, Šuplinska, Ilga (2022). *Latgalian. The Latgalian language in education in Latvia*. 2nd Edition. The Netherlands: Mercator, Fryske Akademy.
- McEnery, Tony, Hardie, Andrew (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- McEnery, Tony, Ostler, Nicholas (2000). A new agenda for corpus linguistics – working with all of the world's languages. *Literary and Linguistic Computing*, No. 15(4), pp. 403–420.
- Millour, Alice, Fort, Karën (2020). Text Corpora and the Challenge of Newly Written Languages. *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, Marseille: European Language Resources association, pp. 111–120. Available: <https://aclanthology.org/2020.sltu-1.15/> [accessed 12.11.2021.].
- Ostler, Nicholas (2009). Corpora of less studied languages. *Corpus linguistics: an international handbook*. Lüdeling Anke, Kytö, Merja (eds.). Berlin, New York: Walter de Gruyter, pp. 457–483.
- Pomierska, Justyna, Stanulewicz, Danuta (2019). Tworzenie korpusu języka kaszubskiego – dostępne źródła zdigitalizowanych tekstów. *Acta Casubiana*, No. 21, pp. 159–178
- Sampson, Geoffrey, McCarthy, Diana (2007). *Corpus linguistics: Readings in a widening discipline*. London: Continuum, pp. 1–2.

Scannel, Kevin (2007). The Crúbadán Project: Corpus building for under-resourced languages. *Cahiers du Cental*, No. 4, pp. 5–15.

Šuplinska, Ilga (zyn. red.) (2012). *Latgolys lingvoteritorialuo vuordineica. Lingvoterritorial Dictionary of Latgale*. II. Rēzekne: Rēzeknis Augstškola.

UNESCO (2017). *Endangered languages. A methodology for assessing language vitality and endangerment*. Available: <http://www.unesco.org/new/en/culture/themes/endangered-languages/language-vitality/#topPage> [accessed 28.10.2021.].

Viejo, X.; Sauri, R; Neira, A. (2008). Eslema. Towards a corpus for Asturian. *Collaboration: Interoperability between people in the creation of language resources for less-resourced languages. A SALTMI workshop. LREC 2008*. Marrakesh.

The Corpus of Latgalian in the Context of Other Lesser Used Languages of Europe: Characterization, Usage and Potential

Sanita Martena, Anna Briška, Nicole Nau

Keywords: Latgalian, corpus, corpus of Latgalian, MuLa, regional languages, lesser-used languages of Europe

The article reports on the current process of completing and further developing the corpus of contemporary written Latgalian (MuLa). It gives an overview over the sources and the principles of compiling the corpus and compares it with corpora of other lesser used languages of Europe. In addition, it analyses the profile of current and potential users, users' experience with the corpus so far and their wishes for the future. The first version of the corpus MuLa, which has been publicly available since 2012, contains 1 million running words. It will be enlarged to 2 million words in the extended and corrected version prepared 2020–2022.

Corpora of lesser used languages play an important role in the documentation and development of the language. They are also a valuable resource for the preparation of linguistic tools and teaching materials. While MuLa does not have either the size or the functionality of corpora of such well-cared-for European regional languages as Basque, Welsh, or Sami, the fact that a corpus exists and is being further developed puts Latgalian in a better position than some other regional languages in Europe. Due to a shortage of financial and human resources, corpora of lesser used languages are often compiled from sources automatically gathered from the Internet. MuLa, in turn, is still compiled manually, which allows higher control of register diversity and balance, as well as linguistic quality. The corpus is not tagged, but it is accessible to everyone.

Data about the usage and users of MuLa since 2012 have been collected with an online questionnaire answered by 214 respondents. The study shows that the corpus has been used very little, mostly by researchers and a few other professionals. On the other hand, many respondents expressed a potential interest and ideas about potential uses of the corpus for learning about Latgalian as well as further developing their linguistic skills. In order for the corpus to be used more broadly, promotion and the spread of information within society is indispensable. Still more important is cooperation and a constant dialogue with teachers and university students.

Mūsdienu latgaliešu valodas runas korpusa izveide mazāk lietoto valodu dokumentēšanas kontekstā

Angelika Juško-Štekele,
Antra Kļavinska

Raksts tapis Valsts pētījuma programmas “Humanitāro zinātņu digitālie resursi” projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001).

Ievads

Mazāk lietoto valodu dokumentēšanai pēdējā laikā tiek veidoti dažādi rakstītu un mutvārdu tekstu korpusi, audio, videoieraksti, tiek veikti korpusā balstīti pētījumi. Atšķirībā no rakstu valodas, runātā valoda ir spontānāka un dabiskāka, turklāt nestandartizēti valodas varianti atklāj valodas daudzveidību un sniedz ieskatu pārmaiņās, kas notiek valodā laikā un telpā, kā arī saskarē ar citām valodām un valodas variantiem. Salīdzinot ar standartizētas rakstu valodas korpusu, reģionālā valodas varianta runas korpusa izveide ir daudz komplicētāka pat lielākajās pasaules valodās, piemēram, angļu valodā (Anderson et al. 2007; Anderwald, Wagner 2007), un vēl jo vairāk tādām valodām kā latviešu vai latgaliešu valoda.

Saskaņā ar 2011. gada tautas skaitīšanas datiem latgaliešu valodu ikdienā lieto 164 510 Latvijas iedzīvotāju, kas ir apmēram 8 % no kopējā Latvijas iedzīvotāju skaita. Pēc UNESCO datiem latgaliešu valoda ar 150 000 uzrādītiem lietotājiem 2013. gadā atzīta par vienu no pasaules apdraudētajām valodām, kas pakļauta riskam (*Vulnerable*), jo visas paaudzes joprojām lieto mutvārdu formu, bet valodas ilgtspēja ir nopietni apdraudēta (Moseley 2020), tāpēc ka gados jaunāko valodas lietotāju skaits samazinās. Plašs pasākumu loks reģionālo vai mazāk lietoto valodu aizsargāšanai un lietošanas veicināšanai ietverts Eiropas Reģionālo vai minoritāšu valodu hartā (pieņemta 1992. gadā)¹. Savukārt Eiropas Parlamenta 2013. gada 11. septembra rezolūcija par Eiropas apdraudētajām valodām un lingvistisko daudzveidību Eiropas Savienībā uzsver, ka efektīvus rīkus Eiropas apdraudēto valodu prasmes, izplatīšanas, mācīšanas un saglabāšanas veicināšanai var nodrošināt jaunās tehnoloģijas, multivides platformas, kas ietver atbalstu gan saturam, gan lietojumprogrammām².

Ievērojot ES direktīvas un ieteikumus reģionālo un riskam pakļauto valodu saglabāšanai, izpētei un attīstībai, kā arī Valsts valodas politikas nostādnes 2021.–2027. gadam attiecībā uz daudzveidīgu tekstu korpusu izstrādi, publiskošanu tīmeklī un pieejamību, Rēzeknes Tehnoloģiju akadēmijas pētnieku grupa valsts pētījumu

1 Eiropas Reģionālo vai minoritāšu valodu harta. Pieejams: <https://m.likumi.lv/doc.php?id=209969> [skatīts 25.10.2021.].

2 Eiropas Parlamenta 2013. gada 11. septembra rezolūcija par Eiropas apdraudētajām valodām un lingvistisko daudzveidību Eiropas Savienībā (2016/C 093/07) Eiropas Savienības Oficiālais Vēstnesis. 2016. gada 9. marts, Nr. 59., 52.–58.lpp. Pieejams: <https://eur-lex.europa.eu/legal-content/LV/TXT/PDF/?uri=OJ:C:2016:093:FULL&from=ES> [skatīts 25.10.2021.].

programmas projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001)” 2020. gadā uzsākusi darbu pie mūsdienu latgaliešu valodas runas korpusa (MuLaR) izveides, un tas paredzēts latgaliešu valodas dokumentēšanai, apguvei, studiju un pētniecības vajadzībām.

Dialektu un izlokšņu daudzveidības dēļ runas korpusa izveide ir laikietilpīgs process, kas ietver daudz manuāla darba, lai transkribētu un apstrādātu datus, pirms tie iegūst korpusa formu (Vuković 2021). Neraugoties uz to, saskatāmas daudzas apdraudēto, mazāk lietoto valodu dokumentēšanas priekšrocības, veidojot runas korpusus: pieejamība un ilgtspējība; liels oriģinālu, digitāli arhivētu, transkribētu, anotētu un katalogizētu datu apjoms; daudzfunkcionalitāte, aptverot potenciāli dažādas tēmas, tekstu žānrus; kontekstualizācija, kas iespējama, izmantojot metadatus; iespēja atspoguļot dažādos laika posmos dokumentētus datus u. tml. (Rießler, Wilbur 2017: 60).

Raksta mērķis ir, lietojot zinātniskās literatūras referatīvās analīzes un salīdzinošo metodiku, identificēt un analizēt problēmjautājumus, kas nozīmīgi MuLaR izveides procesā, proti:

- 1) kā izveidot reprezentatīvu latgaliešu valodas runas korpusu, ņemot vērā latgalisko izlokšņu daudzveidību;
- 2) kādu metodoloģiju izmantot dabiskas, spontānas valodas dokumentēšanai (jaunu datu vākšana, pastāvošo ierakstu izmantošana);
- 3) kā veikt runas ierakstu transkribēšanu, lai pēc iespējas precīzāk atklātu mutvārdu teksta iezīmes;
- 4) kā izveidot atvērtā piekļuvē pieejamu, ērti izmantojamu runas korpusa platformu.

Eiropas reģionālo valodu un dialektu runas korpusi

Pēc EP oficiālās informācijas, ES ir vairāk nekā 60 vietējo reģionālo un mazākumtautību valodu, kurās runā apmēram 40 miljoni cilvēku (Pasikowska-Schnass 2016: 1). Eiropas vienotajā valodas resursu un tehnoloģiju infrastruktūrā CLARIN pieejami 129 runas korpusi, no tiem 118 veidoti ar paralēlu audioierakstu un tā transkripcijas funkciju, 11 – tikai ar transkribētiem tekstiem³. CLARIN platformā atrodam, piemēram, katalāņu valodas fonoprosodisku runas korpusu (PhonCAT), kas satur 8719 minūšu spontānas runas ierakstus no 234 Barselonā dzīvojošiem katalāņu valodā runājošiem vīriešiem vecumā no 5 līdz 45 gadiem. Igaņu dialektu korpus satur fonētiski transkribētus dialogus 1 284 000 vārdu apjomā, kas

3 CLARIN Spoken Corpora. Introduction (2021). Available: <https://www.clarin.eu/resource-families/spoken-corpora> [accessed 14.11.2021.].

ierakstīti ilgākā laika periodā, sākot no 1938. gada līdz 20. gadsimta 70. gadiem. Pie masīvākajiem dialektu korpusiem jānosauca, piemēram, Freiburgas angļu valodas dialektu korpus (FRED), kas ietver Britu salu iedzīvotāju ķeltisko dialektu morfosintaktiskās variācijas, ieraksti veikti no 2000. līdz 2005. gadam. FRED korpus ietver 2,5 miljonus vārdlietojumu, 300 stundu ierakstus no 372 intervējamām personām 163 dažādās apdzīvotās vietās, kas administratīvi atbilst 9 dialektu apgabaliem (Hernández 2006).

Veidojot MuLaR, tiek ņemti vērā labās prakses piemēri citu mazāk lietoto valodu un dialektu runas korpusu izstrādē. Viens no tādiem labi strukturētiem un ērti lietojamiem resursiem ir Spišas dialektu korpus (*Korpus Spiški*) – mutvārdu tekstu kolekcija, kas dokumentē Polijas Spišas reģiona iedzīvotāju runu. Korpus ir paredzēts ikvienam reģionālās valodas un kultūras entuziastam, lietotājiem ir brīva piekļuve teksta transkripcijām un audio ierakstiem. Teksti ir pierakstīti ortogrāfiskajā transkripcijā, izmantojot poļu alfabētu, tāpēc tos var lasīt un meklēt arī lietotāji, kas nepārzina fonētisko transkripciju. Meklētājprogramma nodrošina dažādas meklēšanas kategorijas, piemēram, lemmas, dažādas gramatiskās formas, atbilstošus audioieraksta segmentus un metadatus. Leksika, kas raksturīga šim reģionam, ir skaidrota vārdnīcā. Teksti ir morfoloģiski marķēti, tas zinātniekiem sniedz iespējas veikt pētījumus par gramatisko sistēmu, vārddarināšanas procesiem, sintaksi un leksiku. Augstas kvalitātes audioieraksti ir piemēroti fonētikas pētījumiem, savukārt metadati (par runātāju vecumu, dzimumu, dzīvesvietu, izglītību, tautību) sniedz iespējas sociolingvistikas pētījumiem (Grochola-Szczepanek et al. 2019).

Korpusa mērķa un tehnoloģisko risinājumu ziņā MuLaR ir saistoša arī čehu valodas dialektu korpusa DIALEKT pieredze. DIALEKT veidotāji to piesaka kā viegli izmantojamu korpusu, kas piemērots ne tikai dialektoloģijas pētniekiem, bet arī valodniecības robežzinātņu speciālistiem, skolotājiem un citiem interesentiem. DIALEKT iekļauti gan vēsturiski dati no 20. gadsimta 50.–80. gadu ekspedīcijām, kas jau iestrādāti valodniecības atlantos u. c. zinātniskos izdevumos, gan jaunāki ieraksti, kas neparedz striktu apjoma ierobežojumu, bet, līdzīgi kā MuLaR iecerē, ir orientēti uz regulāru datu papildināšanu, nepretendējot uz stingru datu kopu balansu. Tomēr DIALEKT saglabā dialektu apgabalu principu, kas redzams korpusa metadatos. Ierakstu transkribēšanai korpusā tiek izmantots komplekss ortogrāfiski dialektoloģiskais princips ar dialektoloģiskā principa pārsvaru, kas pamatots ar tradicionālām dialektoloģiskā pieraksta konvencijām Čehijā (Komrsková et al. 2017).

Iestrādes

Lai nodrošinātu mūsdienīgus digitālos resursus un rīkus latgaliešu valodas dokumentēšanai, pētniecībai un apguvei, tiek veidots Mūsdienu latgaliešu valodas tekstu korpus (MuLa). Patlaban pieejamajā versijā noteiktās proporcijās ievietoti latgaliešu rakstu valodā publicēti (1988–2012) teksti ar metadatiem, apjoms – viens miljons vārdlietojumu. Projektā

“Humanitāro zinātņu digitālie resursi: integrācija un attīstība” korpuss tiek pilnveidots un papildināts ar jaunākiem tekstiem, kā arī tiek veidots jauns mūsdienu latgaliešu valodas runas korpuss (MuLaR)⁴.

Ar vienošanās līgumu MuLaR tiek integrēti projektā “Triangulation Approach for Modelling Convergence with a High Zoom-In Factor” (TriMCo⁵) un projektā “Diachronic Typology of Differential Argument Marking”⁶ transkribētie latgaliešu runas teksti. TriMCo projekta empīriskās bāzes vajadzībām, izmantojot ELAN programmatūru, tika veidots multilingvāls Dialektu korpuss ar krievu, baltkrievu, lietuviešu un latgaliešu tekstiem, kas projektu ierobežotā laika un cilvēkresursu trūkuma dēļ nevainagojās ar lietotājiem pieejamu latgaliešu runas datni.

MuLaR tiek izmantotas arī Rēzeknes Tehnoloģijas akadēmijas zinātnisko grantu konkursā finansētā projekta “Austrumu diasporas latgalisko izlokšņu ierakstu datubāzes izveide” iestrādes (projekta partneris – Krasnojarskas latviešu biedrība “Dzintars”). Projektā tika izveidots publiski pieejams interneta resurss (<https://reiti.rta.lv/>), kur ievietoti lauka pētījumos Sibīrijā (galvenokārt Krasnojarskas un Tomskas apgabalā) no latgalisko izlokšņu runātājiem iegūtie videoieraksti, kā arī transkribēti audioierakstu teksti. Projekta tīmekļvietnē ir īstenotas meklēšanas iespējas ar atslēgvārdiem, pieejami metadati (respondenta dzīvesvieta, vecums, dzimums, izglītība), kā arī karte, kurā apzīmētas intervēšanas vietas. Ne visi lauka pētījumos iegūtie ieraksti projekta laikā tika transkribēti un ievietoti datubāzē, šis darbs tiek turpināts MuLaR izstrādes procesā.

Reprezentatīva latgaliešu runas korpusa problemātika

Reprezentativitāte ir viens no plašāk un daudzveidīgāk skatītajiem korpuslingvistikas jautājumiem, kas visbiežāk korelē ar korpusa līdzsvarotības un datu ieguves jautājumiem. Berlīnes Universitātes lingvistikas profesors Anatols Stefanovičs (*Anatol Stefanowitsch*) valodas korpusa reprezentativitāti saista ar tā līdzsvarotību un blakus korpusa autentiskumam un apjomam atzīst to par vienu no būtiskākajām valodas korpusa pazīmēm (Stefanowitsch 2020: 22–23), kas ir saistīta ar korpusa mērķi un plānoto lietojumu. MuLaR saskaņā ar projekta sākotnējo ieceri ir paredzēts latgaliešu valodas dokumentēšanas, pētniecības, studiju un apguves vajadzībām. Šim nolūkam MuLaR reprezentativitātes kritēriju lokā tiek vērtēti: aptverošs

- 4 Skat šajā izdevumā Sanitas Martenas, Annas Briškas un Nikoles Nauas rakstu “Latgaliešu valodas korpuss citu Eiropas mazāk lietoto valodu kontekstā: korpusa raksturojums, lietojums un potenciālā iespējošana”.
- 5 Prof. Dr. Björn Wiemer, German Research Foundation, Johannes-Gutenberg-University Mainz, projekta numurs: WI 1286/16-1.
- 6 Dr. Ilja A. Seržant, Incoming Fellowship Programme, Grant Agreement Number: 291784, Zukunfts Kolleg, University of Konstanz.

mūsdienu latgaliešu valodas lietojuma areāls, autentiski teksti, kas fiksēti dialoga, poliloga vai stāstījuma situācijā un papildināti ar pētniecībai saistošiem valodas lietotāju sociodemogrāfiskiem metadatiem.

Lai nodrošinātu reprezentatīvu mūsdienu latgaliešu runātās valodas aptveri, ir būtiski fiksēt autentisku valodas lietojumu visā latgaliešu valodas lietošanas areālā. Tipoloģiski iespējams izdalīt vismaz četrus šādus areālus: Latgales kopiena (Ltg), pārējo Latvijas novadu (pārnovadu) kopiena (Lv), Austrumu diaspora Krievijā, Baltkrievijā u. c. (Ad) un Rietumu diaspora ASV, Kanādā, Zviedrijā, Austrālijā, Vācijā, Polijā, Lietuvā u. c. (Rd). Pēc 2011. gada tautas skaitīšanas datiem, ikdienā latgaliešu valodu lieto 35,5 % no Latgales iedzīvotājiem (Ltg), 9,2 % no iedzīvotājiem Rīgā un Pierīgā (LvRP), 4,8 % iedzīvotāju Vidzemē (LvV), 4,3 % Zemgalē (LvZ) un 1,5 % Kurzemē (LvK). Latgaliešu valodas lietošanas aptvere Austrumu diasporā un Rietumu diasporā bez papildu statistikas datiem nav precīzi nosakāma, tāpēc MuLaR nevar pretendēt uz pilnīgi precīzu datu līdzsvaru. Atsaucoties uz Ziemeļarizonas Universitātes profesora Dagleša Baibera (*Douglas Biber*) viedokli, MuLaR pētnieku grupa pieņem, ka korpusu reprezentativitāte ir ne tik daudz atkarīga no filigrānām skaitliskajām proporcijām, cik no tā, vai korpusā iekļautie teksti atklāj iespējami pilnīgu ainu par valodas lietojumu (Biber 1993).

MuLaR izstrādes pirmajā posmā iekļaušanai korpusā tiek apstrādāti 20 stundu audioieraksti, no kuriem 4 stundas ir Austrumu diasporas (Sibīrijas) audioieraksti, 16 stundas – ieraksti, kas veikti Latgales teritorijā. Balstoties uz aprēķina, ka dominējošais latgaliešu valodas lietotāju skaits ir Latgales iedzīvotāji, MuLaR Latgales ierakstu apjomam ilgtermiņā jānodrošina iespējami adekvāts balanss ar latgaliešu valodas ierakstiem pārējos Latvijas novados, kas pēc tautskaites datiem ir nosacīti nedaudz vairāk nekā puse (55,7 %) no Latgales kopienas. Austrumu un Rietumu diasporas ieraksti attiecīgi veido atlikušo nosacīto pusi. Tādējādi MuLaR valodas teritoriālās aptveres balansam iespējams izmantot formulu: $MuLaR/100\% = 50\%bLtg + 26\%bLv + 12\%bAd + 12\%bRd$, kur Ltg – Latgales kopienas audioieraksti, Lv – pārnovadu kopienas audioieraksti, Ad – Austrumu diasporas audioieraksti, Rd – rietumu kopienas audioieraksti, *b* – mainīgā audioieraksta stundu daļa. Turpmākajā korpusa izstrādes gaitā balanss var mainīties, jo sevišķi Austrumu un Rietumu diasporas sadaļās, kur apjoma palielināšanās varētu nebūt būtiska, saglabājot reprezentativitātes principu, cik tas iespējams. Iespējami pilnīgam balansam Latgales kopienas un pārnovadu kopienas sadaļās tiek veidota sava iekšējā struktūra, kas, izmantojot tautskaites datus, sīkāk sadalās pēc latgaliešu valodas lietotāju administratīvi teritoriālā pārkļājuma.

Latgaliešu valodas lietotāju administratīvi teritoriālais sadalījums savā ziņā korelē ar korpusa izstrādē aktuālu jautājumu par latgaliešu valodas izlokšņu skaitu, kas Latgalē ir apmēram 70. Atsaucoties uz valodnieces Annas Stafeckas pētījumiem, MuLaR pētnieku grupa pieturas pie atzinuma, ka Latgalē izlokšņu robežas nosacīti atbilst viena pagasta teritorijai (pēc 1939. g. administratīvi teritoriālā iedalījuma) (Stafecka 2017: 58). Līdz ar to MuLaR reprezentativitātes nodrošināšanai nav plānots piemērot izlokšņu kritēriju, atstājot iespēju pētniekiem pētīt izlokšņu iezīmes pēc administratīvi teritoriālā dalījuma un pētniecībai pieejamiem teicēju metadatiem.

Būtisks jautājums MuLaR izstrādes procesā ir runātāju metadati: dzimums, vecums, dzīvesvieta, kas tiek dokumentēti ar mērķi padarīt korpusu saistošu turpmākiem sociolingvistikas pētījumiem, jo sevišķi, analizējot latgaliešu valodas lietotāju skaitu un pašas valodas īpatnības dažādās vecuma grupās un to ietekmi uz iespējamo latgaliešu valodas lietojumu nākotnē. Jau 2011. gada tautskaites dati rāda, ka vairumā Latgales novadu latgaliešu valodā runājošo skaits samazinās vecumā līdz 19 gadiem un palielinās vecumā virs 70 gadiem, atsevišķos novados (Baltinavas, Rugāju) vecumā virs 90 gadiem pat sasniedzot 100 %. Balstoties uz tautskaites datiem par Latgales iedzīvotāju – ikdienas latgaliešu valodas lietotāju – sadalījumu pa vecuma grupām, Latgales kopienas interviju ierakstus, ievērojot iekšējo teritoriālo proporciju, plānots strukturēt šādi: 15 % ieraksti no iedzīvotājiem vecumā līdz 19 gadiem, 45 % – no 20 līdz 65 gadiem, 35 % – vecumā virs 65 gadiem. Pārējos Latvijas novados piemērojamā statistika, kas veidota, pamatojoties uz tautskaites datiem, ir šāda: 10 % – vecumā virs 19 gadiem, 50 % – vecumā no 20 līdz 65 gadiem, 40 % – vecumā virs 65 gadiem. Austrumu un Rietumu diasporas kopienās informantu vecuma līdzsvaru nav plānots īstenot, jo nav drošu datu par populācijas apjomu un vecuma struktūru.

Teritoriālā kopiena	Ltg	Lv	Ad	Rd	
Ierakstu apjoma proporcija	50 %	26 %	12 %	12 %	
Informantu dzimumu struktūra	S:V	50:50	50:50	50:50	50:50
Informantu vecuma struktūra	0–19	15	10		
	20–64	45	50		
	65+	50	40		
Teksta žanrs	Intervijas, dialogi, polilogi, stāstījumi				

1.tabula. MuLaR līdzsvarotības koncepcija.

MuLaR izstrādes posmi

Teorētiskajā literatūrā tiek definēti runas korpusa izveides posmi. Piemēram, Pols Tompsons (*Paul Thompson*) piedāvā četrus posmus: 1) datu vākšanas posms, kurā nepieciešams apspriest gan audio/video ierakstīšanas tehniskos aspektus, gan kontekstuālās informācijas vākšanu un dalībnieku piekrišanas ieguvī; 2) transkribēšanas posms; 3) marķēšanas un anotēšanas posms, kurā transkripcijas tiek pārveidotas mašīnlasāmā formātā, kā arī tiek veikta marķēšana; 4) piekļuves posms, kurā uzsvars vispirmām kārtām likts uz iespējamo korpusa lietojumu, proti, vai tas būs pieejams citiem pētniekiem un kādā formā to var darīt pieejamu, ja to vēlas (Thompson 2005). Dažkārt runas korpusa izstrāde tiek reducēta trīs

pamatposmos: 1) runas ierakstu veikšana, 2) transkribēšana un kodēšana, 3) pārvaldība un analīze (Adolphs, Knoght 2010: 40).

MuLaR ir pirmais latgaliešu valodas runas korpuss Latvijā, RTA pētniekiem līdz šim nav bijis pieredzes runas korpusu izveidē, tāpēc šajā gadījumā kā pirmais būtu izceļams plānošanas posms, kas Tompsona sniegtajā pirmā posmā aprakstā attiecināts galvenokārt uz tehniskajiem aspektiem un metadatiem. MuLaR korpusa plānošanas posms sākas ar citu runas korpusu izpēti, korpusa dizaina izvēli, konsultēšanos ar pieredzējušiem ārzemju kolēģiem, piemērotas transkribēšanas datorprogrammas izvēli un apguvi, iekļaujot tajā arī diskusijas par konvencijām. MuLaR korpusa izveide patlaban tiek turpināta šādos posmos: datu ieguve (interviju ierakstīšana un arhivēšana); datu apstrāde (transkribēšana); korpusa platformas izveide (piekļuves posms).

Datu ieguve

MuLaR ir iecerēts kā multimodāls spontānas runas datu kopums. Primārais uzdevums ir transkribēt jau iepriekš veiktos audioierakstus, kā arī organizēt jaunu datu ieguvu. Korpusa izveides sākumposmā tiek izmantoti Rēzeknes Tehnoloģiju akadēmijā organizētu lauka pētījumu materiāli un plašsaziņas līdzekļu ieraksti. Korpusa metadatos patlaban definēti trīs tekstu pamatavoti:

- 1) dažādos Latgales novados 2009.–2021. gadā ierakstītas intervijas, tostarp “TriMCo” un “Diachronic Typology of Differential Argument Marking” projektu dati (jau iepriekš transkribētās intervijas tiek pārbaudītas un rediģētas saskaņā ar MuLaR konvencijām);
- 2) Sibīrijā dzīvojošo latgaliešu runas ieraksti (lauka pētījumi 2017.–2018. gadā, kas līdz šim daļēji bija ievietoti austrumu diasporas latgalisko izlokšņu datubāzē (<https://reiti.rta.lv/>);
- 3) TV ieraksti (2018.–2020. gadā Latgales Reģionālajā televīzijā veiktas intervijas).

2021. gadā korpusa datu ieguves nolūkā tika organizēts lauka pētījums Aglonā. Tā laikā aptaujāti 22 informanti (3 – vecumā līdz 19 gadiem, 4 – vecumā virs 65 gadiem, 15 – vecumā no 20 līdz 65 gadiem), iegūti ieraksti 15 stundu garumā. MuLaR izveides turpmākajā gaitā plānots veikt papildu datu ieguvu pārējos Latvijas novados un Rietumu diasporā.

Lai nodrošinātu atbilstību Akadēmiskā godīguma vispārējām vadlīnijām un Eiropas Parlamenta un ES Padomes regulas “Par fizisku personu aizsardzību” prasībām, pirms intervijas veikšanas tiek iegūta mutiska (ierakstot audioierakstā) vai rakstiska informēta piekrišana (intervējot bērnus – vecāku vai aizbildņu piekrišana) datu sniegšanai un audioieraksta veikšanai. Rakstisku informēto piekrišanu paraksta ne tikai datu subjekts, bet arī pētnieks, kas piedalās intervijā (Tauginienē et al. 2019).

Konfidencialitātes un personas datu aizsardzības nolūkā intervējamo personu un intervētāju dati tiek šifrēti, ievērojot ES Vispārīgajā datu aizsardzības regulā definēto datu pseidonimizāciju, lai personas datus nebūtu iespējams saistīt ar konkrētu datu subjektu bez papildu informācijas izmantošanas. MuLaR datu uzglabāšanas politika paredz, ka šāda papildu informācijai tiks uzglabāta atsevišķi, nodrošinot tādu tehniskas un organizatoriskas drošības pakāpi, lai personas dati netiktu saistīti ar konkrētu datu subjektu⁷.

Ja personas runā lietoti uz viņu pašu vai kādu citu personu vērsti jutīgi dati, kas dod iespēju identificēt minēto personu, šādi teksta fragmenti korpusā netiek iekļauti. Katram audioierakstam un transkripcijas failam ir pievienoti metadati: ziņas par ieraksta vietu un laiku, ieraksta ilgumu, informanta dzimumu un vecumu. Korpusa izstrādē stingri tiek ievērots datu minimizēšanas princips, kas paredz, ka personas dati tiek apstrādāti tikai tam paredzētajam mērķim un tam nepieciešamajā apjomā.

Papildus runas ierakstīšanas procesam ir svarīgi dokumentēt informāciju par procesa dalībniekiem, runas situāciju, proti, metadatus jeb “datus par datiem”. Kā atzīst digitālo humanitāro zinātņu eksperts Lū Burnards (*Lou Burnard*), metadati precizē kontekstu, tādējādi ļaujot sasaistīt doto teksta paraugu ar tā izcelsmi. Lietojot valodas korpusu, neizbēgami rodas jautājumi par datu precizitāti un autentiskumu, un bez metadatiem pētnieks nevar atbildēt uz šiem jautājumiem (Burnard 2005: 31). Metadati ir ļoti svarīgi korpusam, lai varētu sasniegt reprezentativitātes, līdzsvara un viendabīguma standartus (Adolphs, Knoght 2010: 42).

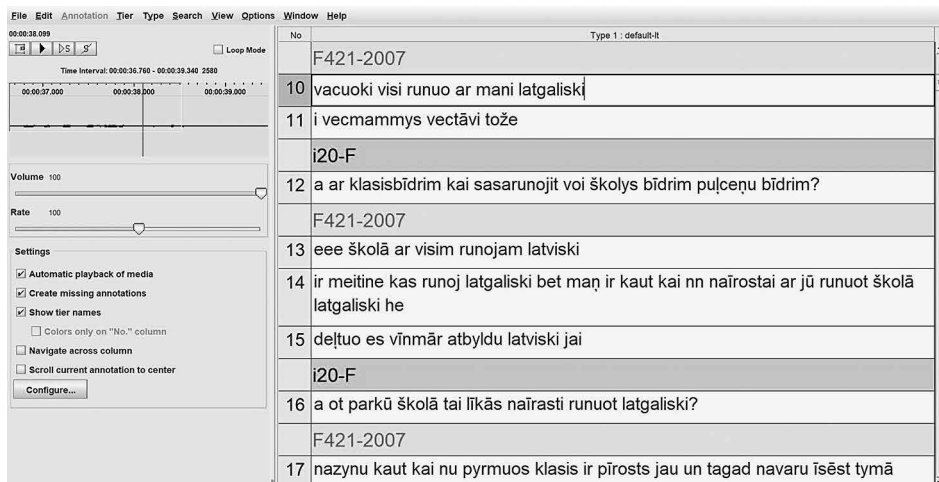
MuLaR plānošanas un izveides sākumposmā darba grupa ir vienojusies par metadatu kategorijām, piemēram, par to, kā tiks ģenerēti failu nosaukumi, kā tiks kodēti runātāju un intervētāju nosaukumi, kuri metadati būs izmantojami tikai korpusa administrēšanā, kuri būs publiski pieejami korpusā u. tml.

Datu apstrāde

Uzsākot plānot transkribēšanas procesu, vispirmām kārtām jāizlemj, kāds transkripcijas veids tiks izmantots: fonētiskā, prosodiskā, ortogrāfiskā transkripcija vai to kombinācija. Tas lielā mērā atkarīgs no runas korpusa mērķa (Thompson 2005).

Freiburgas angļu valodas dialektu korpusa (*FRED*) veidotāji, analizējot dialektu mutvārdu tekstu transkribēšanas problēmas runas korpusa vajadzībām, pamato ortogrāfiskās transkripcijas priekšrocības, salīdzinot ar fonētisko transkripciju. Ja konkrētā reģionā pastāv vairāki

7 Eiropas Parlamenta un Padomes Regula (ES) 2016/679 (2016) par fizisku personu aizsardzību attiecībā uz personas datu apstrādi un šādu datu brīvu apriti un ar ko atceļ Direktīvu 95/46/EK. Pieejams: <https://eur-lex.europa.eu/legal-content/LV/TXT/HTML/?uri=CELEX:32016R0679&from=LV> [skatīts 14.11.2021.].



1. attēls. Transkribēšana ELAN programmā.

dialekti un izloksnes, būtu sarežģīti atainot visas fonētiskās iezīmes vienkopus, turklāt tas ir arī ļoti laikietilpīgs process. Specifisku fonētiskās transkripcijas zīmju lietojums rada arī acīmredzamas tehniskas problēmas (fonētiskās transkripcijas lietojums datorrakstā, neiespējama visa veida marķieru meklēšana, jo pētniekam nebūtu ne jausmas, kādas formas meklēt). Būtu jāizmanto tāda transkripcija, kas ir dabisks kompromiss starp pieraksta detalizāciju (dziļumu) un pārklājumu (plašumu). Korpusa prasībām atbilstošāka ir ortogrāfiskā transkripcija, jo tam jābūt mašīnlasāmam, nodrošinot vieglu piekļuvi, dažādas meklēšanas iespējas un salīdzināmību ar citiem korpusiem (Anderwald, Wagner 2007).

Ņemot vērā citu runas korpusu izveides pieredzi, MuLaR korpusa veidotāji ir vienojušies mutvārdu tekstus atveidot latgaliešu valodas ortogrāfiskajā transkripcijā. Tā kā korpusa platformā plānots sasaistīt transkribēto tekstu ar audiofailu, ortogrāfiskā transkripcija atvieglo meklēšanas iespējas, tajā pašā laikā neizslēdz iespēju veikt pētījumus fonētikā, fonoloģijā, kā arī dialektoloģijā.

Transkribēšanai izmantota ELAN programmatūra, kas izstrādāta Maksa Planka Psiholingvistikas institūtā Neimegenā (Nīderlandē). ELAN ir rīks, kas paredzēts audio un video ierakstu transkribēšanai, nodrošina audioierakstu segmentēšanu, vairākus veidus, kā skatīt anotācijas, katrs skats ir savienots un sinhronizēts ar multivides laika skalu⁸.

8 ELAN (Version 6.2) [Computer software]. (2021). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>.

Rindas	Katram teicējam un katram intervētājam ir sava rinda, rindas nosaukums ir teicēja vai intervētāja kods. Pirmās pakāpes rindās tiek rakstīts tikai runas atšifrējums. Komentāri, skaidrojumi iespējami asociētajās rindās, piemēram, norāde uz vārdu vai izteikumu citā valodā.
Ortogrāfija	- Tiek lietota mūsdienu latgaliešu valodas ortogrāfija, bet runas pieraksts netiek koriģēts atbilstoši latgaliešu rakstības noteikumiem⁹, piemēram, vārds 'vējš' var tikt pierakstīts kā <i>viejs, viejš, vējs</i> u. tml. (atbilstoši izrunai). - Lielie sākumburti netiek lietoti segmenta sākumā, bet tiek lietoti īpašvārdos, piemēram, <i>Leldīne, Zīmyssuātķi, Dīvs</i>. - Skaitļi rakstāmi ar vārdiem. - Svarīgākie pieraksta nosacījumi (t. sk. problēmgadījumi). - Divskanis [uo] tiek pierakstīts ar diviem burtiem: <i>uo</i>, piemēram, <i>muosa</i>. Garais patskanis [ō] tiek atveidots ar diviem burtiem <i>oo</i>, burts <i>ō</i> netiek lietots. - Līdzskaņu burtus nemīkstina, kad aiz tiem seko patskaņi <i>e, ē, i, ī</i> vai divskaņi <i>ei, ie, iu</i>, arī aizguvumos no citām valodām, piemēram, no krievu valodas: <i>posle</i> (nevis <i>posle</i>). Ja skaidri saprotams, ka runātājs izrunā kādu vārdu vai frāzi latviešu literārajā valodā, tad līdzskani šajā pozīcijā mīkstina, piemēram, <i>atsevišķi gadījumi</i> (latgaliski būtu <i>atsevišķi gadījumi</i>). - Latgaliešu valodas sistēma lietojama arī īpašvārdos: <i>Andeni, Antonina</i> (nevis <i>Andeņi, Antoņina</i>). - Mīkstu līdzskaņu burtus <i>k, ģ, ņ, l</i> raksta pirms patskaņiem <i>a, ā, o, u, ū</i> un divskaņiem <i>au, uo</i>, citu patskaņu mīkstinājumu apzīmē ar <i>j</i>, piemēram, <i>ċjoce, acjom</i>. - Plato patskani apzīmē ar <i>e, ē</i> (attiecīgi arī divskani <i>eļ</i>), pat ja tas ir ļoti plats un tā izruna līdzinās [a], piemēram, <i>zeme</i> (nevis <i>zjamja</i>), <i>meita</i> (nevis <i>mjaita</i>). - Pozicionālās skaņu pārmaiņas pierakstā netiek atspoguļotas, piemēram, <i>atsasoka</i> (nevis <i>acasoka</i>). - Ja informanta runā lietotas saīsinātās formas, tās tā arī tiek pierakstītas nepārveidojot, piemēram, <i>guo, guoa, vāg, tāt, pascēja</i>.
Interpunkcija	- Pieturzīmes netiek lietotas, izņemot jautājuma zīmes, kas tiek lietotas īstā jautājuma teikuma beigās, piemēram, <i>a sātyš veinu taisiejoj?</i> - Pēdīņas tiek lietotas nosaukumos ar īpašvārdisku nozīmi, piemēram, <i>gruomota "Syltuo mola"</i>.
Prosodija	- Pauzes netiek norādītas (ja pauze ir gara, tā veido atsevišķu segmentu, ELAN programma automātiski mēra segmenta garumu). - Uzsvara vieta netiek fiksēta. Ja runā uzsvara vieta atšķiras no normas, to atzīmē asociētajā rindā.
Vārdu fragmenti	- Tiek rakstīta vārda daļa, izlaistā vārda fragmenta vietā tiek lietota defise. Piemēram, <i>runoj latv- latgaliešu volūdā</i>. - Aiz aprautās daļas ir atstarpe, piemēram, <i>s- s- smējās</i>.
Abreviācijas	- Ja tiek izrunāta vispārzināma abreviātūra (gan burtu, gan zillbju), tā tiek pierakstīta ar burtiem: <i>PVN</i> (izrunā <i>pē vē en</i>), arī tad, ja izrunā abreviātūru citā valodā, piemēram, <i>PR</i> [izrunā <i>pī ār</i>]. - Ja tiek lietots nezināms saīsinājums, tiek pierakstīts pēc izrunas, piemēram, <i>pē pē kā</i>.
Neskaidrs teksts	Tiek lietots apzīmējums [<i>unint.</i>]
Neverbālie elementi	- Apstiprinājuma signāls: <i>mbm</i>. - Pildītas pauzes, vilcināšanās, piemēram: <i>eee, aaa</i>. - Smieklī starp vārdiem: <i>haha, ha ha ha, be be, hibi</i> (atbilstoši skanējumam). - Smieklī vienlaicīgi ar vārdiem netiek atzīmēti.

2. tabula. MuLaR korpusa tekstu transkribēšanas pamatprincipi¹⁰.

- 9 Valsts valodas centra Latviešu valodas ekspertu komisijas Latgaliešu ortogrāfijas apakškomisijas lēmums Nr.1. 2007. gada 28. septembrī. "Par Latgaliešu rakstības noteikumiem". Pieejams: <https://likumi.lv/ta/id/164904-par-latgaliesu-rakstibas-noteikumiem> [skatīts 12.11.2021.].
- 10 Pateicamies Adama Mickeviča universitātes Poznaņā (Polijā) profesorei Nikolei Nauai (*Nicole Nau*) par konsultācijām un ieteikumiem transkribēšanas konvenciju izstrādē.

Oriģinālie audiofaili pirms transkribēšanas tiek apstrādāti: 1) tiek izgriezti fragmenti, kas satur nevajadzīgas detaļas, piemēram, personas datus, mūziku, runu citā valodā (izņemot atsevišķus vārdus vai teikumus) u. tml., 2) liela apjoma faili tiek sadalīti mazākos (10–30 minūtes); 3) lai atšifrēšana būtu vieglāka, faili tiek konvertēti WAV formātā.

Strādājot ELAN programmā, katram runātājam tiek piešķirta sava rinda, kurā ieraksta runas atšifrējumu. Rindas tiek segmentētas, ņemot vērā loģiskās izteikuma robežas, intonatīvās vienības, runas pauzes. Līdz ar to segmentu garums variējas. Katram runātājam (gan intervētajiem, gan informantiem) tiek piešķirts kods, lai atvērtajā piekļuvē pieejamais teksts būtu pseidonimizēts, bet projekta dalībnieki varētu viņus identificēt un saistīt ar metadatiem. Piemēram, runātāju kodi *M400-1971* vai *F406-1960*, kur redzama norāde uz dzimumu, kārtas numuru un dzimšanas gadu. Intervētāja kodā ir redzams kārtas numurs un norāde uz dzimumu, piemēram, *i01-F* vai *i04-M* (sk. 1. attēlu).

Sekojo citu ortogrāfiski transkribētu runas korpusu piemēriem (Anderwald, Wagner 2007; Grochola-Szczepanek 2019; Pinnis et al. 2014 u. c.), latgaliešu runas korpusa veidotāji ir izstrādājuši savas tekstu transkribēšanas konvencijas (sk. 2. tabulu). Darba grupā daudz diskutēts, cenšoties veidot tādu pierakstu, kas būtu saprotams plašam interesentu lokam, un tajā pašā laikā tiecoties pēc iespējas precīzāk atspoguļot mutvārdu teksta (tostarp individuālās runātāja) iezīmes. Diskusiju rezultātā nonāks pie lēmuma mutvārdu tekstus nepārveidot standartizētā latgaliešu literārajā rakstu valodā. Problēmas sagādā prosodijas atveidojuma detalizācijas pakāpes noteikšana transkribēšanas procesā, fonētisko un morfoloģisko variantu daudzveidības (dažādu latgalisko izlokšņu vai individuālo runas īpatnību) atspoguļojums. Tas, iespējams, var apgrūtināt meklēšanu korpusā, tāpēc plānots veidot tādu korpusa platformu, kas piedāvātu pēc iespējas plašākas iespējas meklēt atsevišķas vārda daļas. Ja, veicot transkribēšanu, tiek fiksēts kāds problēmgadījums, tas tiek apspriests darba grupā, tādējādi konvenciju saraksts pakāpeniski tiek papildināts vai precizēts. Lai pēc iespējas precīzāk sekotu konvencijām un novērstu atkāpes no tām, pirms ievietošanas korpusa platformā katru transkribējumu vēlreiz pārbauda un, ja nepieciešams, rediģē cits darba grupas dalībnieks.

Korpusa platformas izveide

Kā minēts iepriekš, viens no runas korpusiem, ko MuLaR korpusa veidotāji ir atzinuši par vērā ņemamu paraugu, ir Polijas Spišas dialekta korpus. Šī korpusa platformas veidotāji ir izmantojuši tīmekļa saskarni SpoCo.

SpoCo veidotāji to raksturo kā viegli uzturamu, efektīvu sistēmu valodas datu meklēšanai tīmeklī, kas piemērota dialektu runas korpusiem un salāgojama ar ELAN transkribēšanas programmu. Platformas lietojuma vienkāršība ir svarīga, lai atvieglotu pētniecību un pieejamību plašam lietotāju lokam, tostarp tiem, kam ir ierobežoti tehniskie un finanšu resursi. Sākotnēji SpoCo tika izstrādāta konkrēta krievu valodas dialekta runas korpusa mērķim, taču

tā ir viegli pielāgojama un izmantojama citām valodām un dialektiem. Patlaban šai saskarnei ir izveidotas divas versijas, kuru galvenā atšķirība ir vērojama to uzbūves tehnoloģijā, taču abām versijām ir kopīga lielākā daļa funkciju: lietotāju pārvaldība, meklēšana korpusā, korigēšanas iespējas, pilna teksta pārlūkošana (Waldenfels, Woźniak 2016).

MuLaR korpus ir veidots uz SpoCo platformas otrās versijas bāzes, šī versija tika veidota tieši Spišas dialektu korpusam. Platforma piedāvā meklēšanas iespējas, ierakstot pilnu vārdu, vārda sākuma vai beigu daļu. Ir iespējams ierobežot meklēšanu, metadatu sadaļā atlasot datus pēc avota (Latgalē, Austrumu diasporā vai medijos iegūti runas dati), pēc runātāju dzimuma, vecuma, dzīvesvietas, kā arī viena informanta runā (sk. 2. attēlu).

Meklēšanas rezultāts ir redzams segmenta līmenī, vajadzības gadījumā ir iespēja paplašināt kontekstu. Paralēli pieejams audiofails un transkripcija (sk. 3. attēlu).

Secinājumi un ieceres

Aspektam, ka MuLaR ir pirmais latgaliešu valodas runas korpus Latvijā, ir gan pozitīvās, gan negatīvās puses. Pozitīvās: pētnieki var radoši izpausties, pētīt un izvēlēties viņiem saistošāko korpusa dizainu, datorprogrammas, iegūt pieredzi no ārzemju kolēģiem, pētniekiem u. tml.; negatīvās: darbs pie runas korpusa ir komplicēts un laikietilpīgs.

Ļoti svarīgs ir plānošanas posms, lai nodrošinātu korpusa reprezentativitāti, lai tiktu savākta un precīzi dokumentēta visa korpusa izmantošanai būtiskā informācija (vienošanās ar informantiem, vajadzīgie metadati, audiofaili). Svarīgi ir plānot, kāds būs tehniskais nodrošinājums, lai tiktu veikti kvalitatīvi audioieraksti un droša datu glabāšana. Jau plānošanas posmā ir svarīgi izlemēt, kāda programmatūra tiks izmantota tekstu transkribēšanai un korpusa platformas izveidei (meklēšanas iespējas, iespējas transkribēto tekstu sasaistīt ar audiofailiem). Nozīmīgs darbs tiek veikts, izstrādājot mutvārdu tekstu transkribēšanas konvencijas, ir jāizdara izvēle starp plašumu un dziļumu, proti, starp lielu mutvārdu tekstu datu apjomu un detalizētu datu anotāciju. Ilgstošākais un darbietilpīgākais posms ir tekstu transkribēšana (vienas minūtes audioieraksta atšifrēšana atkarībā no pētnieka pieredzes prasa laiku līdz pat vienai stundai). Tā kā transkribēšanu veic galvenokārt trīs projekta dalībnieki, ļoti svarīgi ir saglabāt augstu konsekvences līmeni transkribēšanas procesā.

Līdz 2022. g. beigām plānots pabeigt lauka pētījumos iegūto audioierakstu transkribēšanu un pēc izvērtēšanas integrēt tos MuLaR korpusā. Projekta noslēgumā paredzēts veikt plānveidīgu MuLaR korpusa validāciju un tā darbības efektivitātes izvērtēšanu. 2021. gada decembrī ir apstiprināts valsts pētījumu programmas "Letonika latviskas un eiropiskas sabiedrības attīstībai" projekts "Latviešu valodas daudzveidība laikā un telpā", kurā viens no uzdevumiem ir turpināt MuLaR izstrādi, balstoties projekta "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" īstenošanas gaitā gūtajā pieredzē. Projektā, kurš turpināsies līdz 2024. gada decembrim,

MuLaR Corpus ▾ Statistics ▾ About ▾ Help [ltg lv en]

Basic Search ?

Search

☐ begins with ☐ ends with ☐ case sensitive Add / Delete Row + -

☐ exclude
 ☐ exclude
 ☐ exclude

☐ exclude
 ☐ exclude

CQL Search ?

Search

↑ 2. attēls. MuLaR korpasa saskarne.

↓ 3. attēls. Meklēšanā izgūstamo datu paraugs MuLaR korpasā.

MuLaR Corpus ▾ Statistics ▾ About ▾ Help [ltg lv en]

▶ 0:00 / 0:00 — 🔊 ⋮	a ka īrauga tāvs sasītu tiuleņ gīva rūkuos mītu	
	Person ID: F124-1929 Year of birth: 1929 Sex: f Place: Cibla	▼ □
	Source: L	
▶ 0:00 / 0:00 — 🔊 ⋮	<i>F124-1929:</i> mobilizēja syutēja a - ha tod nū - nūskreja ▶ no tuo i atgruo - i pajēme meitu vot i navajadzēja jai iz tom šahtom braukt ▶ <i>io8-F:</i> a kai jius nūklivot Krīveji ? ▶ <i>F124-1929:</i> nu tāvs bāga nu vuocišim ▶ nabeja vēl parteji īleids bet jau kandidatūs ▶ pats bogotuokis i pats komūnists beja īstuostēja kai ▶ nu tāvs beja jis redzi tymā ▶	⬆ □
	Person ID: F124-1929 Year of birth: 1929 Sex: f Place: Cibla	
	Source: L	
▶ 0:00 / 0:00 — 🔊 ⋮	nu tāvs beja jis redzi tymā	
	Person ID: F124-1929 Year of birth: 1929 Sex: f Place: Cibla	▼ □
	Source: L	

plānota jaunu latgaliešu runas datu ieguve (intervijas iepriekš neaptvertos Latvijas novados, radio un TV ieraksti), tekstu transkribēšana un ievietošana mutvārdu korpasā. Viena laikus tiek plānoti arī turpmākie korpasa pilnveides posmi, izvērtējot transkribēto tekstu morfoloģiskās marķēšanas iespējas.

- Adolphs, Svenja, Dawn, Knoght (2010). Building a spoken corpus: what are the basics? O'Keeffe, Anne, McCarthy, Michael (eds.). *The Routledge Handbook of Corpus Linguistics*. London, New York: Routledge Taylor & Francis group, pp. 38–52.
- Anderson, Jean, Beavan, Dave, Kay, Christian (2007). SCOTS: Scottish corpus of texts and speech. Beal, Joan, Corrigan, Karen, Moisl, Hermann (eds.). *Creating and digitizing language corpora. Volume 1: Synchronic databases*. Basingstoke: Palgrave Macmillan, pp. 17–34. Available: <https://core.ac.uk/download/pdf/110386.pdf> [accessed 09.11.2021.].
- Anderwald, Lieselotte, Wagner, Susanne (2007). FRED – The Freiburg English Dialect Corpus: Applying corpus-linguistic research tools to the analysis of dialect data. Beal, Joan, Corrigan, Karen, Moisl, Hermann (eds.). *Creating and digitizing language corpora. Volume 1: Synchronic databases*. Basingstoke: Palgrave Macmillan, pp. 35–53. Available: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.519.4935&rep=rep1&type=pdf> [accessed 09.11.2021.].
- Biber, Douglas (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, No. 8 (4), pp. 243–257. Available: <http://otl-pl.philol.msu.ru/media/biber930.pdf> [accessed 09.11.2021.].
- Burnard, Lou (2005). Metadata for Corpus Work. Wynne, Martin (ed.). *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxford Books, pp. 30–46. Available: <http://ota.ox.ac.uk/documents/creating/dlc/> [accessed 09.11.2021.].
- Grochola-Szczepanek, Helena, Górski L. Rafał, von Waldenfels, Ruprecht, Woźniak, Michał (2019). Korpus języka mówionego mieszkańców Spisza. *LingVaria*, No. 14(27), s. 165–180. Available: <https://w.akademicka.pl/ojs/lv/article/view/727/712> [accessed 14.11.2021.].
- Hernández, Nuria (2006). *User's Guide to F R E D Freiburg Corpus of English Dialects*. Freiburg: English Dialects Research Group, Albert-Ludwigs-Universität. Available: <https://freidok.uni-freiburg.de/fedora/objects/freidok:2489/datastreams/FILE1/content> [accessed 14.11.2021.].
- Komrsková, Zuzana, Kopřivová, Marie, Lukeš, David, Poukarová, Petra, Golánová, Hana (2017). New Spoken Corpora of Czech: ORTOFON and DIALEKT. *Journal of Linguistics/Jazykovedný časopis*, No. 68(2), pp. 219–228. Available: <https://doi.org/10.1515/jazcas-2017-0031> [accessed 14.11.2021.].
- Moseley, Christopher (ed.) (2010). *Atlas of the World's Languages in Danger*, 3rd edn. Paris, UNESCO Publishing. Available: <http://www.unesco.org/culture/en/endangeredlanguages/atlas> [accessed 14.11.2021.].
- Pasikowska-Schnass, Magdalena (2016). Regional and minority languages in the European Union. *European Parliamentary Research Service*. September 2016. Available: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589794/EPRS_BRI\(2016\)589794_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589794/EPRS_BRI(2016)589794_EN.pdf) [accessed 14.11.2021.].
- Pinnis, Mārcis, Auziņa, Ilze, Goba, Kārlis (2014). Designing the Latvian Speech Recognition Corpus. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik: European Language Resources Association (ELRA), pp. 1547–1553. Available: http://www.lrec-conf.org/proceedings/lrec2014/pdf/284_Paper.pdf [accessed 09.11.2021.].
- Pošeko, Solvita (2016). Latgaliešu valodas attīstība. Lauze, Linda (zin. red.). *Valodas situācija Latvijā: 2010–2015. Sociolingvistisks pētījums*. Rīga: LVA, 173.–194. lpp.
- Riefler, Michael, Wilbur, Joshua (2017). Documenting endangered oral histories of the Arctic: A proposed symbiosis for endangered language documentation and oral history research, illustrated by Saami and Komi examples. Kasten, Erich, Roller, Katja, Wilbur, Joshua (eds.). *Oral History Meets Linguistics*. Fürstenberg, Havel: Kulturstiftung Sibirien, pp. 31–64. Available: https://www.researchgate.net/publication/314577167_Documenting_endangered_oral_histories_of_the_Arctic_A_proposed_symbiosis_for_language_documentation_and_oral_history_research_illustrated_by_Saami_and_Komi_examples [accessed 14.11.2021.].
- Stefanowitsch, Anatol (2020). *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press. Available: https://books.google.lv/books?id=3ZHeDwAAQBAJ&printsec=frontcover&hl=lv&source=gbg_summary_r&cad=0#v=onepage&q&f=true [accessed 14.11.2021.].
- Tauginienē, Loreta, Ojsteršek, Milan, Foltýnek, Tomáš, Marino, Franca, Cosentino, Marco, Gaizauskaitė, Inga, Glendinning, Irene, Sivasubramaniam, Shiva, Razi, Salim, Ribeiro Laura, Odiņeca, Tatjana,

- Trevisiol, Oliver (2019). *Akadēmiskā godīguma vispārējās vadlīnijas*. Pieejams: https://www.academicintegrity.eu/wp/wp-content/uploads/2019/10/RED_Guidelines_RTU_VS_amended_v2.pdf [skatīts 09.11.2021.].
- Thompson, Paul (2005). Spoken language corpora. Martin Wynne (ed.). *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, pp. 59–70. Available: <http://ota.ox.ac.uk/documents/creating/dlc/> [accessed 09.11.2021.]
- von Waldenfels, Ruprecht, Woźniak, Michał (2016). SpoCo – a simple and adaptable web interface for dialect corpora. Kupietz, Marc, Geyken, Alexander (eds.). *Journal for Language Technology and Computational Linguistics. Corpus Linguistic Software Tools*, No. 31 (1), pp. 155–170. Available: https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/6218/file/Journal_for_language_technology_and_computational_linguistics_1_2016.pdf [accessed 14.11.2021.].
- Vuković, Teodora (2021). Representing variation in a spoken corpus of an endangered dialect: the case of Torlak. *Lang Resources & Evaluation*, No. 55, pp. 731–756. Available: Representing variation in a spoken corpus of an endangered dialect: the case of Torlak | SpringerLink [accessed 25.10.2021.].

Creation of Contemporary Latgalian Speech Corpus in the Context of Documenting Lesser Used Languages

Angelika Juško-Štekele, Antra Kļavinska

Keywords: corpus linguistics, representativeness, corpus design, metadata, transcription, convention

According to data of UNESCO, in 2013, Latgalian language with 150,000 users was recognised as one of the world's endangered and vulnerable languages, as all generations still use the oral form, but the sustainability of the language is seriously jeopardised, since the number of young language users decreases. Pursuant to the EU directives and recommendations for preservation, research and development of regional and endangered languages, as well as the Guidelines for the State Language Policy 2021–2027 regarding development, disclosure on the web and accessibility of varied text corpus, in 2020, a group of researchers of the Rēzekne Academy of Technologies in the Project of State Research Programme *Digital Resources of Humanities: Integration and Development* (No. VPP-IZM-DH-2020/1-0001) started its work on the development of the Contemporary Latgalian Speech Corpus (MuLaR) aimed at the documentation, research, studies and acquisition of Latgalian.

The aim of the article is to identify and analyse the issues that are important in the process of creating MuLaR, applying the referential analysis of the scientific literature and comparative methodology. In turn, applying the analytical-synthetic method and based on the experience accumulated by the corpus creators, there was developed an initial model for the corpus architectonics and technological solutions, covering such issues as ensuring a representative Latgalian speech corpus, bearing in mind the territorial distribution of Latgalian language communities and diversity of Latgalian patois; the most appropriate methods to document natural, spontaneous language: collection of new data, opportunities to use the existing recordings (interviews, TV, radio broadcasts, field research data collections), other databases (reiti.rta.lv); understanding metadata; ethical aspects of the speech corpus; transcribing (software, conventions to reveal the features of spoken text as accurately as possible); creation of an accessible, easy-to-use open-access platform, using the experience of creating oral speech corpora for lesser-used languages / dialects in other countries. The article declares the main challenges for the corpus development after the initial validation of the corpus data, including in relation to the morphological tagging possibilities of the corpus.

Specializēta latviešu valodas runas korpusa un izrunas vārdnīcas izveide vizuālās diagnostikas izmeklējumu lingvistiskai analīzei un sistemātiskai transkribēšanai

Ilze Auziņa, Roberts Dargis, Baiba Saulīte,
Normunds Grūzītis, Mikus Grasmanis,
Andrejs Spektors, Kaspars Stepanovs

Šis darbs ir daļa no ERAF praktiskas ievirzes pētījumu projekta "Latviešu valodas runas atpazīšana un sintēze medicīnas lietojumiem" (Nr. 1.1.1.1/18/A/153) rezultātiem sinerģijā ar Valsts pētījumu programmas "Humanitāro zinātņu digitālie resursi" projektu Nr. VPP-IZM-DH-2020/1-0001.

Ievads

Latvijā dažādās medicīnas nozarēs, t. sk. diagnostiskajā radioloģijā, plaši tiek izmantotas mūsdienīgas medicīnas tehnoloģijas, tomēr izmeklējumu rezultātu apraksti un epikrīzes vēl aizvien tiek sagatavotas pilnībā manuāli, izmantojot diktofonu centru pakalpojumus. Taču, pastāvīgi pieaugot diagnostisko izmeklējumu skaitam, ārstiem un pacientiem izmeklējumu rezultātu apraksti jāgaida pat vairākas dienas, ja vien rezultāti nav ļoti steidzami. Ir nepieciešams jauns un ilgtspējīgs risinājums un tehniskā infrastruktūra, kas nodrošinātu latviešu valodas runas tehnoloģiju integrēšanu medicīnā un ļautu atvieglot un paātrināt izmeklējumu rezultātu aprakstu ieguvu.

LU Matemātikas un informātikas institūtā (LU MII), sadarbojoties ar Rīgas Austrumu klīniskās universitātes slimnīcu (RAKUS), tiek īstenots Eiropas Reģionālās attīstības fonda praktiskas ievirzes pētījums “Latviešu valodas runas atpazīšana un sintēze medicīnas lietojumiem”. Projekta mērķis ir radīt būtiskus latviešu valodas resursus (t. sk. specializētu latviešu valodas runas korpusu un izrunas vārdnīcu) un tehnoloģiju komponentes automatiskai runas atpazīšanai un runas sintēzei medicīnas jomā, kā arī demonstrēt šo resursu un tehnoloģiju izmantošanu, izstrādājot un aprobējot izmeklējumu diktēšanas un transkribēšanas platformu diagnostiskās radioloģijas vajadzībām.

Platforma būs pielāgota diviem lietojuma scenārijiem. Pirmajā scenārijā paredzēts, ka noteiktiem radioloģijas izmeklējumiem, piemēram, rentgenam un ultrasonogrāfijai, kuru apraksti ir relatīvi vienkārši un kuru diktātu automatiska atpazīšana ir precīzāka (salīdzinājumā ar, piemēram, datortomogrāfijas un magnētiskās rezonanses aprakstiem), ārsti izmantos diktēšanas platformu. Tajā pēc diktāta automatiskas atpazīšanas viņi paši tekstu varēs rediģēt un apstiprināt, tādējādi ātri iegūstot jau gatavu izmeklējuma aprakstu. Savukārt otrajā scenārijā paredzēts, ka diktofonu centra darbinieki pārskatīs un rediģēs automatiski atpazītus diktātus, nevis rakstīs visu no jauna.

Līdzīgi kā daudzās citās valodās, arī latviešu valodai pēc pietiekami liela vispārīga runas korpusa izveides (Pinnis et al. 2014) ir izstrādātas vairākas automatiskas runas atpazīšanas (*Automatic Speech Recognition*; turpmāk tekstā – ASR) sistēmas (Salimbajevs & Strigins 2015; Znotiņš et al. 2015). Tomēr šādas sistēmas nav izmantojamas medicīnā (t. sk. radioloģijā), jo tās ir paredzētas vispārīgam lietojumam un nespēj pietiekami precīzi atpazīt specifiskus tekstus, konkrēti – radioloģisko izmeklējumu un epikrīžu diktātus.

Lai varētu izstrādāt medicīnā izmantojamu ASR sistēmu, ir jāsagatavo specializēti valodu resursi un jāveic papildu priekšapstrādes un pēcapstrādes darbības:

- valodas modeļa izstrādei ir nepieciešams anonimizētu medicīnisku ziņojumu, galvenokārt radioloģisko izmeklējumu aprakstu un epikrīzu, tekstu korpusu (sk. nodaļu “Teksta korpusa izveide”);
- ortogrāfiski transkribēts un anonimizēts runas korpus, kas nepieciešams ASR sistēmas novērtēšanai un teksta priekšapstrādes un pēcapstrādes likumu izstrādei (sk. nodaļu “Runas korpusa izveide”);
- medicīnas terminu (t. sk. latīnisko), saīsinājumu, nosaukto entitāšu, kā arī vispārlietojamās leksikas izrunas vārdnīca (sk. nodaļu “Izrunas vārdnīca”);
- datu priekšapstrādes un pēcapstrādes likumi, kuri nodrošina, ka ziņojuma tekstā automātiski tiek izvērsti saīsinājumi, cipari tiek pierakstīti ar vārdiem u. tml. un automātiski transkribētajā tekstā tiek iekļauti atbilstoši apzīmējumi, saīsinājumi u. tml. (sk. nodaļu “Teksta izvēšanas un savēršanas gramatika”).

Akustiskā modeļa pielāgošana medicīnas tekstiem var nedaudz ietekmēt ASR transkripciju precizitāti, tomēr ASR precizitāti galvenokārt ietekmē pielāgotais valodas modelis un specializētā izrunas vārdnīca (Blackley et al. 2019).

Šajā pētījumā specializēto korpusu un izrunas vārdnīcas izejas materiāls ir anonimizēti RAKUS radioloģijas izmeklējumu arhīva dati: audioieraksti (radioloģisko izmeklējumu un epikrīzu audio ieraksti) un teksti (radioloģisko izmeklējumu un epikrīzu apraksti).

Teksta korpusa izveide

Lai izstrādātu medicīnas jomas korpusus, kas izmantoti par pamatu tālākiem pētījumiem un sarežģītāku resursu izstrādei, šajā projektā pielāgotas esošās datu kopas. Korpusos iekļauti dati no lielākās slimnīcas Latvijā, t. i., RAKUS, radioloģijas izmeklējumu arhīva.

Medicīnas valodu reprezentējošais teksta korpus veidots, izmantojot anonimizētus izmeklējumu, operāciju un epikrīzu aprakstus no RAKUS arhīva. Teksta korpusa izveide ietver trīs galvenos soļus:

1. teksta izguve;
2. teksta anonimizēšana;
3. teksta normalizēšana.

Šobrīd runas un teksta datu plūsma RAKUS ir šāda:

- radiologi, ārsti un viņu palīgi diktē izmeklējumu vai epikrīzi, izmantojot dažādas ierīces un kanālus, t. sk. telefona zvanus;
- diktāta audioieraksti tiek iesniegti diktofonu centrā;
- audioieraksti tiek manuāli pārrakstīti, izveidojot *Microsoft Word* dokumentus, kuru pamatā ir iepriekš definētas dokumentu veidnes;
- sagatavotie dokumenti tiek nosūtīti ārstiem apstiprināšanai;
- apstiprinātie dokumenti tiek pievienoti arhīvam.

Lai izvairītos no sensitīvu datu nonākšanas ārpus ārstniecības iestādes, pirmie divi soļi veikti RAKUS IT infrastruktūrā. Teksta segmenti izgūti no vairāk nekā 100 tūkstošiem dažādu veidu medicīnisko izrakstu, kas RAKUS tikuši sagatavoti pēdējo astoņu gadu laikā. Sagatavoto izrakstu skaits katru gadu pieaug un šobrīd sasniedz ap 15 tūkstošiem izrakstu gadā.

Teksta izguve

Ārstniecības iestādēs izveidotajos *MS Word* dokumentos bieži ir vairākas savstarpēji pakārtotas tabulas un dažādi teksta formatējuma elementi, piemēram, treknraksts, slīpraksts, pasvītrojums. Teksta izguvē ir svarīgi saglabāt teksta dalījumu teikumos un rindkopās, jo vairāki teksta analīzes rīki (piemēram, morfoloģiskās marķēšanas un sintaktiskās analīzes rīki) paļaujas uz korektu teksta dalījumu teikumos un nekorekts dalījums var būtiski ietekmēt šo rīku precizitāti. Dalījums teikumos un rindkopās tiek izmantots, arī lai apmācītu valodas modeli, kas automātiski ģenerētās diktātu transkripcijas sadala teikumos un rindkopās.

Jāņem vērā arī tas, ka medicīniskajos izrakstos ir dažādi īsi segmenti bez pieturzīmēm, piemēram, sadaļu virsraksti, tabulu, kolonnu, rindu un citu teksta lauku nosaukumi. Ja visu tekstu apvienotu vienā un dalītu teikumos tikai pēc pieturzīmēm, šādi īsie segmenti nekorekti tiktu pievienoti teikumiem, kas ir aiz tiem. Pretēja rakstura problēma rastos, ja katru teksta segmentu uzskatītu par atsevišķu rindkopu. Šādā gadījumā īsie teksta segmenti, t. sk. specifiski noformētie segmenti (treknraksts, kursīvs u. c.), tiktu uzskatīti par atsevišķām rindkopām un daudzi teikumi tiktu sadalīti vairākās neloģiskās, īsās rindkopās.

Teksta izguvei no RAKUS izrakstu arhīva tika izstrādāts pielāgots rīks, kas segmentus izgūst, balstoties uz dokumenta iekšējo XML datu struktūru (*Microsoft Word Open XML* jeb *DOCX* formāts, kas RAKUS izrakstos tiek lietots kopš 2010. gada).

Teksta anonimizēšana

Nākamais būtiskais solis ir izgūtā teksta anonimizēšana, lai izvairītos no personas datu (vārds, uzvārds, personas kods, adrese u. c.) iekļaušanas korpusā, pat ja tam nav plānota publiska

piekļuve. Kaut gan teksts ir automātiski anonimizēts, šis korpuss joprojām tiek uzskatīts par potenciāli sensitīviem datiem, jo korpusa lielā apjoma dēļ to nav iespējams manuāli caurskatīt un pārbaudīt.

Izgūtais teksta apjoms ir pietiekami liels tālākajiem pētījumiem (medicīnas valodas modelēšanai), tāpēc teksta anonimizēšana tika izvirzīta par augstāku prioritāti nekā teksta saglabāšana, proti, ja bija vismazākās šaubas, ka teksta segments varētu saturēt personas datus, tas tika izņemts no korpusa, pat ja pastāvēja liela varbūtība, ka šajā segmentā personas datu, visticamāk, nemaz nav.

Teksts tika anonimizēts segmentu (aptuveni: rindkopu) līmenī. Ja segmentā konstatēts kaut viens vārds, kas varētu norādīt uz potenciāliem personas datiem, viss segments tika izņemts no teksta korpusa. Papildus tam pēc tekstu normalizēšanas (skat. nākamo sadaļu) visas atlikušās rindkopas teksta korpusā ir sadalītas teikumos, bet teikumi ir sakārtoti alfabētiskā secībā. Tādējādi ir pilnībā sajaukta rindkopu un teikumu sākotnējā secība: pat ja kāds izteikums satur kādus personas datus, tiem nav pieejams plašāks konteksts.

Teksta normalizēšana

Tekstu korpusa normalizēšana notiek vairākos posmos. Vispirms automātiski tiek labotas tipiskās drukas un pareizrakstības kļūdas. Pirmajā korpusa apstrādes iterācijā šādu kļūdu labošana vēl nebija iespējama, bet pēc pirmās iterācijas tika izgūts tekstu korpusa vārdformu biežumsaraksts un biežāk lietotās vārdformas, par kuru pareizrakstību un izrunu nebija iespējams pārlielināties automātiski (sk. nodaļu “Izrunas vārdnīca”), un tās tika caurskatītas manuāli, tostarp kļūdainajām vārdformām norādot korektu rakstību (sk. nodaļu “Izrunas vārdnīca”). Rakstības normalizēšana ilustrēta ar diviem ļoti izplatītiem piemēriem RAKUS korpusā:

- 1) termins “dupleksdoplerogrāfija” izrakstos tiek lietots visdažādākajos rakstības variantos, t. sk. ar pareizrakstības kļūdām: “dupleks-doplerogrāfija”, “dupleks doplerogrāfija”, “duplekss doplerogrāfija”, “duplex-doplerogrāfija”, “duplex – doplerogrāfija” utt. – tie visi (visās formās) tika automātiski normalizēti kā “dupleksdoplerogrāfija” (attiecīgajās formās);
- 2) savukārt termins “folikuls” tiek bieži lietots ar vienu vai divām pareizrakstības kļūdām: “folikuls”, “folikuļi”, “folīkuļi” utt. – kļūdainās formas korpusā tika automātiski izlabotas.

Nākamais solis pēc dažāda veida rakstības un segmentācijas kļūdu labošanas un teksta normalizēšanas ir teksta automātiska sadalīšana teikumos un tekstvienībās (vārdformas, akronīmi un saīsinājumi, simboli un apzīmējumi, interpunkcija). Strukturāli sadalītajam tekstam tika veikta morfoloģiskā analīze, izmantojot LU MII atvērtā pirmkoda rīkoku NLP-PIPE (Paikens et al. 2013; Znotiņš & Cīrule 2018). Tādējādi tekstu korpuss ir papildināts ar informāciju par vārdlietojumu pamatformām un morfoloģiskajām pazīmēm, kas ir noderīgi korpusa tālākā lingvistiskajā analīzē, piemēram, pamatformu pārbaudei un lietojuma biežuma salīdzināšanai citos latviešu valodas korpusos un leksiskas resursos.

Trešais būtiskais normalizēšanas apakšsolis ir tekstu korpusa automātiska izvēšana (verbalizēšana): skaitļu, datumu, dažādu saīsinājumu un apzīmējumu pārrakstīšana pilnos vārdos, kontekstuāli saskaņojot locījumus. Papildus tam gadījuma vietās tiek verbalizētas arī pieturzīmes (piemēram, komats tiek aizvietots ar “komats”, punkts – ar “punkts” vai “nākamajā teikumā”) un strukturālais noformējums (piemēram, rindkopu pārnesumi – “jaunā rindā” u. tml.; sadaļu virsraksti – “secinājumos rakstām”, “slēdzienā [lūdzu] rakstām” u. tml.). Tādējādi izvērstajā tekstu korpusā tiek reprezentētas arī tipiskās instrukcijas un to biežāk lietotie, alternatīvie verbalizācijas varianti, ko ārsti līdzšinējā darbplūsmā mēdz norādīt dikt fonu centra operatoriem. Tas tiek darīts, lai teksta korpuss būtu pēc iespējas līdzīgāks runas korpusam (sk. nodaļu “Runas korpusa izveide”) un lai runas atpazīšanas sistēmas valodas modelis pēc iespējas labāk reprezentētu to, kā apraksti un izraksti tiek diktēti. Šis apakšsolis un tam pretējais virziens (transkripciju automātiska savēršana) detalizētāk ir aprakstīts nodaļā “Teksta izvēšanas un savēršanas gramatika”.

Teksta normalizēšanai un tipisko kļūdu novēršanai ir būtiska loma turpmākajā valodas statistiskā modeļa izveidē un izmantošanā. Lai gan visas pareizrakstības kļūdas izlabot nav iespējams, svarīgi ir novērst tipiskās. Tā kā RAKUS korpuss ir ļoti apjomīgs (vairāk nekā 100 milj. tekstvienību), manuāli analizētas un automātiski labotas tikai tās kļūdas un nekonsekvences, kuru biežums korpusā ir vismaz 1000 (t. i., relatīvais biežums vismaz 0,00001). Par retākiem gadījumiem: tiek pieņemts, ka valodas modelim, kas runas atpazīšanas sistēmai palīdz ģenerēt diktātu automātiskās transkripcijas, būs tendence izvēlēties pareizākus rakstības variantus, t. i., tiek pieņemts, ka normalizētajā korpusā pastāv spēcīga korelācija starp statistiski biežākiem un leksiski pareizākiem vārdlietojumu kontekstiem (n-grammām). Lai gan praksē tā ne vienmēr notiek, sistēmas prototipa testēšanā iesaistītie ārsti norāda, ka automātiskajās transkripcijās bieži vien vārdi tiek pierakstīti pareizāk, nekā viņi tos ir diktējuši.

Runas korpusa izveide

Lielais teksta korpusa apjoms ļoti labi atspoguļo rakstītā teksta dažādību. Diemžēl ar to vien ir par maz, lai izveidotu izrunas vārdnīcas. Tekstā ir sastopami daudzi svešvārdi un saīsinājumi, kuru izruna nav viennozīmīga. Lai sīkāk varētu pētīt izrunas nianšes, ir nepieciešams runas korpuss, tāpēc projektā tika izstrādāts 30 stundu ortogrāfiski transkribēts runas korpuss.

Runas korpusa izstrāde sastāv no vairākiem soļiem:

1. runas datu atlase;
2. runas datu automatizēta apstrāde un anonimizēšana;
3. automātiski atpazītā teksta pārskatīšana, rediģēšana un marķēšana.

Runas datu atlase

Runas dati tika ņemti no pēdējo sešu mēnešu diktofonu centra arhīva. Lai nodrošinātu korpusa kvalitāti, tajā netika iekļauti ieraksti ar zemu kvalitāti. Korpusā netika iekļauti diktāti, kas veikti pa telefonu, ieraksti ar skaļu fona troksni vai ieraksti, kurus ierunājis cilvēks ar spēcīgu akcentu un daudzām izrunas kļūdām.

Lai nodrošinātu runas korpusa reprezentativitāti, proporcionāli tika atlasīti dažādu runātāju ieraksti. Korpusā tika iekļauti ieraksti no 71 runātāja, atlasot ierakstus vienmērīgi no visa perioda.

Automātiska runas datu apstrāde un anonimizēšana

Lai paātrinātu runas korpusa izveidi, no izrakstiem izgūtais teksts tika automātiski sastatīts ar ierakstu. Automātiskais sastatījums nav precīzs, tāpēc automātiski atpazītais teksts tika pārskatīts un manuāli rediģēts.

Audiodatu anonimizēšana ir konceptuāli un tehniski sarežģītāka nekā teksta anonimizēšana. Lai anonimizētu audiofailu (diktātu), vispirms būtu jāiegūst tā automātiska transkripcija un jānosaka, kuros laika intervālos audiofailā ir pieminēti personas dati, attiecīgos segmentus no audiofaila pēc tam izgriežot. Tomēr audiodatu anonimizēšanā nedrīkstam paļauties uz automātiskām transkripcijām, kas ir salīdzinoši neprecīzas tieši attiecībā uz personas datiem (jo šāda veida dati ir apzināti izņemti no valodas modeļa un izrunas vārdnīcas; sk. nodaļu “Teksta korpusa izveide” un “Izrunas vārdnīca”). Šāda pieeja – anonimizēt tikai atsevišķus, īsus audio segmentus, balstoties automātiskajās transkripcijās, – nenodrošinātu pietiekamu anonimizēšanas precizitāti.

Vairumā gadījumu personas dati tiek diktēti audioieraksta sākumā, tāpēc tika izmantota vienkāršāka un anonimizēšanas ziņā uzticamāka metode, lai arī šādā veidā audiofails bieži vien tiek apgriezts vairāk, nekā tas būtu nepieciešams. Automātisko transkripciju vietā tika izmantoti secīgi teksta segmenti no anonimizētā, bet alfabētiski nesajauktā tekstu korpusa (sk. nodaļu “Runas datu atlase”). Izmantojot anonimizētā teksta un audioieraksta automātisku sastatījumu, atlasītajiem audiofailiem automātiski tika nogriezts viss nesastatītais sākums un beigas.

Manuāla runas datu transkribēšana un marķēšana

Radioloģisko izmeklējumu un epikrīžu apraksti tiek sastatīti ar atbilstošu audio ierakstu, kuram, izmantojot ASR sistēmu, tiek piedāvāts aptuvens atšifrējums. Klausoties diktāta ierakstu, diktāta transkripcija tiek pierakstīta atbilstoši izstrādātajām vadlīnijām. Diktāta transkribēšanas režīmā tiek pārbaudīts, vai transkripcijā ir pierakstīts viss, kas diktātā tiek teikts, un vai tiek pierakstīts tā, kā tas tiek pateikts (sk. 1. attēlu).

Runas korpusā visi automātiski atpazītie teksti ir pārskatīti un pierakstīti t. s. ortogrāfiskajā transkripcijā. Ortogrāfiskā transkripcija ir burtiska sacītā atveide rakstos mašīnlasāmā formā,

Annotate Admin

ID 4

☐ Pacienta dati
☐ Skati trokšņi / neskaidra runa
☐ Nesaisītā runa

Akcenta*
Bez akcenta

Modalitāte
MR - Magnetic Resonance

Nozare
13. Urogenitālā radioloģija

Parādīt/Pasliptot referenci

1
2 MR [em er] prostatai ar kontrastvielu.
3
4 prostata trīs komats deviņi reiz trīs komats trīs reiz trīs komats astoņi centimetri. tilpums divdesmit seši komats piecpadsmit kubikcentimetri. PSA [pē es ā] līmenis nav ziņams. stāvoklis pēc TUR TUR-P [tur pē] operācijas ar tipisku defektu ar tipisku defektu. dziedzera pārejas daļā redzami labdabīgas hiperplāzijas mezgli. audus ar difūzijas ierobežojumu dotajā izmeklējumā nesaskatu. sēklas pūslīši simetriski bez strukturālām izmaiņām. urīnpūslis gludām sienām bez intralumināliem ieslēgumiem. patoloģiska lieluma limfmezglus nesaskatu.
5 (un rakstām [rakstām] lūdzu slēdzienu)
6 labdabīga prostatas hiperplāzija. stāvoklis pēc TUR-P [tur pē] operācijas. pašreiz nav datu par augstas malignitātes tumora audiemi prostatā. (paldies)

Komentāri

☒ Pabeigts

Save Save & Next

00:00 01:32

1. attēls. Diktāta transkribēšanas saskarne.

ievērojot valodas ortogrāfijas principus (Oostdijk & Boves 2008: 644; Schiel & Draxler 2004; Goedertier et al. 2000: 909). Atveidojot runātās valodas piemērus rakstu formā, tiek saglabāts runas fakta autentiskums, t. i., tiek rakstīts tieši tā, kā tiek teikts, paturot visus vārdu atkārtojumus, ja nepieciešams, parādot arī vārdu izrunas atkāpes no literārās valodas normas. Teksti tiek pierakstīti ortogrāfiskajā transkripcijā, izmantojot korpusa transkribēšanas vadlīnijas, kas balstītas uz iepriekšējo pieredzi (Pinnis et al. 2014) un papildinātas atbilstoši specializētā runas korpusa vajadzībām, piemēram, norādot atbilstošu terminu, saīsinājumu izrunu. Transkripcijā netiek norādītas pauzes – klusuma pauzes un aizpildītās pauzes (*āā, ēē, em klm* u. tml.), tāpat netiek norādīti neverbālie elementi, piemēram, ieelpa un izelpa, cilvēka radīts fizioloģisks troksnis – mēles klakšķināšana, šņaukšanās, čapstināšana u. tml. Šī informācija jau ir uzkrāta, izmantojot “Latviešu valodas runas atpazīšanas korpusu” (Pinnis et al. 2014), un tiek izmantota ASR sistēmu izstrādē.

Ortogrāfiski marķējot runas audiosignālu, tiek ievēroti vairāki ortogrāfiskās transkripcijas pamatprincipi: (1) runātais tiek pierakstīts burtiski, norādot arī vārdu atkārtojumus, pārteikšanos u. tml.; (2) pēc iespējas tiek ievērota interpunkcija un pareizrakstība; (3) norādītas būtiskas atkāpes no pareizrūnas normām, piemēram, *Aknu izmērs četrdesmit [čēsmīt] milimetri, tālāk rakstām [rakstām]*; (4) visi vārdi, ja vien tie nav īpašvārdi vai akronīmi, tiek rakstīti ar mazajiem burtiem; (5) cipari, saīsinājumi tiek rakstīti ar vārdiem atbilstoši to izrunai audioierakstā.

Nekorekta izruna

Nekorektai vārdu izrunai var būt dažādi iemesli, piemēram, latviešu valoda nav runātāja dzimtā valoda, runā vērojama latviešu valodas izlokšņu ietekme. Tas var būt saistīts arī ar īpatnēju atsevišķu vārdu izrunu, kas raksturīga katram indivīdam. Samērā bieži neprecīzi tiek izrunāti skaitļa vārdi, piemēram, *četrdesmit* [čēsmīt], *piecdesmit* [piesmīt], *divi* [div]. Atkāpes no literārās valodas normām vērojamas arī patskaņu kvantitatē, t. sk. noteiktās darbības vārdu formās, piemēram *rakstām* [rak-stam], kā arī patskaņu un līdzskaņu kvalitātē, piemēram, *dorsāls* [dorzāls], *patoloģija* [patologija].

Saīsinājumi un svešvārdi

Ja vārds tiek izrunāts kā saīsinājums (t. sk. mērvienība) vai svešvārds (latīnisms, zāļu nosaukums u. tml.), tad tā izruna tiek likta kvadrātiekvās, piemēram:

USG [ū es gā] veikta trešajā martā
RTG [er tē gā] gramma krūškurvīm taisnā projekcijā
Spinālā kanāla salīdzinoša stenoze *C3* [cē trīs] *C7* [cē septiņi] zonā –
 uzskatāmāk *C6* [cē seši] *C7* [cē septiņi] *dextra* [dekstra].

Ja terminu vai nosaukumu veido vairāki vārdi, izruna tiek norādīta aiz katra vārda, kuram tiek precizēta izruna, piemēram:

Lietot Deferasirox [deferasiroks] *Mylan* [milan] reizi dienā
 ..*truncus* [trunkus] *coeliacus* [celiakus] līmenī..
distālais gals vena cava [kava] *superior* [supērior] projekcijā

Ja mērvienību apzīmējumus (piemēram, *cm*, *s*, *g*, *mg*, *mm*, *diam.*) izrunā kā pilnu vārdu, tad raksta pilnu vārdu, piemēram:

Necelt smagumus vairāk par pieciem kilogramiem aptuveni mēnesi.
Necelt smagumus vairāk par pieci kilogrami aptuveni mēnesi.
 ..*postinfarkta zona ar diametru divdesmit divi reiz piecpadsmit milimetri.*

Ja svešvārds vai saīsinājums tiek izrunāts tieši tā, kā pierakstīts, izrunu kvadrātiekvās nepievieno, piemēram, *luteum*, *TUR*.

Matemātiskie simboli

Arī matemātiskie simboli tiek rakstīti vārdiem, piemēram:

sin > *dxt* → *sinistra* vairāk kā *dextra* [dekstra]
Xefo 8 mg x 1 pēc vajadzības → *Xefo* [ksefo] *astoni miligrami* reiz viens pēc vajadzības

Cipari un datumi

Ciparus (arī decimāldaļskaitļus, daļskaitļus) raksta ar vārdiem tā, kā tie tiek izrunāti:

centrālajā zonā ar kopējo apjomu LL [el el] seši komats septiņi centimetri, AP [ā pē] pieci komats četri centimetri un CC [cē cē] septiņi cm [cē em]

Arī datums tiek rakstīts tā, kā tas tiek izrunāts:

*Pacientam plānots konsilijs nulle pirmajā nulle otrajā divtūkstoš deviņpadsmitajā.
Pacientam plānots konsilijs divtūkstoš deviņpadsmitā gada pirmajā februārī.
Izmeklējums veikts devītā pirmā divtūkstoš divdesmitā gadā.*

Speciāli simboli

Speciāli simboli tiek izmantoti, lai parādītu aprautus vārdus un nesaprotamu runu, stostīšanos. Ja vārdam netiek izrunātas beigas, tā vietā tiek likta zvaigznīte. Pareizā forma šajos gadījumos netiek pierakstīta, piemēram, *Lab* kreisajā pusē neliels izaugums*. Nesaprotama pārteikšanās un stostīšanās, ko nevar transkribēt kā vārdu(s) vai aprautu(s) vārdu(s), tiek atzīmēta ar restīti, piemēram, *pirmajā divtūkstoš divdesmitajā # {kur izmeklējums uzrakstām}*.

Formatēšanas komandas

Transkripcijā tiek norādītas arī teksta formatēšanas komandas, kas tiek marķētas, pamatojoties uz iepriekšējo darbu pie vispārējās nozīmes latviešu valodas diktēšanas korpusa izstrādes (Pinnis et al. 2016). Ārstu norādes par formatējumu, apraksta sākumu, gaitu vai beigām ir liktas figūr- iekavās, piemēram, *{nākamajā rindā lūdz rakstām}, {komats}, {rakstām tā}, {tas arī viss}, {paldies}, {paldies par šo visu}, {pirms slēdziena jāpieraksta, ka}, {pie slēdziena pierakstām}*.

2. attēlā parādīti sastatīti teksta un runas korpusa fragmenti: kreisajā pusē (A) dots teksta fragments no izmeklējumu arhīva, savukārt labajā pusē (B) – ortogrāfiski transkribēts diktāta fragments. Ortogrāfiskajā transkripcijā ir izvērsti saīsinājumi (*CT* → *datortomogrāfija*, *Kr.* → *kreisā* u. c.), norādītas formatēšanas komandas (*{un rakstām}*, *{komats}*, *{domuzīme}*, *{un tālāk rakstām}* u. c.) un precizēta terminu un vārdformu izruna (*oculi* [okuli], *opticus* [optikus], *diferencējas* [diferencējās] u. c.).

Ortogrāfiskā transkripcija tiek izmantota, lai kvantitatīvi novērtētu runas atpazīšanas sistēmas precizitāti, izstrādātu medicīnas jomai pielāgotāku akustisko modeli, papildinātu izrunas vārdnīcu ar vārdu, akronīmu, simbolu u. c. tekstvienību izrunas alternatīvām. Runas korpusā ļauj novērtēt, cik precīzi darbojas visi ASR komponenti: akustiskais modelis, izrunas vārdnīca un valodas modelis.

1 Izmeklējums

2 Orbitu un galvas CT.

3 Apraksts

4 Stāvoklis pēc kr. puses orbitas traumas saskaņā ar anamnēzi – orbitu veidojošo kaulu traumatiska rakstura bojājumi netiek konstatēti.

5 Bulbus oculi abās pusēs vienādi labi diferenciējas, priekšējās kamera saglabātas, lēcas luksāciju, subluksāciju nekonstatē. Intrabulbāri svešķermeņus vai hemorāģiju nekonstatē. Retrobulbārās telpas simetriskas, n. opticus diferenciējas. Acu grozītājmuskulji bez atzīmējamām izmaiņām. Kr. pusē prebulbāri mīksto audu tūska.

6 Galvas smadzeņu ventrikulārā sistēma simetriska, nav paplašināta. Vidusstruktūras centrālas.

7 Galvas smadzenēs perēkļus, intrakraniālu hemorāģiju vai tilpuma procesa pazīmes nekonstatē.

8 Intrakraniālie maģistrālie asinsvadi sklerotizēti.

9 Kortikālās rievas paplašinātas frontoparietālajās daļās.

10 Secinājumi

11 Stāvoklis pēc kr. puses orbitas traumas – orbitu veidojošo kaulu traumatiska rakstura bojājumi netiek konstatēti.

12 Deguna blakusdobumi labi pneimatizēti.

13 Galvas smadzenēs perēkļus, intrakraniālu hemorāģiju vai tilpuma procesa pazīmes nekonstatē.

A

1

2 orbitu un galvas datortomogrāfija.

3 (un rakstām)

4 stāvoklis pēc kreisās puses orbitas traumas saskaņā ar anamnēzi (komats) (domuzīme) orbitu veidojošo kaulu traumatiska rakstura bojājumi netiek konstatēti.

5 bulbus oculi [okuli] abās pusēs vienādi labi diferenciējas [diferenciējas], priekšējās kamera saglabātas, lēcas luksāciju, subluksāciju nekonstatē. intrabulbāri svešķermeņus vai hemorāģiju nekonstatē. retrobulbārās telpas simetriskas [komats] nervus opticus [optikus] diferenciējas [diferenciējas]. acu grozītājmuskulji bez atzīmējamām izmaiņām. kreisā pusē prebulbāri mīksto audu tūska [punkts]

6 (un tālāk rakstām) galvas smadzeņu ventrikulārā [ventikulārā] sistēma simetriska, nav paplašināta. vidusstruktūras centrālas.

7 galvas smadzenēs perēkļus, intrakraniālu [intra* kraniālu] hemorāģiju vai tilpumprocesa pazīmes nekonstatē.

8 intrakraniālie maģistrālie asinsvadi sklerotizēti.

9 kortikālās rievas paplašinātas frontoparietālās daļās.

10 (un rakstām slēdziens)

11 stāvoklis pēc kreisās puses orbitas traumas (komats) (domuzīme) orbitu veidojošo kaulu traumatiska rakstura bojājumi netiek konstatēti.

12 (tālāk rakstām) deguna blakusdobumi labi pneimatizēti.

13 galvas smadzenēs perēkļus, intrakraniālu hemorāģiju vai tilpuma procesa pazīmes netiek konstatētas [punkts]

B

2. attēls. Sastatīti references teksta (A) un runas korpusa (B) fragmenti.

Diktātu kategorizēšana

Atšifrējot diktātus, par katru audioierakstu tiek norādīta papildu informācija – izmeklējuma veids (modalitāte) un radioloģijas nozare (sk. 1. tabulu). Šobrīd korpusā iekļauti 900 dažādu veidu izmeklējumu apraksti: datortomogrāfijas izmeklējumi (CT), magnētiskās rezonanses izmeklējumi (MR), mamogrāfijas izmeklējumi (MG), rentgenogrammu apraksti (CR) un ultrasonogrāfijas izmeklējumi (US), kuru sagatavošanai tiek bieži izmantots RAKUS diktofonu centrs. Pašlaik dati tiek kategorizēti manuāli, un šo informāciju plānots izmantot, lai apmācītu automātisku izmeklējuma modalitāšu un nozaru klasifikatoru, kā arī lai noteiktu, kura izmeklējumos lietotā leksika ir kopīga visām modalitātēm un radioloģijas nozarēm, bet kura – specifiski katrai no modalitātēm un nozarēm. Tas ļaus arī papildināt izrunas vārdnīcu ar informāciju par vārdu lietojumu.

Modalitāte		Biežums	Nozare	Biežums
CT	Datortomogrāfija	489	Onkoloģiskā radioloģija	425
MR	Magnētiskā rezonanse	129	Torakālā radioloģija	130
MG	Mamogrāfija	113	Krūts radioloģija	117

CR	Rentgenogramma	109	Neiroradioloģija	67
US	Ultrasonogrāfija	60	Sirds un kardiovaskulārā radioloģija	59
			Gastrointestinālā un abdominālā radioloģija	54
			Uroģenitālā radioloģija	22
			Muskuloskeletālā radioloģija	15
			Galvas un kakla radioloģija	9
			Neatliekamā radioloģija	2

1. tabula. Modalitāšu un nozaru
īpatsvars (reprezentācija) runas korpusā.

Izrunas vārdnīca

Pēc teksta izvēšanas un automātiski iegūtās runas transkripcijas pārskatīšanas nākamais solis ir izrunas vārdnīcas izveide. Izrunas vārdnīcu veido tekstvienību (vārdu, saīsinājumu un abreviatūru u. tml.) saraksts un norādes par to izrunu, kas runas atpazīšanas ASR sistēmai būtu jāatpazīst, kā arī automātiski ģenerēta fonētiskā transkripcija. Tā drīzāk ir datu kopa, nevis izrunas vārdnīca tradicionālā izpratnē, kur katrai vārdnīcā iekļautajai vienībai ir pievienota minimāla gramatiskā informācija un fonētiskā transkripcija.

Izrunas vārdnīcā iekļaujamo vārdu saraksts tiek iegūts no tekstu korpusa: teksti tiek marķēti un no tiem tiek izgūtas visas unikālās tekstvienības. Sarakstā iekļautās tekstvienības var iedalīt četrās grupās:

- latviešu valodas vārdi, kas tiek izrunāti atbilstoši latviešu valodas pareizrunas normām (lielākā daļa vārdu),
- kļūdaini uzrakstīti vārdi, kurus nevajadzētu iekļaut izrunas vārdnīcā,
- saīsinājumi un simboli, kas jāizvērs, norādot to izrunu vai lasījumu,
- vārdi citās valodās (ne latviešu), piemēram, medicīnas termini latīņu valodā, zāļu nosaukumi, kuru izruna latviešu valodā atšķiras no rakstības – tiem jānodrošina fonētiskā transkripcija, izmantojot latviešu valodas fonēmas.

Kopumā specializētajā tekstu korpusā ir 1,8 miljoni tekstvienību, no kurām aptuveni 1,1 miljonom tekstvienību ir vismaz viens burts, tādēļ tās var uzskatīt par izrunas vārdnīcā iekļaujamām vienībām. Manuāla visu vārdu iedalīšana iepriekš minētajās četrās grupās būtu ļoti laikietilpīgs uzdevums, tāpēc tika izmantoti vairāki latviešu valodas resursi un rīki, lai sagrupētu lielāko daļu vārdu.

Vispirms vārdi, kas bija atrodamī lielākajā tiešsaistes latviešu valodas vārdnīcā *Tēzaurs.lv* (Spektors et al. 2016), tika atzīmēti kā standartvalodas vārdi. Vārdu locījumi tika atpazīti tikai tiem vārdiem, kuriem *Tēzaurs.lv* ir norādītas locīšanas paradigmas. Specializētajā medicīnas jomas korpusā bija sastopami arī tādi vārdi, kuri ir *Tēzaurs.lv*, bet kuriem nav norādītas locīšanas paradigmas. Lai varētu automātiski atpazīt vēl vairāk vārdu, šiem vārdiem locīšanas paradigmas tika pievienotas manuāli. Šī informācija savukārt tika integrēta *Tēzaurs.lv*.

Lai atpazītu vārdus un frāzes, kas, iespējams, ir personas dati, ir izmantots atvērtā pirmkoda latviešu valodas nosaukto entitāšu atpazīnējs – daļa no jau pieminētās NLP-PIPE rīkkopas (Znotiņš & Cīrule 2018). Īpašvārdi, kas uzskatāmi par personas datiem (vārds, uzvārds, adrese), tiek pusautomātiski atlasīti un izmantoti kā papildu filtrs teksta anonimizēšanā.

Vārdi, kas netika klasificēti automātiski, bija jāpārskata un jāsagrupē manuāli. Tekstvienības, kas teksta korpusā sastopamas vismaz 1000 reižu (aptuveni 14 tūkst.), tika atlasītas manuāli pārskatīšanai. Vārdi, kuru nav *Tēzaurs.lv* un kuri manuāli tika atzīmēti kā standartvalodas vārdi, ir pievienoti kandidātu sarakstam iekļaušanai *Tēzaurs.lv*. To vidū ir arī diktātos lietoto standartvalodas vārdu abreviatūras, piemēram, vārda *ultrasonogrāfija* abreviatūra USG, vārdu savienojuma *magnētiskā rezonanse* abreviatūra MG, vārda *rentgenogramma* abreviatūra RTG.

Atlasītajām tekstvienībām izrunas vārdnīcā tiek norādīti izrunas varianti, ja tie atšķiras no vispārpieņemtajām izrunas normām, vai tekstvienību (gk. abreviatūru vai citvalodu vārdu) lasījums. Manuāli pārskatot tekstvienību sarakstu, tiek norādīts, 1) vai vārds ir svešvalodā (pazīme – [svešvaloda], 2) vai vārds ir īpašvārds (pazīme – [personas dati]), 3) vai tas ir saīsinājums, abreviatūra vai matemātisks simbols (pazīme – [saīsinājums]), 4) vai vārds ir kļūdaini uzrakstīts (pazīme – [drukas kļūda]). Ja sarakstā iekļautajai tekstvienībai bez plašākas apkaimes nav iespējams norādīt izrunu vai saprast, vai tas ir saīsinājums, īpašvārds u. tml., tiek pievienota pazīme [konteksts], piemēram, *l, e, g, pac*.

Pazīme [personas dati] pievienota tad, ja tekstvienība ir īpašvārds – vārds, uzvārds, apdzīvotas vietas nosaukums u. tml., kas savukārt liek pārbaudīt, vai tā nav daļa no personas datiem, kam korpusā nevajadzētu parādīties.

Ja tekstvienībai ir pievienota pazīme [saīsinājums], tad parasti tiek norādīts gan lasījums/izruna, gan izvērsta forma, piemēram, *PVN* [pē vē en] – *pievienotās vērtības nodoklis*, *RAKUS* [rakus] – *Rīgas Austrumu klīniskās universitātes slimnīca*, *EKG* [ē kā gā] – *elektrokardiogramma*.

Vārdnīcā netiek izmantota fonētiskā transkripcija. Fonētiskā izruna starptautiskajā fonētiskajā alfabētā (IPA) tiek automātiski ģenerēta, izmantojot *Tēzaurs API*¹.

Izrunas vārdnīca un informācija par tekstvienību biežumu tiek izmantota valodas modeļa izstrādē. Reti sastopami vai kļūdaini vārdi un tekstvienības, kas potenciāli ir daļa no personas datiem, netiek iekļauti valodas modelī.

1 <https://api.tezaurs.lv>

Tekstvienība	Tips	Izruna / lasījums / tekstvienība bez drukas kļūdām	Izvēsta forma
ķīmijterapija	parasts vārds		
Spirix	svešvaloda	spiriks	
EKG	saīsinājums	ē kā gā	elektrokardiogramma
RAKUS	saīsinājums	rakus	rīgas austrumu klīniskās universitātes slimnīca
Neu	nepieciešams konteksts		
mkmol	saīsinājums	mikromoli	
ārstēšanas	drukā kļūda	ārstēšanas	
Kalniņa	personas dati		
g	nepieciešams konteksts		
Valmiera	personas dati		
Mydocalm	svešvaloda	midokalm	

2. tabula. Izrunas vārdnīcas fragments.

Teksta izvēršanas un savēršanas gramatika

Tekstu korpusa (anonimizētā RAKUS arhīva) izvēršanai un transkripciju savēršanai tiek izmantota līdzīga pieeja, kāda tika izmantota radioloģijas izmeklējumu diktātu transkribēšanas sistēmas izveidē igauņu valodai (Paats et al. 2018) – kontekstuāli formālas un skaitļojamas gramatikas likumi. Gramatikas likumi tika definēti, izmantojot paralēlos tekstus, kas tika iegūti runas korpusa izveides rezultātā: oriģinālie apraksti un tiem atbilstošās segmentu līmenī sastatītās ortogrāfiskās transkripcijas. Valodas modelis automātiskajai runas transkribēšanai tiek apmācīts uz izvērstā (verbalizētā) tekstu korpusa datiem, savukārt automātiskās transkripcijas tiek automātiski savērstas (deverbalizētas), izmantojot pretējus gramatikas likumus un tādējādi būtiski samazinot transkripciju pēcreģistrēšanai nepieciešamo darba apjomu. Gramatikas likumu principi parādīti 3. attēlā.

1 CT krūškurvja orgāniem ,
 2 CT krūšu kurvja orgāniem komats
 3 CT krūšu kurvja orgāniem ,

1 pēc Ultravist 300 - 100 ml i/v ievades
 2 pēc Ultravist trīssimt simt mililitru intravenozas ievades
 3 pēc Ultravist 300 100 ml i/v ievades

1 nierēs konkrements 1,89 cm / diam.
 2 nierēs konkrements viens komats astoņdesmit deviņi centimetri diametrā
 3 nierēs konkrements 1,89 cm / Ø

↑ 3. attēls. Fragmenti no sastatītā teksta un runas korpusa, kas ilustrē izvēršanas un savēršanas gramatikas lomu. Rindu apzīmējumi: 1 – fragments no teksta korpusa (references teksts); 2 – atbilstošais ortogrāfiskās transkripcijas fragments no runas korpusa (izvēršanas gramatikas mērķis); 3 – sagaidāmais automātiskās transkribēšanas sistēmas galarezultāts (savēršanas gramatikas mērķis).

↓ 4. attēls. Izmeklējumu diktātu automātiskās transkribēšanas un manuālās pēcreģistrēšanas sistēmas prototips: kreisajā pusē – automātiski atpazītā runa; labajā pusē – automātiski pēcapstrādāta transkripcija.

RUTA^{med}

* Saglabāt 5s Pabeigt B I U x² x₂ ≡ doc json

MR prostatai ar k/v.

Prostata 3,9 x 3,3 x 3,8 cm. Tilpums 26,15 cm³.

PSA līmenis nav zināms.

Stāvoklis pēc TUR-P operācijas ar tipisku defektu.

Dziedzera pārejas daļā redzami labdabīgas hiperplāzijas mezgli.

Audus ar difūzijas ierobežojumu dotajā izmeklējumā nesaskatu.

Sēklas pūslīši simetriski, bez strukturālām izmaiņām.

Urīnpūslis gludām sienām, bez intralūmenāliem ieslēgumiem.

Patoloģiska lieluma l/m nesaskatu.

Slēdziens

Labdabīga prostatas hiperplāzija.

Stāvoklis pēc TUR-P operācijas.

Pašreiz nav datu par augstas malignitātes tumora audiem prostatā.

▶ 00:01:21 / 00:01:32

Teksta divvirzienu pārrakstīšanas gramatika sastāv no vairākiem simtiem likumu, kas apraksta bezkonteksta un kontekstuālu teksta šablonu atpazīšanu un apstrādi – izvēršanu vai savēršanu. Tā kā latviešu valoda ir morfoloģiski bagāta, izvēršanas gramatikai ir jānodrošina vārdformu kontekstuāla sintaktiska saskaņošana, savukārt savēršanas gramatikai jāatpazīst vārdi locījumos.

Izstrādātās runas automātiskās transkribēšanas un manuālās pēcreidīgēšanas sistēmas prototips ilustrēts 4. attēlā. Šim automātiski transkribētajam un savērstajam MR izmeklējuma diktātam bija nepieciešami vien daži vienkārši manuāli labojumi. Dalījums teikumos un rindkopās, kā arī interpunkcija ir sistēmas ģenerēta; divās vietās manuāli tika precizēts, bet interpunkciju nebija nepieciešams koriģēt. Akronīmi MR un PSA tika korekti automātiski transkribēti ar izrunas vārdnīcu. Akronīms TUR-P tika transkribēts neprecīzi; ņemot to vērā, ir pilnveidota tekstu korpusa normalizēšana un papildināta izrunas vārdnīca. Skaitļu, mērvienību (*cm*, *cm³*), simbolu (*x*) u. c. atslēgvārdu (*k/v*, *l/m*) automātiska atpazīšana un saīsināšana veikta korekti. Formatējuma instrukcija “un rakstām lūdzu slēdzienu” atpazīta un automātiski interpretēta, izveidojot slēdziena sadaļu.

Secinājumi

Rakstā iezīmēta komplicētu specializēto korpusu izstrāde un adaptācija, kas ir nozīmīga latviešu valodas korpuslingvistikas resursu paplašināšanai. Specializētais teksta un runas korpus ir ne vien ļāvis izstrādāt automatizētu sistēmu vizuālās diagnostikas izmeklējumu transkribēšanai, bet arī veicinās medicīnas valodas sistemātiskāku izpēti un lietojumu.

Līdzšinējie eksperimenti, pielāgojot runas sintēzes un runas atpazīšanas sistēmas medicīnas lietojumiem, ir pierādījuši nozarei specifisku datu nozīmi. Vispārīgu ASR sistēmu var pielāgot specifiskai jomai, konkrēti – radioloģijai, izmantojot specializētas datu kopas – specializētu teksta un runas korpusu, izrunas vārdnīcu. Papildus korpusiem, kas nepieciešami valodas modeļa un akustiskā modeļa izstrādei, būtiska ir arī teksta izvēršanas un savēršanas gramatika, kas ļauj samazināt automātiski ģenerēto izmeklējumu un slimību aprakstu manuālās pēcapstrādes apjomu. Šobrīd tiek pārskatīta runas korpusa datu ortogrāfiskā transkripcija un vienādots pieraksts, kas palīdzēs objektīvāk novērtēt (kvantitatīvi) runas atpazīšanas precizitāti, kā arī ļaus pilnveidot izrunas vārdnīcu.

Turpmākais darbs ir saistīts ar automatizētās diktēšanas platformas pilnveidi un tās efektivitātes novērtēšanu reālos apstākļos. Jau sākotnējā prototipa testēšanas fāzē esam novērojuši, ka automātiskās transkripcijas kopumā veidojas konsekventākas nekā diktofonu centra dažādu operatoru sagatavotās transkripcijas. Automātiskās transkripcijas, protams, nav bez kļūdām, taču atpazīšanas precizitāte un pēcreidīgēšanas darba apjoms ir ļoti atkarīgi arī no diktēšanas stila. Tāpēc radiologiem tiek izstrādāti ieteikumi diktātu ierunāšanai, lai iegūtu pēc iespējas kvalitatīvāku rezultātu.

Kopumā būtiskākais jautājums un kritērijs sistēmas izmantošanai praksē ir: vai automātisko transkripciju pēcreidīgēšana, izmantojot specializēto redaktoru, kas nodrošina arī automātisku teksta un audio sastatīšanu, ir ātrāka un ērtāka nekā pilnībā manuāla izmeklējuma apraksta sagatavošana. Uz šo jautājumu jau šobrīd ir pozitīva atbilde, jo daļa radio-
logu, kas piedalās sistēmas testēšanā, būtu gatavi turpmāk vismaz daļu no izmeklējumu transkripcijām rediģēt paši, negaidot rindā uz diktofonu centru un vienlaikus samazinot diktofonu centra noslodzi.

- Blackley, Suzanne V., Huynh, Jessica, Wang, Liqin, Korach, Zfania, Zhou, Li (2019). Speech recognition for clinical documentation from 1990 to 2018: a systematic review. *Journal of the American Medical Informatics Association*, No. 26(4), pp. 324–338.
- Goedertier, Wim, Goddijn, Simo, Martens, Jean-Pierre (2000). Orthographic Transcription of the Spoken Dutch Corpus. *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC)*, Athens, pp. 909–914.
- Oostdijk, Nelleke, Boves, Lou (2008). Preprocessing speech corpora: Transcription and phonological annotation. Lüdeling, Anke, Kytö, Merja (eds.). *Corpus Linguistics: An International Handbook*. Volume I, Berlin, New York: W. de Gruyter, pp. 642–663.
- Paats, Andrus, Alumäe, Tanel, Meister, Einar, Fridolin, Ivo (2018). Retrospective Analysis of Clinical Performance of an Estonian Speech Recognition System for Radiology: Effects of Different Acoustic and Language Models. *Journal of Digital Imaging*, No. 31(5), pp. 615–621.
- Paikens, Pēteris, Rituma, Lauma, Pretkalniņa, Lauma (2013). Morphological analysis with limited resources: Latvian example. *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA)*, pp. 267–277.
- Pinnis, Mārcis, Auziņa, Ilze, Goba, Kārlis (2014). Designing the Latvian Speech Recognition Corpus. *Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland: ELRA, pp. 1547–1553.
- Pinnis, Mārcis, Salimbajevs, Askars, Auziņa, Ilze (2016). Designing a Speech Corpus for the. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*, Portorož, Slovenia: ELRA, pp. 775–780.
- Salimbajevs, Askars and Strigins, Jevgenijs (2015). Latvian speech-to-text transcription service. *Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 772–723.
- Schiel, Florian and Draxler, Christoph (2004). The Production of Speech Corpora. Version 2.5, 01.06. Available: <http://www.phonetik.uni-muenchen.de/forschung/BITS/TP1/Cookbook/TP1.html> [accessed 12.04.2022.].
- Spektors, Andrejs, Auziņa, Ilze, Dargis, Roberts, Grūzītis, Normunds, Paikens, Pēteris, Pretkalniņa, Lauma, Rituma, Laura, Saulīte, Baiba (2016). Tezaurs.lv: the largest open lexical database for Latvian. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*, pp. 2568–2571.
- Znotiņš, Artūrs, Cīrule, Elita (2018). NLP-PIPE: Latvian NLP Tool Pipeline. *Human Language Technologies – The Baltic Perspective*, Vol. 307, pp. 183–189.
- Znotiņš, Artūrs, Polis, Kaspars, Dargis, Roberts (2015). Media monitoring system for Latvian radio and TV broadcasts. *Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 732–733.

Development of a Specialized Latvian Speech Corpus and Pronunciation Dictionary for the Linguistic Analysis and Systematic Transcription of Visual Diagnostic Examinations

Ilze Auziņa, Roberts Dargis, Baiba Saulīte,
Normunds Grūzītis, Mikus Grasmanis,
Andrejs Spektors, Kaspars Stepanovs

Keywords: speech corpus, pronunciation dictionary, medical terminology, digital language resources, automatic speech recognition, natural language processing, post-editing

The Laboratory of Artificial Intelligence (AiLab) at the Institute of Mathematics and Computer Science of the University of Latvia (IMCS UL) in cooperation with the Riga East University Hospital (REUH) has developed the RUTA:MED platform for automated transcription of medical audio recordings. This was done within an ERDF-funded industry-driven research project aimed at developing specific Latvian speech recognition systems for the medical domain.

This paper describes the creation of Latvian language resources for the medical domain focusing on digital imaging to develop a medical speech recognition system for Latvian. The language resources include a pronunciation dictionary, a text corpus for language modelling, and an orthographically transcribed speech corpus for the (i) adaptation of the acoustic model, (ii) evaluation of the speech recognition accuracy, (iii) development and testing of rewrite rules for automatic text conversion to the spoken form and back to the written form.

Experiments to date in adapting speech synthesis and speech recognition systems to medical applications have demonstrated the importance of industry-specific data. The general ASR system can be adapted to a specific field, namely radiology, using specialized data sets – a specialized text and speech corpus, pronunciation dictionary. In addition to the corpora required for the development of the language model and the acoustic model, the grammar of text expansion and compression is also important, which allows to reduce the amount of manual post-processing of automatically generated examinations and disease descriptions.

Valodas korpusu izmantošana latviešu valodas uzdevumu automātiskā ģenerēšanā

Ilze Auziņa, Roberts Darģis,
Inga Kaija, Kristīne Levāne-Petrova,
Kristīne Pokratniece

Šis darbs ir FLPP projekta “Latviešu valodas apguvēju korpusa izveide: metodes, rīki un izmantojums” (projekta Nr. Izp-2018/1-0527) rezultāts sinerģijā ar Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi” projektu Nr. VPP-IZM-DH-2020/1-0001.

Ievads

Mūsdienās valodas korpusi nav tikai pētījumu empīrisks pamats, bet tie var būt noderīgi dažādu datus balstītu mācību materiālu un rīku izveidē. Korpusu izmantošanai valodas izpētē un apguvē, mācību un metodisko materiālu izstrādē ir lielas priekšrocības, jo korpusi piedāvā autentisku valodas materiālu un objektīvu skatījumu uz valodas parādībām, vairākkārt izmantojamus datus un iespēju pārbaudīt vai apstrīdēt hipotēzes par kādu valodas parādību vai valodu kopumā.

Lai pētītu valodu apguves īpatnības, analizētu valodas apguvēju pieļautās kļūdas un nodrošinātu datus balstītu valodas mācību un metodisko materiālu izstrādi, tiek veidoti valodas apguvēju korpusi, kuros tiek iekļauti valodas apguvēju (gan svešvalodas, gan otrās valodas) teksti un/ vai runas dati (Leech 1998: xiv; Nesselhauf 2005: 40; Granger 2003). Valodas apguvēju korpusi ir specializēti korpusi, t. i., korpusi, kuros dati apkopoti pēc noteiktas pazīmes. Arī Latvijā pēdējo gadu laikā ir tapuši vairāki latviešu valodas apguvēju korpusi, t. sk. otrās baltu (latviešu vai lietuviešu) valodas apguvēju korpusi ESAM¹, “Valsts valodas prasmes pārbaudes darbu korpus” (VVPP)² un jaunizveidotais “Latviešu valodas apguvēju korpus” (turpmāk tekstā – LaVA)³ (Auziņa u. c. 2021; Dargis et al. 2020), kura datu analīze ir pamatā rakstā aprakstītās pašmācības uzdevumu kopas izstrādē. Tā ir pirmā latviešu valodas apguvei paredzētā pašpārbaudes uzdevumu kopa, kurā iekļautie uzdevumi tiek ģenerēti automātiski no vairākiem valodas korpusiem un atbilst latviešu valodas prasmes pamatlīmenim.

Ar dabiskās valodas apstrādes (NLP) rīkiem (konformanču programmu, morfoloģisko analizatoru, tekstu sastatīšanu u. c.) var daļēji vai pilnībā automatizēt vairākus procesus, kas saistīti ar svešvalodu un otrās valodas (L2) apguvi. Viens no procesiem, ko arvien biežāk veiksmīgi automatizē, ir pašmācības uzdevumu ģenerēšana (O’Keeffe 2007; Pilán et al. 2016; Pilán et al. 2017; Smith et al. 2010; Volodina et al. 2012 u. c.). Uzdevumu izveidē tiek izmantoti ne tikai valodas apguvēju korpusi, bet arī paralēlie korpusi, kuros savstarpēji sastatīti teikumi un to tulkojumi citā valodā (Zanetti 2020). Par pamatu ņemot valodas korpusos atrodamos teikumus, leksiku, analizējot valodas apguvēju pieļautās kļūdas, tiek automātiski ģenerēti uzdevumi gan pašvadītai valodas apguvei, gan izmantošanai mācību procesā.

- 1 Otrās baltu valodas apguvēju korpus ESAM <http://esam.korpus.lv>
- 2 Valsts valodas prasmes pārbaudes darbu korpus <http://www.korpus.lv/id/VVPP>
- 3 Latviešu valodas apguvēju korpus (LaVA) <http://lava.korpus.lv>

Ja ir izstrādāta atbilstoša datu atlasē metodika un kritēriji, izmantojamo korpusu skaits var būt neierobežots. Tādējādi tiek nodrošināta uzdevumu daudzveidība un atbilstība konkrētam mērķim, turklāt ir iespējams izvairīties no subjektivitātes.

Lai varētu automatiski ģenerēt uzdevumus, ir jāveic pamatīga datu analīze un jānosaka uzdevumos iekļaujamo vārdu, vārdformu un teikumu atlasē kritēriji un atbilstība noteikta valodas prasmes līmeņa vingrinājumu elementiem. Izvēloties teikumus no korpusiem, ir vairāki papildu aspekti, kas jāņem vērā – 1) vai teikums ir saprotams izolēti, ārpus plašāka konteksta, 2) vai teikuma struktūra un lingvistiskā sarežģītība ir piemērota atbilstošam valodas prasmes līmenim. Arī runājot par tādu uzdevumu automatisku izveidi, kuros tiek piedāvāta lokāmo vārdšķiru vārdu paradīgmu apguve, jāņem vērā lingvistiskā atbilstība atbilstošam valodas prasmes līmenim.

Uzdevumu izveidē ir vairāki posmi: 1) kvantitatīva un kvalitatīva LaVA kļūdu analīze un tipisko kļūdu noteikšana, 2) lingvistisko paraugteikumu ieguve no dažādiem latviešu valodas korpusiem, t. sk. “Līdzsvarotā mūsdienu latviešu valodas tekstu korpusa” (LVK2018) un “Skolēnu pārsapņemošu korpusa” (SPK), ietverot tikai LaVA korpusā lietotos vārdus, 3) dažādu tipu uzdevumu ģenerēšana, izmantojot atlasītos paraugteikumus un vārdu sarakstus.

Datu analīzē un atlasē izmantotas korpuslingvistikas metodes. Kvantitatīvai LaVA datu analīzei izgūti dažādi statistiskie rādītāji, t. sk. absolūtais un relatīvais vārdu un vārdformu biežums korpusā, vārdu rangs. Kvalitatīvai datu analīzei un arī atbilstošu teikumu atlasē izmantotas konkordances.

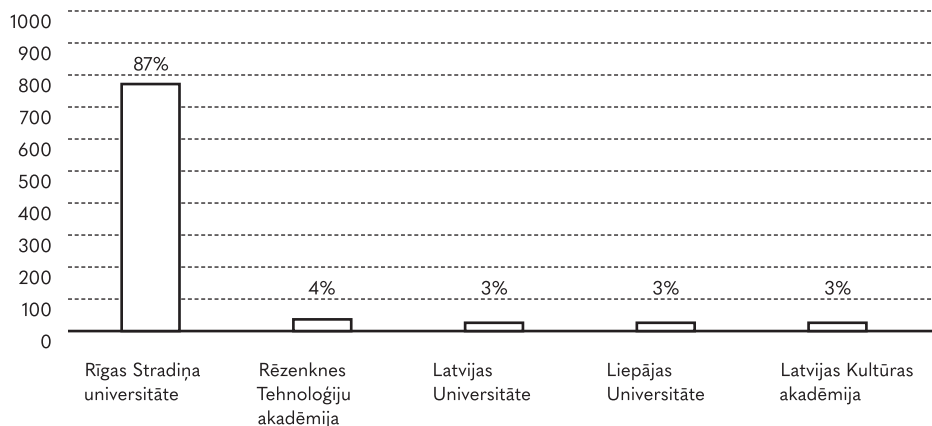
Raksta 1. nodaļā tiek raksturots “Latviešu valodas apguvēju korpus” un analizēti tā dati, raksta 2. nodaļā aprakstīta uzdevumos iekļaujamo un izmantojamo valodas vienību atlasē metodoloģija, savukārt raksta 3. nodaļā – uzdevumu kopas izstrāde un tajā iekļauto uzdevumu veidi.

LaVA korpus

La VA korpusa vispārīgs raksturojums

LaVA korpusā, kas fundamentālo un lietišķo pētījuma projektā (FLPP) “Latviešu valodas apguvēju korpusa izveide: metodes, rīki un izmantojums” (projekta Nr. lzp-2018/1-0527)⁴ laikā (2018–2021) izstrādāts LU MII, ir iekļauti 1015 to Latvijas augstākajās mācību iestādēs studējošo ārvalstnieku darbi, kuri latviešu valodu apgūst kā svešvalodu pirmo vai otro semestri, sasniedzot A1 (iespējams, A2) latviešu valodas prasmes līmeni. Korpusa LaVA apjoms ir gandrīz 190 000 vārdlietojumu.

4 Finansējuma avots: FLPP projekts “Latviešu valodas apguvēju korpusa izveide: metodes, rīki un izmantojums” (Nr. lzp-2018/1-0527).



1. attēls. Datu ieguves avoti.

LaVA dati ir iegūti no piecām Latvijas augstskolām: Rīgas Stradiņa universitātes (RSU), Latvijas Universitātes (LU), Latvijas Kultūras akadēmijas (LKA), Liepājas Universitātes (LiepU) un Rēzeknes tehnoloģiju akadēmijas (RTA). Visvairāk tekstu ir no RSU (87 %, 882 teksti), no pārējām augstskolām daudz mazāk: RTA – 45 teksti (4 %), LU – 33 teksti (3 %), LiepU – 30 teksti (3 %) un LKA – 25 teksti (3 %), sk. 1. att.

Korpusu LaVA veido datu kopa (sk. <http://lava.korpuss.lv/corpus/>), kurā ir 1) studentu rakstītās esejas kopija, 2) digitalizētā esēja, 3) labotais teksts. Turklāt katram tekstam ir pievienota informācija jeb metadati par personas dzimumu, vecumu, dzimto valodu, laiku, cik ilgi tiek apgūta latviešu valoda (sk. 2. att.).

Tā kā lielākā daļa tekstu ir rokrakstā, tie vispirms tika digitalizēti, manuāli pārrakstot datorrakstā. Detalizētāk par korpusa LaVA izveides procesu sk. Levāne-Petrova et al. 2020. Savukārt digitalizētie teksti pēc korpusa veidotāju izstrādātās metodikas ir pārrakstīti atbilstoši valodas normām, t. i., ir izvirzīta mērķhipotēze (Auziņa u. c. 2020). Tādējādi teksti ir laboti, interpretējot valodas apguvēju rakstīto un izlemjot, ko valodas apguvējs katrā konkrētā gadījumā ir gribējis teikt, piem., *Es vinmer ēdu rīsu.* → *Es vienmēr ēdu rīsus.*

Gan oriģinālie, gan labotie teksti ir automātiski morfoloģiski marķēti – katrai vārdformai pievienojot morfoloģiskās pazīmes un pamatformu, piem., vārdformai *Rīgā* tiek pievienotas morfoloģiskās pazīmes [npfsl4], kur *n* ir lietvārds, *p* – īpašvārds, *f* – sieviešu dzimte, *s* – vienskaitlis, *l* – lokatīvs, *4* – 4. deklinācija, un norādīta pamatforma – *Rīga*. Automātiskā

Dokuments ID 28

Vecums: 26

Dzimums: Vīrietis

Originalais teksts

No rīta, es bieži esot pusstoņos. Es šū brokastis astoņos. Parasti, es gatavoju putru ar ābolu un banānu. Man ļoti garšo arī augļi piemēram zemenes un mango. Bet man arī ļoti garšo maize un olu kultenis brokastīs. Es bieži dzeru tēju no rīta. Pēc brokastis es parasti braucu uz universitāti ar pirmo tramvaju, vai velosipēdu vasarā. Universitātē es mācos. No pirmdienas līdz piektdienai es šū pusdienas trīspadsmitos Stockpotā. Es vienmēr šū rīsu. Pēc pusdienas es dzeru kafiju kopā ar draugiem kafējnīcā. Sestdienā un Svētdienā es gatavoju makaronus vai kartupeļus pusdienām. Es sportoju četras reizes nedēļas vakarā. Man arī ļoti patīk spēlēt futbolu un badmintonu vai laist. Vakārā es gatavoju maizi vai salāti. Es parasti šū vakariņas astoņos. Man ļoti patīk gatavot un dzeru vīnu kopā ar draugiem. Es šū gulēt pusnaktis vakarā. Nedēļas nogale es bieži mācos nacionālā bibliotēkā vai mājās. Dažreiz es un mani draugi spēlējām futbolu parkā un vakarā mēs ejam uz krogu vai naktsklubu, piemēram Folk klubs ala pagrabā. Bet Sestdienās vai svētdienās es arī tīru dzīvokli.

Labotais teksts

No rīta es bieži esot pusstoņos. Es šū brokastis astoņos. Parasti, es gatavoju putru ar ābolu un banānu. Man ļoti garšo arī augļi, piemēram, zemenes un mango. Bet man arī ļoti garšo maize un olu kultenis brokastīs. Es bieži dzeru tēju no rīta. Pēc brokastīm es parasti braucu uz universitāti ar pirmo tramvaju vai velosipēdu vasarā. Universitātē es mācos. No pirmdienas līdz piektdienai es šū pusdienas trīspadsmitos Stockpotā. Es vienmēr šū rīsus. Pēc pusdienām es dzeru kafiju kopā ar draugiem kafējnīcā. Sestdienās un svētdienās es gatavoju makaronus vai kartupeļus pusdienām. Es sportoju četras reizes nedēļā vakarā. Man arī ļoti patīk spēlēt futbolu un badmintonu vai laist. Vakārā es gatavoju maizi vai salātus. Es parasti šū vakariņas astoņos. Man ļoti patīk gatavot un dzeru vīnu kopā ar draugiem. Es šū gulēt pusnakti. Nedēļas nogalē es bieži mācos Nacionālajā bibliotēkā vai mājās. Dažreiz es un mani draugi spēlējām futbolu parkā, un vakarā mēs ejam uz krogu vai naktsklubu, piemēram, folkkluba Ala pagrabu. Bet sestdienās vai svētdienās es arī tīru dzīvokli.

Semestris: 2

Dzīvotāji (mēneši): 12

Dzimtā valoda: German

Citas valodas: English

Ļo rīta, es bieži esot pusstoņos. Es šū brokastis astoņos. Parasti, es gatavoju putru ar ābolu un banānu. Man ļoti garšo arī augļi piemēram zemenes un mango. Bet man arī ļoti garšo maize un olu kultenis brokastīs. Es bieži dzeru tēju no rīta. Pēc brokastis es parasti braucu uz universitāti ar pirmo tramvaju, vai velosipēdu vasarā. Universitātē es mācos. No pirmdienas līdz piektdienai es šū pusdienas trīspadsmitos Stockpotā. Es vienmēr šū rīsu. Pēc pusdienas es dzeru kafiju kopā ar draugiem kafējnīcā. Sestdienā un Svētdienā es gatavoju makaronus vai kartupeļus pusdienām. Es sportoju četras reizes nedēļas vakarā. Man arī ļoti patīk spēlēt futbolu un badmintonu vai laist. Vakārā es gatavoju maizi vai salāti. Es parasti šū vakariņas astoņos. Man ļoti patīk gatavot un dzeru vīnu kopā ar draugiem. Es šū gulēt pusnaktis vakarā. Nedēļas nogale es bieži mācos nacionālā bibliotēkā vai mājās. Dažreiz es un mani draugi spēlējām futbolu parkā un vakarā mēs ejam uz krogu vai naktsklubu, piemēram Folk klubs ala pagrabā. Bet Sestdienās vai svētdienās es arī tīru dzīvokli.

Edit report

Original	Es	domāju	īstairīti	no	neveselīgi	pārtika
Without types		domāju	īstairīti		neveselīgi	pārtika
Original lemma		domāt	īstairīties		neveselīgs	pārtika
Original tag		vnmip_21san	vmyopt3000n		af_n_p	ncf_4
Corrected	Es	domāju	īstairīti	no	neveselīgi	pārtika
Corrected lemma	es	domāt	īstairīties	no	neveselīgs	pārtika
Corrected tag	pp10ann	vnmip_21san	vmyo0_3000n	spag	affsgnp	ncfag0
Unclear						
Misalignment						

↑ 2. attēls. Korpusa LaVA saskarnes ekrānuzņēmums.

↓ 3. attēls. Korpusa LaVA anotēšanas saskarne.

morfoloģiskā marķēšana ir veikta, izmantojot latviešu valodai izstrādāto automātiskās morfoloģiskās anotēšanas rīku (Paikens et al. 2013) un latviešu valodas morfoloģisko pazīmju kopu⁵. Tā kā latviešu valoda ir morfoloģiski daudznozīmīga un automātiskās marķēšanas procesā reizēm nav iespējams atšķirt konkrētas vārdformas vārdšķiru, dažkārt novērojamas kļūdas morfoloģiskajā marķējumā, un morfoloģiskais marķējums tika manuāli pārskatīts. Lai marķētu valodas apguvēju oriģinālajos tekstos visas lietotās tekstvienības, piem., paradigmai neatbilstošas formas (*universitātī, Francijī*) vai neeksistējošus vārdus un vārdformas (*bizeni, dārzē*), ir izmantota reducēta morfoloģisko pazīmju kopa, parasti norādot tikai vārdšķiras identifikatoru: *Francijī* – tag [n], lemma [Francija]; *universitātī* – tag [n], lemma [universitāte]; *aizguļošamies* – tag [v], lemma [aizgulēties].

5 Korpus.lv latviešu valodas morfoloģisko pazīmju kopa (v.2.2.1 (22.09.2021.)) http://www.korpus.lv/LV_TagSet_v.2.2.1_22092021.pdf

Korpusā LaVA ir arī marķētas valodas apguvēju pieļautās kļūdas atbilstoši noteiktiem kļūdu tiptiem: pareizrakstības kļūdas (*Spelling*), formveidošanas un vārddarināšanas kļūdas (*WordFormation*), leksikas kļūdas (*Lexical*), interpunkcijas kļūdas (*Punctuation*), sintakses kļūdas (*Syntactic*), kombinētas kļūdas (*Complex*). Sasatot korpusa oriģinālos un labotos tekstus, daļa kļūdu – pareizrakstības, formveidošanas un vārddarināšanas, interpunkcijas un leksikas kļūdas – ir marķētas pusautomātiski, pēc tam marķējumu manuāli pārskatot (sk. 3. att.). Savukārt sintakses un kombinētās kļūdas ir marķētas manuāli.

LaVA korpusā lietoto vārdu un vārdformu lietojuma biežums

Korpusa LaVA iekļautajos 1015 tekstos – gan oriģinālajos, gan labotajos – ir sastopami vairāk nekā 192 tūkstoši tekstvienību (t. sk. vārdformu, pieturzīmju, simbolu). Valodas apguvēji lietojuši 8400 dažādu vārdu un vairāk nekā 20 tūkstošus dažādu vārdformu.

Biežāk lietotie 20 vārdi korpusā ir *es* (16198), *būt* (11247), *un* (8220), *mana* (4119), *viņš* (3996), *patikt* (3886), *mans* (3580), *viņa* (2874), *dzīvot* (2476), *gads* (2365), *Rīga* (2280), *ar* (2071), *universitāte* (2044), *studēt* (1895), *saukt* (1866), *mēs* (1844), *uz* (1585), *bet* (1581), *no* (1538), *loti* (1532). Biežāk lietotie vārdi ir cieši saistīti ar studentu eseju tematiem – “Es un mana ģimene” (piem., *ģimene, brālis, draugs, māsa, tēvs, māte; saukt, studēt; liels, vecs, mīļš, mazs*), “Mana ikdiena” (piem., *studēt, iet, braukt; brīvs*), “Manas studijas” u. c.

	Lietvārds	Darbības vārds	Īpašības vārds	Vietniekvārds	Skaitļa vārds
1.	gads	būt	labs	es	divi
2.	Rīga	patikt	liels	mana	viens
3.	universitāte	dzīvot	vecs	viņš	divdesmit
4.	ģimene	studēt	mīļš	mans	trīs
5.	brālis	saukt	brīvs	viņa	astoņi
6.	draugs	ēst	jauns	mēs	četri
7.	māsa	runāt	mazs	tas	pieci
8.	māja	iet	skaists	katra	septiņi
9.	tēvs	garšot	grūts	tā	deviņi
10.	māte	braukt	nākamais	viss	seši

1. tabula. Biežāk lietoto lokāmo vārdšķiru vārdi.

Likumsakarīgi, ka vērojama korelācija starp vārdu lietojuma biežumu un kļūdaini lietotajām vārda formām, piem., visbiežāk valodas apguvēji kļūdījušies, lietojot lietvārdu *māja, brālis, darbs, draudzene, universitāte, māsa, nodarbība, bioķīmija, grāmata, bibliotēka* formas, no kuriem vārdi *gads, universitāte, brālis, māsa, māja* ir pirmajā desmitniekā.

LaVA korpusa kļūdu analīze

Izmantojot kļūdu marķējumu, *SketchEngine* programmā (Rychlý et al. 2007) ir iespējams iegūt informāciju par valodas apguvēju pieļautajām kļūdām. Kļūdu analīze liecina, ka apguvēju tekstos izplatītākās ir formveidošanas un vārddarināšanas kļūdas (40 %) un pareizrakstības kļūdas (33 %) (sk. 4 att.).

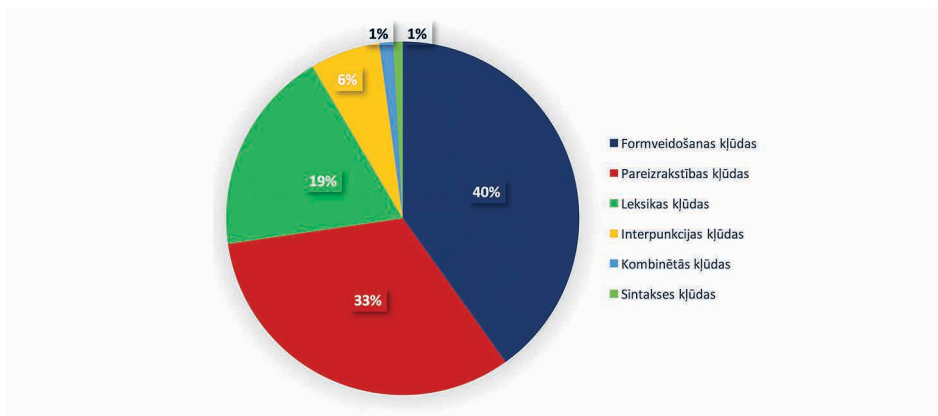
Kļūdaini lietoto vārdformu biežums korelē ar vārdu lietojumu biežumu korpusā, t. i., formveidošanas un vārddarināšanas kļūdas visvairāk sastopamas LaVA korpusā biežāk lietotajos vārdos: *mans, mana, viņš, gads, es, viņa, brālis, draugs, valoda, māte, māsa, divi, tēvs, Rīga, ģimene*. Formveidošanas un vārddarināšanas kļūdu grupā 61,7 % gadījumu ir kļūdaini lietoti lietvārdi, 17,7 % gadījumu – vietniekvārdi, 9,7 % gadījumu – darbības vārdi, 4,2 % gadījumu – īpašības vārdi, 3,9 % gadījumu – skaitļa vārdi un 2,8 % gadījumu – citu vārdsķiru vārdi. Visbiežāk valodas apguvēji nav atbilstoši lietojuši vienskaitļa akuzatīva formas 4. deklinācijas (*medicīnu, mūsu, valodu, dienu, zobārstniecību, kafiju, anatomiju, bibliotēku* utt.) un 1. deklinācijas lietvārdiem (*tēvu, draugu, futbolu, autobusu, gadu, sieru* utt.). Nav pareizi lietotas arī 4. deklinācijas lietvārda vienskaitļa lokatīva, 1. konjugācijas vienskaitļa un daudzskaitļa nominatīva formas.

Visbiežāk valodas apguvēji kļūdās vārdu *patikt* (*patik* → *patik, patick* → *patik, patika* → *patika* u. c.), *dzīvot* (*dzīvo* → *dzīvo, dzivot* → *dzīvot* u. c.), *viņš* (*vīna* → *viņa, viņam* → *viņam, viņi* → *viņi* u. c.), *studēt* (*stude* → *studē, studejam* → *studējam* u. c.), *Rīga* (*riga* → *Rīga, Rīga* → *Rīga* u. c.), *ļoti* (*loti* → *ļoti, lotī* → *ļoti* u. c.), *arī* (*ari* → *arī*), *universitāte* (*universitate* → *universitāte, universitatē* → *universitātē* u. c.), *viņa* (*vīna* → *viņa, viņa* → *Viņa* u. c.), *ēst* (*ed* → *ēd, edišu* → *ēdišu, edu* → *ēdu* u. c.), *mēst* (*mes* → *mēs, musu* → *mūsu* u. c.), *runāt* (*runa* → *runā, runajam* → *runājam* u. c.), *nepatikt* (*nepatik* → *nepatik* u. c.), *ģimene* (*giemene* → *ģimene, gimenē* → *ģimenē* u. c.) formu pareizrakstībā – visbiežāk tās ir ar diakritisko zīmju lietojumu (neesamību) saistītas kļūdas.

Lielākā daļa pareizrakstības, formveidošanas un vārddarināšanas, kā arī leksikas kļūdu ir saistītas ar vienu tekstvienību – parasti vienu vārdu vai vārdformu, bet ir analītiskas gramatiskās formas, kurās gramatiskās nozīmes izteikšanai tiek izmantots patstāvīgs vārds kopā ar kādu palīgvārdu vai palīgvārda funkcijā lietotu vārdu, piem., darbības vārdu salikto laiku formas (*esmu lasījis, ir bijis* u. tml.), prepozicionāli savienojumi (*ar draugiem, uz skolu* u. tml.). Kļūdas analītisko gramatisko formu lietojumā marķētas kā sintakses kļūdas.

Latviešu valodā atšķirībā, piem., no angļu valodas, pieturzīmju lietojums teikumā ir ļoti nozīmīgs. Latviešu valodā interpunkcija ir balstīta uz gramatikas principiem, un atšķirīga pieturzīmju izmantošana bieži pilnībā maina teikuma nozīmi. Tādēļ korpusā, lai gan iesācēju līmenī interpunkcijas zināšanas valodas apguvējiem netiek prasītas, labotajos tekstos pieturzīmes liktas atbilstoši latviešu valodas interpunkcijas likumiem. Tas savukārt ļauj automātiski noteikt interpunkcijas kļūdas.

Bieži vienā vārdformā ir sastopamas vairāku tipu kļūdas – pareizrakstības un formveidošanas kļūda.



4. attēls. Kļūdu izplatība korpusā LaVA.

Paraugteikumu atlase un datu sagatavošana uzdevumu ģenerēšanai

Balstoties uz korpusa LaVA datu, t. sk. valodas apguvēju kļūdu, analīzi, pēc noteiktas metodoloģijas tiek ģenerēti uzdevumi. Uzdevumi paredzēti, lai palīdzētu valodas apguvējam nostiprināt latviešu valodas lingvistisko kompetenci, piem., darbības vārdu personu formu lietojumu īstenības izteiksmes vienkāršajos un saliktajos laikos, vārdu saskaņošanu, prievārdisko konstrukciju lietojumu.

Lai mācību uzdevumus varētu ģenerēt automātiski, tiek apzināts, kādas lingvistiskās prasmes valodas apguvējiem pamatlīmenī (A1, A2) jāsasniedz (sk. 2.1). Tas savukārt, izmantojot morfoloģiskās pazīmes, ļauj atlasīt teikumus ar atbilstošām vārdformām (sk. 2.3.) un LaVA korpusā sastopamajiem vārdiem (sk. 2.2.).

Lingvistisko prasmju definēšana

Šobrīd LaVA korpusā apkopotie dati atspoguļo to valodas apguvēju latviešu valodas prasmi, kuri tikai sāk apgūt valodu un iegūst pamatzināšanas, mācību laikā sasniedzot pamatlīmeņa 1. pakāpi (*Breakthrough*, (A1)) vai 2. pakāpi (*Waystage*, (A2)). Līdz ar to arī korpusā balstītie pašpārbaudes uzdevumi ir paredzēti tiem, kas latviešu valodu apgūst A1–A2 līmenī.

A1 ir zemākais valodas prasmes līmenis, ar kuru sākas jaunas, šajā gadījumā – latviešu, valodas apguve. A1 līmenī valodas apguvējs ir apguvis vienkāršas valodas formas (piem., darbības

vārda vienkāršās tagadnes, pagātnes un nākotnes formas, lietvārdu locīšanas paradigmu), spēj jautāt un atbildēt uz vienkāršiem jautājumiem par sevi, dzīvesvietu, pazīstamiem cilvēkiem u. tml., spēj veidot vienkāršus izteikumus par labi zināmiem tematiem, lieto ļoti vienkāršas valodas līdzekļus, nodarbības apgūtu un ar konkrētiem tematiem un situācijām saistītu leksiku un frāzes. Valodas sniegums vēl ir ļoti nepilnīgs (Šalme, Auziņa 2016: 17). Savukārt A2 līmenī mācās izmantot valodu sociālo funkciju un komunikatīvās darbības īstenošanai, turpina pilnveidot valodas prasmes, mācoties saziņā lietot vienkāršas pieklājības formas, apjautāties par saziņas partnera sajūtām un emocionālo stāvokli, uzzināt jaunumus un iegūt citu informāciju, uzdot jautājumus un atbildēt uz jautājumiem par dažādām tēmām (Šalme, Auziņa 2016: 17).

Korpusā LaVA iekļauto tekstu atšķirības daļēji nosaka tas, ka tie tapuši dažādās augstskolās vai vienas augstskolas dažādu studiju laikā, piem., no Rīgas Stradiņa universitātes tekstus ievākuši kursu “Latviešu valoda medicīnā”, “Latviešu valoda zobārstniecībā” un “Latviešu valoda apmaiņas studentiem” docētāji. Katra kursa saturs var būt atšķirīgs atkarībā no kursa apjoma un apguvēju profesionālajām vajadzībām (Laizāne 2018: 36). Arī tekstu rakstīšanas laiks, t. i., 1. vai 2. semestris, semestra sākums vai beigas, ietekmē tekstos iekļauto saturu, it sevišķi gramatikas aspektā.

Tekstos atrodamās gramatiskās formas un konstrukcijas lielākoties atspoguļo līdz rakstīšanas brīdim kursā apgūto. Tā kā dažādosursos un dažādos viena kursa posmos tas var būtiski atšķirties, uzdevumu veidošanai tika izveidots iekļaujamo gramatisko vienību saraksts, proti, tas veidots pēc A1–A2 līmeņa apraksta (Šalme, Auziņa 2016). Saskaņā ar šo sarakstu uzdevumos var būt iekļauti:

- 1.–6. deklinācijas lietvārdi;
- darbības vārdi darāmās kārtas īstenības izteiksmē;
- īpašības vārdi ar nenoteiktajām galotnēm;
- vietniekvārdi:
 - personu vietniekvārdi nominatīvā, ģenitīvā (izņemot *manis* un *tevis*), datīvā un akuzatīvā;
 - norādāmie, piederības, jautājamie, nenoteiktie un noteiktie vietniekvārdi nominatīvā, ģenitīvā, datīvā, akuzatīvā un lokatīvā;
 - attieksmes, atgriezeniskais vietniekvārds un nolieguma vietniekvārdi netiek iekļauti;
- apstākļa vārdi;
- prievārdi prepozitīvā lietojumā ar patstāvīgajiem vārdiem vienskaitlī:
 - *aiz, virs, zem, pie, no, ārpus, pirms, pēc, kopš, bez* saistījumā ar vienskaitļa ģenitīva formām;
 - *līdz* ar vienskaitļa datīva formām;
 - *ap, gar, pa, caur, pret, starp, pār, ar, par* ar vienskaitļa akuzatīva formām;

- *uz* ar vienskaitļa akuzatīva vai vienskaitļa ģenitīva formām;
- vienkārši saikļi *un, bet, vai, ka, jo, tāpēc ka, ja, kā, lai*;
- pamata vai kārtas skaitļa vārdi (vienkārši vai salikteņi) visos locījumos, izņemot vokatīvu;
- izsaukmes vārdi;
- vienkāršas partikulas *vai* (tikai kā jautājuma partikula), *jā, nē, varbūt, arī, diemžēl, kā* (tikai salīdzinājumā), *nekā*
- saīsinājumi netiek iekļauti.

Korpusā LaVA iekļauto tekstu tematiku nosaka docētāja uzdevums (temats, par kādu autoriem uzdots tekstu rakstīt), piem., “Es un mana ģimene”; “Manas mājas un mana ikdiena”; “Mani vaļasprieki” u. c. Šie temati atbilst A1–A2 līmenim, taču tas, kādus konkrētus valodas elementus izvēlēties, ir autora ziņā. Tā kā, tekstus rakstot, ir atļauts izmantot palīg līdzekļus, piem., vārdnīcas, tekstos iekļautā leksika neaprobežojas ar A1–A2 līmeni tipisko, bet gan iekļauj tādu leksiku, kāda katram autoram ir šķitusi nepieciešama temata izklāstā. Tas nozīmē, ka pēc šī korpusa nevar izdarīt visaptverošus secinājumus par to, kādu leksiku autori zina, bet var pētīt, kādu leksiku apguvēji uzskata par lietderīgu un tiecas izmantot, rakstot par noteiktiem tematiem. Tāpēc leksikas ziņā uzdevumi ir balstīti nevis A1–A2 līmeņa aprakstā, bet gan korpusa LaVA datos.

Vārdu saraksta izveide

Uzdevumu ģenerēšanā ir svarīgi izmantot tikai tos vārdus, kuri valodas apguvējam būtu jāzina. Visprecīzāko informāciju par vārdiem, kurus apguvēji izmanto, var iegūt no apguvēju korpusa.

No apguvēju korpusa tika apkopota informācija par lemmām jeb pamatformām, kuras apguvēji ir izmantojuši tekstos. Apguvēju tekstos var parādīties arī specifiski vārdi, kuri netiek mācīti iesācēju līmenī, bet ko kāds apguvējs ir atradis vārdnīcā, jo tas bijis nepieciešams viņa tekstā. Lai izvairītos no šādu vārdu iekļaušanas vārdu sarakstā, tajā tiek iekļauti tikai tie vārdi, kuri korpusā sastopami vismaz trīs darbos.

Apguvēju korpusa apjoms nav pietiekams, lai iegūtu reprezentatīvu statistiku par katra vārda izmantotajām vārdformām, tāpēc iegūtie vārdi tika automātiski locīti un iegūtās vārdformas tika filtrētas, balstoties uz līmeņa aprakstu. Iegūtais vārdformu saraksts tika tālāk izmantots uzdevumu ģenerēšanā un paraugteikumu atlasē.

Teikumu atlase

Daļai uzdevumu tipu ir nepieciešami teikumi. Tā kā korpusā LaVA atrodami teikumi ir ar dažādiem ierobežojumiem (sk. rakstā iepriekš), lai nodrošinātu pietiekami lielu teikumu

dažādību, teikumi tika atlasīti arī no korpusa LVK2018 un SPK. Teikumu atlasēi izmantotais tekstu korpus LVK2018 ir aptuveni 10 miljonu vārdlietojumu liels, vispārīgs korpus, kas sastāv no dažādu žanru tekstiem (periodikas (60 %), daiļliteratūras (20 %), zinātniskiem tekstiem (10 %), normatīvajiem aktiem (8 %) un Saeimas stenogrammām (2 %)). LVK2018 ir automātiski morfoloģiski marķēts, un tam pievienoti metadati (Levāne-Petrova 2012; 2019). Savukārt SPK ir veidots, lai pētītu skolēnu latviešu valodas prasmes un nodrošinātu dotos balstītus latviešu valodas mācību un metodisko materiālu izstrādi. SPK ir iekļauti 468 pārspriedumi no 12. klases latviešu valodas eksāmenu darbiem (apmēram 185 000 vārdlietojumu). Tie ir latviešu mācībvalodas vidusskolu, mazākumtautību skolu un valsts ģimnāziju audzēkņu darbi no Kurzemes, Latgales un Rīgas. SPK ir iekļauti nelaboti teksti, paturot visas skolēnu pieļautās drukas, interpunkcijas u. c. kļūdas (Levāne-Petrova, Pokratniece 2021).

Atlasītajiem teikumiem jāatbilst valodas apguvēju līmenim. Lai nodrošinātu, ka teikumos izmantotā leksika ir saprotama valodas apguvējiem, pirmajā kārtā tika atlasīti tikai tie teikumi, kuros bija vārdi un vārdformas tikai no iepriekš izveidotā saraksta.

Ar teikuma filtrēšanu tikai pēc vārdu saraksta bieži vien nepietiek. Starp atlasītajiem teikumiem ir gan pārāk gari un sarežģīti teikumi, gan arī pārāk īsi teikumi, kas nesatur pietiekami daudz informācijas uzdevumu ģenerēšanai, piem., teikumi ar nepilnu struktūru. Nākamajā teikumu atlasē kārtā ir nepieciešams veikt sintaktisko filtrēšanu, lai izvērtos no šādu teikumu iekļaušanas uzdevumu ģenerēšanā. Filtrēšana tika veikta, balstoties uz šādiem kritērijiem:

- viens teikums nevar sastāvēt vairāk nekā no trim neatkarīgām teikuma daļām;
- visās teikuma daļās jābūt teikuma gramatiskajam centram – vēlamam, gan teikuma priekšmetam, gan izteicējam;
- neiekļaut vienā teikuma daļā vairāk par diviem apstākļa vārdiem (tomēr tajā var būt vairāk nekā divi paplašinātāji);
- neiekļaut vienā teikumā vairāk par divām partikulām.

Kritēriji ir izveidoti atbilstoši docētāju pedagoģiskajai pieredzei, mācot latviešu valodu kā svešvalodu augstākās izglītības iestādēs.

Lietojot latviešu valodu kā dzimto valodu, teikumi mēdz būt aprauti, ar mainītu vārdu secību vai izlaistu informāciju, kas nosakāma no konteksta. Tāpēc no iepriekš minētajos valodas korpusos iegūtajiem teikumiem tika atlasīti 1000 teikumi (800 teikumi no LVK2018 un 200 teikumi no SPK), pēc nepieciešamības tos pielāgojot A1–A2 līmenim. Veiktas šādas izmaiņas:

- samazināta teikuma atkarība no konteksta (piem., *Un vai tieši Latvijas iedzīvotājiem vienmēr būs šī nauda?* → *Vai Latvijas iedzīvotājiem vienmēr būs nauda?*);
- pievienots teikuma priekšmets un/vai izteicējs (piem., *Gribu pastāstīt, kā šis jautājums virzās.* → *Es gribu pastāstīt, kā šis jautājums virzās.*);

- teikuma vārdu secība mainīta uz neitrālu (piem., *Nu, nebūs tā nekāda izvēles iespēja!* → *Tā nebūs nekāda izvēles iespēja.*);
- izmantota tikai daļa(-as) no sākotnējā teikuma (piem., *Tad, kad mēs teicām: "Bet mēs esam no Latvijas!"* → *Mēs esam no Latvijas.*);
- dažkārt teikums izmantots tikai ierosmei, veidojot tā vietā citu (piem., *Viņa, māte, to nebija varējusi.* → *Viņa māte to nevarēja.*)

Teikumus manuāli atlasīja un pēc vajadzības pārveidoja praktizējoša latviešu valodas kā svešvalodas docētāja pieaugušajiem augstākās izglītības iestādē Latvijā ar pieredzi mācību materiālu gatavošanā.

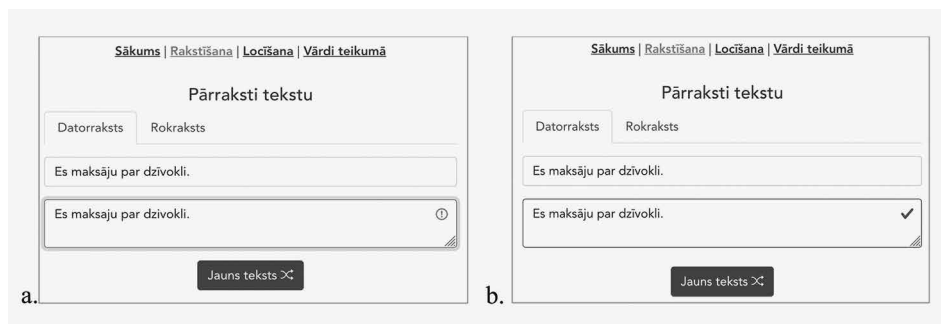
Uzdevumu tipi un uzdevumu ģenerēšana

Balstoties uz kļūdu analīzi, tika izvēlēti trīs dažādu uzdevumu veidi: teksta pārrakstīšanas uzdevumi (sk. 3.1.), vārdu locīšanas uzdevumi (sk. 3.2.) un uzdevumi, kuros valodas apguvējam teikumā jāieraksta vai jāizvēlas atbilstošā vārda forma (sk. 3.3.). Izraudzītie uzdevumu veidi atbilst biežāk sastopamajām apguvēju kļūdām un palīdz izvairīties no tām: rakstīšanas uzdevumi palīdz novērst pareizrakstības kļūdas, locīšanas uzdevumi – formveidošanas kļūdas, savukārt, ierakstot vai ievietojot atbilstošo vārdformu teikumā, valodas apguvēji nostiprina savas zināšanas formveidošanā un sintaksē. Leksikas un interpunkcijas apguvei uzdevumi nav veidoti.

Pārrakstīšanas uzdevumi (Rakstīšana)

Latviešu valodas rakstības pamatā ir latīņu alfabēts, kas papildināts ar diakritiskām zīmēm patskaņu garuma (˘), palatālo līdzskaņu (,) , šņāceņu (˙) apzīmēšanai. Tieši šo papildu diakritisko zīmju kļūdaina izmantošana vai neesamība ir biežs pareizrakstības kļūdu cēlonis. Gan patskaņa garumam neatbilstošā burtu izmantojums (īsa patskaņa vietā tiek rakstīts gara patskaņa burts, gara patskaņa vietā – īsa patskaņa burts), gan līdzskaņu rakstība bez diakritiskajām zīmēm vai neatbilstošas diakritiskās zīmes izmantojums, bieži vien ir saistīts ar valodas apguvēja dzimtās valodas fonētiski fonoloģiskās sistēmas un/vai grafētikas atšķirību no latviešu valodas (Auziņa u. c. 2019). Nepietiekamas zināšanas par latviešu valodas diakritiskajām zīmēm dažkārt noved arī pie latviešu valodas sistēmā neesošas diakritiskās zīmes lietojuma (Kaija 2020). Pareizrakstības kļūdas mēdz būt arī, piem., burtu rakstīšana nepareizā secībā, it sevišķi garākos vārdos.

Teksta pārrakstīšana palīdz labāk apgūt latviešu valodas grafētisko sistēmu, gan vizuāli pievēršot uzmanību burtu secībai un diakritiskajām zīmēm, gan šo secību atkārtojot pašrocīgi. Arī apguvējiem ar disleksiju pārrakstīšana tiek minēta kā viens no noderīgiem pareizrakstības apguves veidiem (Crombie 2000). Šādi uzdevumi arī ļauj trenēt diakritisko zīmju lietošanu datorrakstā, novēršot gadījumus, kad diakritiskās zīmes netiek lietotas tikai tāpēc, ka apguvējs tehniski nav pieradis to darīt.



5. attēls. Rakstīšanas uzdevuma saskarne: kļūdaina izpilde (a) un pareiza izpilde (b).

Pārrakstīšanas uzdevumos valodas apguvējam ir iespēja izvēlēties, vai pārrakstīt jau datorrakstā dotu teikumu vai rokrakstā. Datorrakstā tiek piedāvāts nejauši izvēlēts teikums no iepriekš atlasītajiem (sk. 2.3.). Rokrakstā tiek piedāvāts nejauši izvēlēts teikuma attēls, kas iegūts manuāli, izgriezts no apguvēju darbiem.

Ja valodas apguvējs, pārrakstot teikumu, kļūdās, par to uzreiz tiek ziņots – rāmis ap tekstu kļūst sarkans (sk. 5. (a) att.). Par kļūdu tiek uzskatīts nepareizs burts, burts bez diakritiskās zīmes vai ar lieku/neatbilstošu diakritisko zīmi, neatbilstošs lielo un mazo burtu lietojums, lieka vai nesoša atstarpe, lieka, trūkstošā vai nepareiza pieturzīme. Kad viss teikums ir pārrakstīts pareizi, tas tiek iekrāsots zaļā krāsā. Jebkurā brīdī ir iespējams pieprasīt jaunu teikumu (sk. 5. (b) att.).

Vārdu locīšanas uzdevumi (Locīšana)

Otra grupa ir uzdevumi, kuros valodas apguvējam tiek piedāvāts apgūt latviešu valodas lo-kāmo vārdšķiru – lietvārdu, darbības vārdu, īpašības vārdu, vietniekvārdu un skaitļa vārdu – locīšanas paradigmas. Uzdevumi aptver tikai daļu to gramatikas kategoriju, kuras atbilst latviešu valodas prasmes līmenī dotajam lingvistiskās kompetences raksturojumam, konkrēti – noteiktajām gramatikas zināšanām (Šalme, Auziņa 2016). Valodas apguvējs, izmantojot šos uzdevumus, var apgūt:

- 1.–6. deklinācijas lietvārdu locīšanu piecos locījumos – nominatīvā, ģenitīvā, datīvā, akuzatīvā un lokatīvā – vienskaitlī un daudzskaitlī, arī daudzskaitlinieku vienskaitlinieku locīšanu atbilstošajās formās;
- I–III konjugācijas un nekārtno tiešo un atgriezenisko darbības vārdu locīšanu darāmās kārtas īstenības izteiksmes vienkāršās tagadnes, pagātnes un nākotnes formās;

- nenoteikto īpašības vārdu locīšanu sieviešu un vīriešu dzimtes formās, saskaņojot tos ar atbilstošo lietvārdu dzimtē, skaitlī un locījumā;
- atsevišķu vietniekvārdu grupu, konkrēti – personas vietniekvārdu, norādāmo vietniekvārdu, attieksmes (arī jautājamo) vietniekvārdu –, locīšanu;
- pamata un kārtas skaitļa vārdu locīšanu, t. sk. skaitļa vārdu savienojumus, piem., *divdesmit viens, četrdesmit pieci*.

Izvēloties, kuras vārdšķiras vārdu locīšanu apgūt, tiek piedāvāta papildu izvēle, piem., lietvārdiem – deklinācija, darbības vārdiem – konjugācija, nekārtnie darbības vārdi, tiešie vai atgriezeniskie darbības vārdi. (sk. 6. att.). Mācoties īpašības vārda locīšanas paradigmas, tiek piedāvāts izvēlēties, ar kuras deklinācijas lietvārdu to kopā locīt, saskaņot dzimtē, skaitlī, locījumā. Locīšanai tiek izmantots korpusā LaVA sastopamo vārdu krājums (sk. 2.1.).

Pēc tam, kad ir izraudzīta vārdšķira un atbilstošā kategorija, no LaVA vārdu saraksta automātiski tiek piedāvāts vārds locīšanai (piem., “Izloki *nodarbība*”), kurš pēc morfoloģiskajām pazīmēm atbilst vēlamajam. Piem., ja valodas apguvējs vēlas apgūt 1. deklinācijas lietvārdu locīšanu, tiek atlasīti vārdi, kuru vārdformas identifikators, 1. pazīme, ir [n], savukārt 6. pazīme – [1], t. i., [n....1]. No atlasītajiem vārdiem pēc nejaušības principa tiek piedāvāts viens.

Pildot uzdevumu, tiek piedāvātas vairākas iespējas: 1) pārbaudīt ierakstītās vārdformas (“Pārbaudīt”), 2) apskatīt pareizās atbildes (“Parādīt atbildes”), 3) izvēlēties citus vārdus (“Cits”). Jebkurā laikā var aplūkot pareizās atbildes un pēc vēlēšanās tās atkal paslēpt, nospiežot pogu “Parādīt atbildes”. To var darīt arī tad, ja valodas apguvējs nezina, kā piedāvāto vārdu pareizi locīt. Pēc pārbaudes pogas nospiešanas pareizās vārdformas tiek dotas uz zaļa fona, nepareizās vārdformas vai neaizpildītās vietas – iezīmētas ar sārto fonu.

Vārdformu ievietošanas/ierakstīšanas uzdevumi (Vārdi teikumos)

Valodas apguvē nozīmīgi ir tukšo vietu aizpildīšanas uzdevumi (*gap-fill exercise*), kas tiek izmantoti gan leksikas, gan gramatikas apguvē. Šāda tipa uzdevumi tiek izmantoti arī LaVA uzdevumu kopā, ļaujot valodas apguvējiem ierakstīt atbilstošu vārdformu vai izvēlēties atbilstošo no dotajām.

Korpusi piedāvā plašu, ērti un ātri iegūstamu piemēru klāstu, kas atspoguļo reālo valodu, un ir pierādīts, ka korpusa piemēru izmantošana pozitīvi ietekmē valodas apguvēju lingvistisko prasmju attīstību. Protams, valodas apguvējiem vismaz iesācēju līmenī varētu būt grūti šos teikumus saprast. Otra alternatīva ir mācību līdzekļu autoru mākslīgi veidoti, *neautentiski* teikumi, kas balstās uz pedagogu pieredzi un intuīciju. Veidojot korpusā balstītus uzdevumus, kuros vārds teikumā jāievieto vai jāieraksta atbilstošā formā, mēģinām rast vidusceļu – izmantojot skolotāju intuīciju un zināšanas, latviešu valodas prasmes līmeņa aprakstu un korpusā balstītus pētījumus, uzdevumos tiek ietverti mērķtiecīgi atlasīti, daudzveidīgi teikumi, kas tiek piedāvāti pēc nejaušības principa.

Locīšana

☒ Lietvārds

Deklinācija

[Atzīmēt visu](#) [Nonemt visu](#)

☒ 1. deklinācija
☐ 2. deklinācija
☐ 3. deklinācija
☐ 4. deklinācija
☐ 5. deklinācija
☐ 6. deklinācija

☐ Darbības vārds
☐ Vietniekvārds
☐ Īpašības vārds
☐ Skaitļa vārds

Turpināt

↖ 6. attēls. Gramatisko kategoriju izvēle locīšanas uzdevumos.

Vārdi teikumā

Ievieto doto vārdu pareizajā locījumā

Tu bija viss, ko varētu vēlēties.

Viņa vairs neapmeklēja teātrē rauc uz dažādu slaveno mākslinieku koncertiem Eiropas .

ga mūzika.

ga mūzika.

Deviņpadsmit gadu vecumā man ļoti sāpēja .

Viņai ir labas .

Ap viņu bija cilvēki ar siltām .

Ko tu te darīji?

Ko vajag sievietei, kura dzīvo brīvā ?

Jūs esat tauta bez .

Ko viņš dara ?

Pārbaudīt **Parādīt atbildes** **Citi teikumi**

↗ 7. attēls. Uzdevums 6. deklinācijas lietvārdu formu apguvei.

Teikumos tiek atrasti vārdi, kas ir vārdformu sarakstā, un tiek ģenerēts iespējamo vārdformu saraksts (ievietošanai). Teikumos var būt dažādu vārdšķiru vārdu dažādas formas, bet ievietošanai tiek piedāvātas tās pašas vārdformas, kuru apguvi trenēt ir iespējams locīšanas uzdevumos, tātad 1.–6. deklinācijas lietvārdi utt. Līdzīgi kā locīšanas uzdevumos, arī vārdformu ievietošanas uzdevumos piemēri tiek ģenerēti tikai ar tām vārdšķirām, kuras ir atlasītas izvēlnē.

Uzdevumā pēc nejaušības principa tiek parādīti desmit teikumi, kuros katrā ir viena tukša vieta. Teikumā ne vienmēr pietiek informācijas, lai noteiktu, kāda forma tajā bijusi sākotnēji (piem., vai darbības vārds lietots tagadnes, pagātnes vai nākotnes formā), tāpēc pirms tukšās vietas ir norādīta ievietojamā vārda pamatforma un papildu lingvistiskā informācija (piem., darbības vārdam – laiks, skaitlis, persona). To izmantojot, apguvējam jāieraksta tukšajā vietā pareizā vārdforma. (Sk. 7. att.)

Uzdevumu var pildīt arī kā daudzizvēļu uzdevumu. Aiz tukšā lauka ir poga ar jautājuma zīmi. Nospiežot to, atveras saraksts ar piecām attiecīgajām vārdformām, no kurām viena ir pareizā. Nospiežot uz izvēlētas formas, tā automātiski tiek ierakstīta tukšajā vietā. Jebkurā laikā ir iespējams arī aplūkot vai paslēpt pareizās atbildes, atkārtoti nospiežot pogu “Parādīt atbildes”, vai ģenerēt jaunus piemērus.

Nobeigums

Dažādi valodas korpusi, bet jo īpaši apguvēju valodas korpusi, ir izmantojami mācību uzdevumu izveidē. Ja apguvēju valodas korpusi ir marķēti dažādos līmeņos, uzdevumus, izmantojot dažādus dabiskās apstrādes un analīzes rīkus, kā arī atbilstošu metodoloģiju, ir iespējams veidot automātiski. Valodas apguvēju korpusu parāda, uz kādu valodas elementu pārbaudi vajadzētu koncentrēties, savukārt lielākai piemēru dažādībai var izmantot citus mērķvalodas korpusus, atlasot tajos tekstvienības, kas atbilst valodas apguvēju korpusā konstatētajām.

No korpusa LaVA, LVK2018 un SPK paraugteikumu kopas, kas tika izveidota pēc īpašiem kritērijiem un metodoloģijas, tika automātiski veidoti dažādu veidu pašpārbaudes uzdevumi latviešu valodas apguvei A1 un A2 līmenī. Balstoties uz korpusa LaVA kļūdu analīzi, tika izveidoti trīs dažādu uzdevumu tipi: teksta pārrakstīšanas uzdevumi, vārdu locīšanas uzdevumi un vārdformu ievietošanas un ierakstīšanas uzdevumi. Visu veidu uzdevumus valodas apguvējs var veikt patstāvīgi jebkurā laikā sev ērtā vietā, pēc vajadzības aplūkojot arī pareizās atbildes. Izveidotie mācību uzdevumi ir pieejami ikvienam interesentam korpusa LaVA tīmekļvietnes sadaļā “Uzdevumi” (skat. <http://lava.korpus.lv>). Uzdevumu kopu paredzēts aprobēt, iesaistot latviešu valodas apguvējus, kas sāk mācīties latviešu valodu. Tāpat uzdevumu kopu ir iespējams papildināt ar citiem uzdevumu tiem.

Automātiski ģenerēti uzdevumi ne tikai ļauj docētājam izmantot jau gatavu materiālu bez nepieciešamības to veidot no jauna vai patērēt laiku atbilžu pārskatīšanai un labošanai, bet arī sniedz iespēju apguvējam šos uzdevumus pildīt neierobežotu skaitu reižu. Turklāt tas, ka ir iespējams izvēlēties, piem., lietvārda deklināciju, no kuras atlasīt piemērus, ļauj šos uzdevumus izmantot atkārtoti dažādos valodas apguves posmos, pēc vajadzības atlasot attiecīgi apgūstamās lingvistiskās kategorijas vai arī visas līdz šim apgūtās. Arī uzdevuma grūtības pakāpi var samērot ar pedagoģiskā procesa vajadzībām, piem., uzdevumu ar formu ievietošanu teikumā pildīt kā daudzizvēļu uzdevumu. Uzdevumi ir izmantojami ne tikai docētāja vadītā mācību procesā, bet arī pašmācībā, piem., apgūstot noteiktu vārdšķiru locīšanas paradigmas. Plašās lietojuma iespējas ar iespējām pielāgot uzdevumus konkrēta apguvēja vajadzībām ir nozīmīga automātiski ģenerētu uzdevumu priekšrocība.

Šobrīd uzdevumi tiek piedāvāti tikai latviešu valodas apguvei A1 un A2 līmenī, tomēr, izmantojot rakstā aprakstītos principus, ir iespējams modelēt uzdevumus arī augstāku latviešu valodas prasmes līmeņu apguvei.

- Aužiņa, Ilze, Kaija, Inga, Levāne-Petrova, Kristīne, Dargis, Roberts, Pokratniece, Kristīne (2021). Latviešu valodas apguvēju korpusa (LaVA) izmantošana pētniecībā un mācību uzdevumu izstrādē. *Latviešu valodas apguve. XIII Starptautiskais baltistu kongress: rakstu krājums*. Liepāja: LiePA, 142.–161. lpp.
- Aužiņa, Ilze, Kaija, Inga, Levāne-Petrova, Kristīne (2020). Mērķhipotēžu izvirzīšana latviešu valodas apguvēju korpusā. *Valoda: nozīme un forma*, 11. Rīga: LU Akadēmiskais apgāds, 7.–26. lpp.
- Aužiņa, Ilze, Levāne-Petrova, Kristīne, Dargis, Roberts (2019). Latviešu valodas apguvēju kļūdu analīze: pareizrakstības kļūdas. Smiltieci, Gunta, Lauze, Linda (atb. red.). *Vārds un tā pētīšanas aspekti*, Nr. 23(1/2). Liepāja: LiePA, 220.–227. lpp.
- Crombie, Margaret A. (2000). Dyslexia and the learning of a foreign language in school: where are we going? *Dyslexia*, No. 6(2), pp. 112–123.
- Granger, Sylviane (2003). International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, No. 37(3), pp. 538–546.
- Kaija, Inga (2020). Jaunu burtu veidošana ar diakritiskajām zīmēm latviešu valodas kā svešvalodas apguvēju tekstos. *Valodu apguve: problēmas un perspektīva: zinātnisko rakstu krājums*. Liepāja: LiePA, 102.–110. lpp.
- Laizāne, Inga (2018). Latviešu valodas kā svešvalodas apguve Baltijas valstīs. *Valodu apguve: problēmas un perspektīva: zinātnisko rakstu krājums*. Liepāja: LiePA, 28.–39. lpp.
- Leech, Geoffrey (1998). Preface. *Learner English on computer*. Granger, Sylviane (ed.). London: Addison Wesley Longman, xiv–xx.
- Levāne-Petrova, Kristīne, Aužiņa, Ilze, Pokratniece, Kristīne (2020). Latviešu valodas apguvēju korpusa datu ieguves un apstrādes metodoloģijas izstrāde. *Valodu apguve: problēmas un perspektīva: zinātnisko rakstu krājums*. Liepāja: LiePA, 299.–309. lpp.
- Levāne-Petrova, Kristīne (2012). Līdzsvarots mūsdienu latviešu valodas tekstu korpusu un tā tekstu atlases kritēriji *Baltistica. VIII. Priedas*. Vilnius: Vilniaus Universiteto leidykla, 89.–98. lpp.
- Levāne-Petrova, Kristīne (2019). Līdzsvarotais mūsdienu latviešu valodas tekstu korpusu, tā nozīme gramatikas pētījumos. *Valoda: nozīme un forma*, 10. Rīga: LU Akadēmiskais apgāds, 131.–146. lpp.
- Levāne-Petrova, Kristīne, Pokratniece, Kristīne (2021). Skolēnu pārspriedumu korpusa izveide. *XIII Starptautiskais baltistu kongress*. Liepāja: Liepājas Universitāte, 106.–118. lpp.
- Nesselhauf, Nadja (2005). *Collocations in a learner corpus*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- O'Keeffe, Anne, McCarthy, Michael, Carter Ronald (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge University Press.
- Pilán, Ildikó, Vajjala, Sowmya, Volodina, Elena (2016). A readable read: automatic assessment of language learning materials based on linguistic complexity. *International Journal of Computational Linguistics and Applications*, No. 7(1), pp. 143–159.
- Pilán, Ildikó, Volodina, Elena, Borin, Lars (2017). Candidate sentence selection for language learning exercises: from a comprehensive framework to an empirical evaluation. *arXiv preprint arXiv:1706.03530*
- Paikens, Pēteris, Rituma, Laura, Pretkalniņa, Lauma (2013). Morphological analysis with limited resources: Latvian example. *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA)*.
- Rychlý, Pavel (2007). Manatee/Bonito-A Modular Corpus Manager. *1st Workshop on Recent Advances in Slavonic Natural Language Processing*. Brno: Masaryk University, pp. 65–70.
- Smith, Simon, Avinesh, P.V.S, Kilgarraiff, Adam (2010). Gap-fill tests for language learners: corpus-driven item generation. *Proceedings of ICON-2010: 8th International Conference on Natural Language Processing*, pp. 1–6.
- Šalme, Arvils, Aužiņa, Ilze (2016). *Latviešu valodas prasmes līmeņi: pamatlīmenis A1, A2, vidējais līmenis B1, B2*. Rīga: LVA.
- Volodina, Elena, Johansson, Richard, Johansson, Sofie (2012). Semi-automatic selection of best corpus examples for Swedish: initial algorithm evaluation. *Proceedings of the Workshop on NLP for CALL*, Vol. 80, pp. 59–70.
- Zanetti, Arianna (2020). *NLP methods for the automatic*

generation of exercises for second language learning from parallel corpus data. Master Thesis. Gothenburg University. Available: https://gupea.ub.gu.se/bitstream/handle/2077/66583/gupea_2077_66583_1.pdf?sequence=1&isAllowed=y [accessed 18.03.2022.].

Use of the Language Corpora in Automatic Generation of Latvian Language Exercises

Ilze Auziņa, Roberts Dargis, Inga Kaija,
Kristīne Levāne–Petrova, Kristīne Pokratniece

Keywords: computational linguistics, language corpora, Latvian language acquisition, sentence selection, exercises

Today, language corpora are not only the empirical basis of research but can also be used in developing a variety of data-driven teaching materials and tools. The experience of other countries shows that the development of self-assessment exercises for language learning can be partially or fully automated using language corpora and natural language processing (NLP) tools, thus providing both a variety of exercises and support for teachers in the implementation of the curriculum.

The Latvian Language Learners Corpus (LaVA) developed at the Institute of Mathematics and Computer Science, University of Latvia, includes more than 1000 texts created by foreign Latvian language learners studying at Latvian higher education institutions for the first or second semester reaching A1 (possibly A2) Latvian language proficiency level. The size of the corpus is more than 180 000 words. According to the LaVA data analysis, including learners error analysis, exercises and tests are generated. Data analysis allows us to identify problematic spelling, grammar, and vocabulary issues. The exercises are intended to help the language learner to strengthen the linguistic competence of Latvian language, for example, the use of verb forms in the indicative mood, both in indefinite and perfect tense forms.

The article discusses the methodology according to which, based on the statistical and quantitative analysis of the LaVA corpus data, sample sentences are selected from different corpora of Latvian language, for example, *The Balanced Corpus of Modern Latvian* (LVK2018), *The Corpus of Students' Essays* (SPK), as well describes the task-development algorithms and development of online self-assessment exercises site.

Latviešu valodas *FrameNet* korpuss

Baiba Saulīte, Gunta Nešpore-Bērzkalne,
Laura Rituma, Viesturs Jūlijs Lasmanis,
Normunds Grūzītis

Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi” projekts Nr. VPP-IZM-DH-2020/1-0001 “Humanitāro zinātņu digitālie resursi: integrācija un attīstība”. Darbs uzsākts ERAF Pēcdoktorantūras pētniecības atbalsta projektā Nr. 1.1.1.2/VIAA/1/16/188 “No abstraktās nozīmes reprezentācijas līdz dabiskam teikumam un saistītam tekstam” un ERAF praktiskas ievirzes pētījumu programmas projektā Nr. 1.1.1.1/16/A/219 “Daudzslāņu valodas resursu kopa teksta semantiskai analīzei un sintēzei latviešu valodā”.

FrameNet ir starptautiski atzīta un dažādām valodām izmantojama valodas datu marķēšanas sistēma, kuras pamatā ir freimu semantika – leksiskās nozīmes teorija, kas veidota uz Čārļa Filmora (*Charles J. Fillmore*) un viņa kolēģu darbu pamata (Fillmore 1976; Fillmore et al. 2003; Fillmore and Baker 2010). Atbilstoši *FrameNet* (turpmāk tekstā – FN) pieejai tiek uzskatīts, ka katra vārda nozīmi vislabāk var saprast semantiskā situācijā jeb freimā – vispārinātā notikuma vai attieksmju modelī, kurā aprakstīti iespējamie situācijas dalībnieki un tādi situācijas apstākļi kā laiks, vieta u. c.

Kopš 2017. gada tiek veidots Latviešu valodas *FrameNet* korpus (turpmāk tekstā – *FrameNet-LV*), tajā manuāli marķētas FN sistēmā definētās semantiskās lomas, vienlaikus sastatot biežāk lietotos latviešu valodas verbu (1350 leksēmu) un deverbālos lietvārdus (250 leksēmu) ar tiem atbilstošajiem angļu valodas FN freimiem.

Šajā rakstā aplūkota freimu semantikas pieeja datu marķēšanā, aprakstīta *FrameNet-LV* izveide – datu atlase, marķēšanas principi, kā arī parādīts, kā datus var analizēt, izmantojot korpusa tīmekļvietni.

Freimu semantika

Atbilstoši FN pieejai teksta korpusā katrai izvēlētajai leksiskajai vienībai (nozīmei) tiek norādīts tai atbilstošais freims, kā arī tiek marķēti tā freima elementi (turpmāk tekstā – FE) jeb semantiskās lomas. Piemēram, cilvēka dzimšanas situāciju (BEING BORN ‘piedzimt’) raksturo freims, kurā ir FE Bērns (cilvēks, kurš piedzimis), Laiks un Vieta (situācijas apstākļi). Iespējamās arī sarežģītākas situācijas, kurās ir vairāk FE, piemēram, darba uzsākšanas situācija (HIRING ‘pieņemšana darbā’), kurā ir šādi FE: Darbinieks, Darba devējs, Darba vieta, Amats, Pienākums, Laiks, Vieta, Veids utt.

Atkarībā no tā, vai attiecīgā informācija situācijā ir konceptuāli svarīga vai bez tās situācija var pastāvēt pat tad, ja attiecīgā informācija tieši neparādās teikumā, FN pieejā tiek šķirtas centrālās, perifērās un ārpustematiskās semantiskās lomas (Ruppenhofer et al. 2016). Informācija, ko sniedz ārpustematiskās semantiskās lomas, rāda attieksmes starp vairākām situācijām, bet vienas semantiskās grupas analīzē īpaši aktuālas ir centrālās un perifērās semantiskās lomas.

1. Centrālās semantiskās lomas raksturo situācijā konceptuāli nepieciešamus elementus, kas atšķir situāciju no citām (Ruppenhofer et al. 2016). Piemēram, tirdzniecības situācija nevar pastāvēt bez Pircēja, Pārdevēja, Preces.

2. Perifēro semantisko lomu nosauktā informācija neatšķir situāciju no cita tipa situācijām (tās var palīdzēt nošķirt vienu pārvietošanās situāciju no citas pārvietošanās situācijas), bet sniedz papildinformāciju par situāciju, tās apstākļiem. Šīs semantiskās lomas līdzīgi var raksturot jebkuru semantiski atbilstošu situāciju (Ruppenhofer et al. 2006). Gandrīz jebkuru situāciju var raksturot, piemēram, semantiskās lomas Laiks (šobrīd), Ātrums (*lēnām*), Cēlonis (*aiz garlaicības*) u. tml.:

Jānis šobrīd aiz garlaicības lēnām iet pa ielu.

Jānis šobrīd aiz garlaicības lēnām lasa grāmatu.

Izmantojot šo semantiskā apraksta pieeju, kopš 1997. gada tiek veidota mūsdienu angļu valodas leksiskā datubāze – Bērklījas *FrameNet* (turpmāk tekstā – BFN¹), kurā iekļauti 1224 freimi, 13 685 leksiskās vienības un vairāk nekā 200 tūkstoši manuāli marķētu lietojuma piemēru².

Šobrīd šādi semantiski marķēti dati, kas nepieciešami gan valodniecības pētījumos, gan dabiskās valodas sapratnē un tekstrādē, ir pieejami ne tikai angļu valodai, bet arī vairākām citām valodām, ieskaitot latviešu valodu (Gilardi, Baker 2018).

Latviešu valodas *FrameNet* dati

FrameNet-LV ir daļa no “Daudzslāņu valodas resursu kopas”³ (FullStack-LV; Gruzitis et al. 2018a), kurā ievietotas no “Līdzsvarotā mūsdienu latviešu valodas korpusa” (LVK2018; Levāne-Petrova 2019) atlasītas rindkopas noteiktās reprezentatīvās proporcijās (60 % publicistikas, 20 % daīlliteratūras, 10 % normatīvie akti, 5 % Saeimas stenogrammas, 5 % citi teksti). Šo datu sintaktiskajā un semantiskajā marķēšanā izmantotas šādas dažādām valodām izmantojamās reprezentācijas: *Universal Dependencies* (UD; de Marneffe et al. 2021), FN, *PropBank* (Palmer et al. 2005) un *Abstract Meaning Representation* (AMR; Baranescu et al. 2013).

UD sintaktiskā reprezentācija latviešu valodas korpusam tiek automātiski atvasināta no sarežģītākās hibrīdās sintaktiskās marķēšanas sistēmas, kas ir izmantota “Latviešu valodas sintaktiski marķētajā korpusā” (Pretkalniņa et al. 2016). FN līmenis tiek marķēts uz UD pieejā sintaktiski marķētajiem datiem (Gruzitis et al. 2018b; Pretkalnina et al. 2018), savukārt no

1 *The Berkeley FrameNet lexical database*: <http://framenet.icsi.berkeley.edu>

2 https://framenet.icsi.berkeley.edu/fndrupal/current_status

3 Atvērtie dati pieejami: <https://github.com/LUMII-AILab/FullStack>

FN un sintaktiski marķētajiem datiem tiek atvasināts vispārīgākais *PropBank* semantiskā marķējuma slānis (Gruzītis et al. 2020). Papildus no zemākajiem marķējuma slāņiem daļēji automātiski var konstruēt AMR grafus.

Datu atlase

Latviešu valodas verbu lietojuma biežums ņemts vērā, jau veidojot FullStack-LV datu kopu – no 2000 LVK2018 biežāk lietotajiem latviešu valodas vārdiem tika izveidots biežuma saraksts un, atlasot rindkopas, tika pārbaudīts, lai katrs no šiem vārdiem FullStack-LV parādās vismaz 10 reizes.

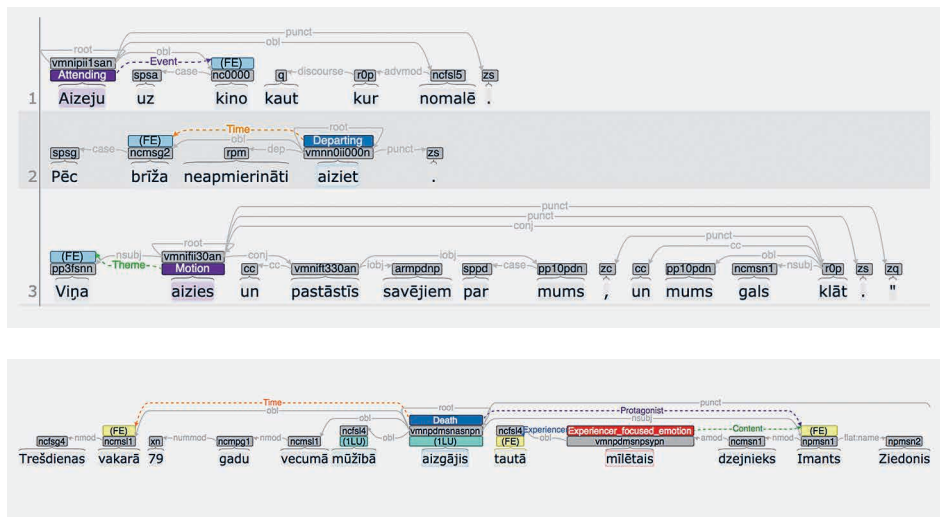
Tas pats verbu biežuma saraksts arī izmantots, tālāk no FullStack-LV izvēloties valodas materiālu semantiskajai marķēšanai. Priedēkļverbi, izņemot noliegtos verbus, tiek uzskatīti par atsevišķām vienībām, savukārt adjektīviski un substantīviski divdabji netiek šķirti no vārdiem, kas ir to pamatā.

Deverbālu lietvārdu atlase bija sarežģītāka, jo lielāko daļu šo atvasinājumu nevar atlasīt no LVK2018 pēc morfoloģiskajām pazīmēm. Tāpēc deverbāli lietvārdi, kuru apkaimē parādās verbam raksturīgi paplašinātāji, pēc biežuma un sintaktiskās apkaimes tika konstatēti jau izveidotajā FullStack-LV. Tie atlasīti, gan ņemot vērā vārda sastāvu (regulāriem atvasinājumiem ar -šana, -šanās), gan “Latviešu valodas sintaktiski marķētajā korpusā” meklējot lietvārdus, kuriem piesaistīti, piemēram, sekundāri predikatīvi komponenti (*vēlme komponent*) vai ar prepozicionālu konstrukciju izteikti apstākļi (*brauciens pie Buša*). Valodas materiāls, kas ar regulārajām izteiksmēm atrasts FullStack-LV, tiek pārskatīts, un deverbālie lietvārdi, kam ir verbam raksturīgi paplašinātāji un vismaz seši lietojuma piemēri, marķēti semantiski.

Datu marķēšana

FN līmenis tiek marķēts, izmantojot nevis sākotnējo rindkopu griezumus, bet gan konkrētā verba vai substantīva konkordances. Tas nozīmē, ka freimu marķēšanai no sintaktiski marķētajiem datiem tiek atlasīti teikumi, kuros parādās freimu marķēšanai izvēlētais verbs vai substantīvs (sk. verba *aiziet* leksiskās vienības 1. att.). Līdz ar to viens un tas pats teikums var parādīties vairākās šādās konkordanču kopās. Pēc tam teikumi un to marķējums atkal tiek sapludināti oriģinālajās rindkopās (2. att. sk. teikumu, kurā sapludināts divu freimu – DEATH ‘nāve’ un EXPERIENCER FOCUSED EMOTION ‘izjutēja emocijas’ – marķējums). Šāda pieeja nepieciešama, lai valodnieks, kas marķē freimus, vienkopus redz visus teikumus ar izvēlēto vārdu un konsekventāk var novērtēt, kurai nozīmei kurš freims vislabāk atbilst.

Tā kā perifērie un ārpustematiskie FE vairāk raksturo tekstā aprakstīto situāciju kopumā, nevis konkrētās leksiskās vienības leksisko nozīmi, latviešu valodas datos tiek marķēti galvenokārt centrālie FE. Perifēros FE marķē tikai gadījumos, ja leksisko vienību raksturo FE Laiks un Vieta, kā arī tad, ja perifērie FE kādai leksiskās vienības nozīmei ir specifiski, piemēram, verba *atgriezties* apkaimē teikumā *Jānis Vasīļonoks nesen atgriezās mājās no tāla ceļojuma*



↑ 1. attēls. Verbs *aiziet* konkordanču griezumā un tā semantiskais marķējums.

↓ 2. attēls. Teikums *Trešdiena vakarā 79 gadu vecumā mužībā aizgājis tautā mīlētais dzejnieks Imants Ziedonis*, kurā apvienots divu freimu marķējums.

blakus centrālajiem FE Tēma (*Jānis Vasiļonoks*) un Mērķis (*mājās*) tiek marķēts arī perifērais FE Avots (*no ceļojuma*).

FN marķēšanai tiek izmantota datu marķēšanas platforma *WebAnno*⁴ (Eckart de Castilho et al. 2016), kurai ir tīmekļa saskarne, plašas konfigurēšanas un pielāgošanas iespējas, kā arī ir piešķirts CLARIN pētnieciskās infrastruktūras atbalsts.

Semantiski tiek marķēts teikums, kas jau iepriekš ir sintaktiski marķēts UD pieejā (sk. 1. un 2. attēlu):

1. tiek atzīmēta marķējamā leksiskā vienība, tai izvēlas atbilstošo freimu (sk. freimus dažādām verba *aiziet* nozīmēm: ATTENDING ‘apmeklēšana’, DEPARTING ‘došanās prom’, MOTION ‘kustība’);
2. no konkrētā freima centrālo semantisko lomu saraksta izvēlas teikumā realizētās centrālās semantiskās lomas (sk. adverbīālu paplašinātāju *uz kino*, kam atbilst centrālā semantiskā loma *Event* ‘notikums’, 1. teikumā un teikuma priekšmetu *viņa*, kam atbilst centrālā semantiskā loma *Theme* ‘tēma’ 3. teikumā);
3. kur nepieciešams, pievieno perifērās semantiskās lomas (sk. adverbīālu paplašinātāju *pēc brīža*, kam atbilst perifērā semantiskā loma *Time* ‘laiks’).

⁴ <https://webanno.github.io/webanno/>

Marķējot semantiskā loma tiek piešķirta vienam konkrētam vārdam teikumā, bet, tā kā marķējums ir veikts kā nākamais līmenis pēc sintaktiskā marķējuma, visu frāzi, kurā ietilpst semantiskā loma, var iegūt automātiski, izvēloties visus no konkrētā vārda atkarīgos vārdus sintaktiskajā struktūrā: sk., piemēram, 2. attēlā – FE atzīmēts uz vārdformas *Imants*. Šai vārdformai sintaktiski ir piesaistītas citas vārdformas, un kopā veidojas frāze *tautā mīlētais dzejnieks Imants Ziedonis*. Pilnu frāzi plānots rādīt lietojuma piemēros FrameNet-LV.

No vienas puses, tas atvieglo marķēšanas procesu, jo valodniekam nav jāieezīmē konkrētās frāzes robežas, kas paildzinātu procesu. No otras puses, tas prasa noteiktas priekšzināšanas par to, kā veidojas sintaktiskā struktūra pielikuma, vienlīdzīgu teikuma locekļu un citos gadījumos, lai lomu piešķirtu tam vārdam, kam sintakses reprezentācijā būs pakārtoti visi pārējie frāzes vārdi (piem., atzīmējot FE, jāzina, ka frāzes *Imants Ziedonis* sintaktiskajā struktūrā otrā vārdforma tehniski ir pakārtota pirmajai, nevis otrādi, sk. 2. att.).

Freimu semantikā leksiskā vienība var sastāvēt arī no vairākiem vārdiem, tāpēc, veidojot latviešu valodas FN korpusu, atsevišķos gadījumos tiek izmantota iespēja norādīt, ka freima leksiskā vienība sastāv no vairākām vārdformām, jo verbs veido frazeoloģisku vienību, kurai atbilst kāds noteikts freims (sk. 2. att.: frazeoloģisms jeb leksiskā vienība *aiziet mūžībā*, kuras nozīmei atbilst freims DEATH ‘nāve’).

Sastatot latviešu valodas vārdu nozīmes ar BFN sistēmā piedāvātajiem freimiem, atsevišķos gadījumos rodas grūtības piemeklēt latviešu valodas verbam atbilstošu BFN freimu, jo:

1. BFN daudziem jēdzieniem vēl nav definēti freimi, piem., nozīmēm *veltīt*, *balsot*, *ziedēt*;
2. nav pietiekami precīzu atbilstmju starp angļu valodas un latviešu valodas jēdzieniem – latviešu valodas verba nozīmei atbilst angļu valodas vārdu savienojums (*kļūdīties*, *maldīties* ‘to be wrong’; ‘to make a mistake’) vai otrādi (*krist ģībonī*, *zaudēt samaņu* ‘to faint’), atšķiras latviešu un angļu verba nozīmes vispārinājuma pakāpe (*braukt* ‘to use vehicle’ – latviešu verba nozīmē nav ietverta informācija, vai persona vada transportlīdzekli vai ir tā pasažieris, bet angļu verbos šis nozīmes elements ir ietverts – attiecīgi verbi iekļaujas freimos RIDE VEHICLE ‘braukt (kā pasažierim) transportlīdzeklī’ vai OPERATE VEHICLE ‘vadīt transportlīdzekli’;
3. latviešu valodas verba nozīmei atbilst nomināls BFN freims, jo verba nozīme angļiski izteikta ar palīgverbu un adjektīvu, piem., – *piedzerties* ‘to get drunk’ ietilpst freimā INTOXICATION ‘saindēšanās’.

Šādos gadījumos tiek meklēti papildu risinājumi, piem., nozīmes, kurām trūkst freimu, tiek marķētas kā UNDEFINED un tiek apkopotas, lai FN sistēmā var ieteikt jaunu freimu izveidi; ja verba nozīmei atbilst nomināls freims, tā tiek aprakstīta ar nominālu freimu (Nešpore-Bērkalne et al. 2018).

Latviešu valodas FN korpusa tīmekļvietne

Līdz šim latviešu valodas korpusa semantiskajā marķēšanā ir izmantoti 570 BFN freimi, marķēts 2900 leksisko vienību (t. sk. 325 lietvārdi; vidēji 5,1 leksiskā vienība uz freimu) un gandrīz 26 000 lietojuma piemēru (vidēji 8,9 piemēri uz leksisko vienību)⁵. Lai šie dati pēc iespējas ērtāk būtu izmantojami lingvistiskiem pētījumiem, LU Matemātikas un informātikas institūtā (LU MII) ir izveidota FrameNet-LV tīmekļvietne⁶, kurai par paraugu ņemta BFN vietne⁷ (sk. 3. attēlu, kur norādīta freima ASSISTANCE ‘palīdzība’ definīcija, kā arī skaidroti centrālie freima elementi).

Freimus un leksiskās vienības var apskatīt kā sarakstu vai meklēt konkrētu freimu vai leksisko vienību. Ja izvēlas freimu, var aplūkot visas leksiskās vienības, kas marķētas, izmantojot šo freimu (sk. 5. att.).

Ja izvēlas leksiskās vienības, var redzēt visus freimus, ar kuriem katra leksiskā vienība marķēta (sk. 6. att.). Sadaļā “Leksiskās vienības” var atlasīt konkrētu vārdšķiru, piemēram, tikai lietvārdus (ierakstot meklētajā *noun*, tiek atlasīts lietvārdu saraksts).

Par konkrētu leksisko vienību pieejama šāda informācija:

1. visi tās lietojuma piemēri ar tajos marķētām semantiskajām lomām (sk. 4. attēla labo pusī);
2. pārskats par to, kādi FE ir šai nozīmei un kā tie realizējas gramatiski UD pieejā (sk. 7. attēlu);
3. pārskats par to, kādi valences modeļi jeb šabloni (visas semantisko lomu gramatiskās realizācijas kombinācijas, kas marķētas piemēros) veidojas šai nozīmei (sk. 8. attēlu).

FE gramatiskā realizācija norādīta saskaņā ar UD pieejā izmantoto vārdšķiru klasifikāciju⁸ un definētajām universālajām sintaktiskajām attieksmēm⁹, sk., piem., vārdšķiras PRON ‘vietniekvārds’, PROPIN ‘īpašvārds’, VERB ‘darbības vārds’ un sintaktiskās attieksmes *nsubj* ‘teikuma priekšmets’, *obj* ‘tiešais papildinātājs’, *iobj* ‘netiešais papildinātājs’ 7. un 8. att. Atsevišķos gadījumos norādīti arī apakštipi, piem., *nsubj:pass* – teikuma priekšmets teikumos ar izteicēju ciešamajā kārtā.

5 Atvērtie dati pieejami: <https://github.com/LUMII-AILab/FullStack/tree/master/FrameNet>

6 <http://framenet.korpuss.lv/>

7 <https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>

8 <https://universaldependencies.org/u/pos/index.html>

9 <https://universaldependencies.org/u/dep/index.html>

Assistance

Definition:

A **Helper** benefits a **Benefited_party** by enabling the culmination of a **Goal** that the **Benefited_party** has. A **Focal_entity** that is involved in reaching the **Goal** may stand in for it.

FEs:

Core:

Benefited_party The **Benefited_party** receives a benefit from the action of the **Helper**.

Focal_entity This FE identifies a **Focal_entity** involved in achieving the **Goal**.

Goal **Goal** The desirable state of affairs that the **Benefited_party** is involved in and which is enabled by the **Helper**.

Helper The **Helper** performs some action that benefits the **Benefited_party**.

Assistance

Leksiskās vienības:

[apkalpot.VERB](#) (3)

[līdzēt.VERB](#) (4)

[palīdzēt.VERB](#) (84)

[uzturēt.VERB](#) (4)

[palikt.VERB](#) (Becoming)

[palikt.VERB](#) (Coming_to_be)

[palikt.VERB](#) (Getting)

[palikt.VERB](#) (Losing)

[palikt.VERB](#) (Remainder)

[palikt.VERB](#) (State_continue)

Semantiskie freimi	Leksiskās vienības	Korpusa piemēri	Atsauces	Latvian FrameNet																			
palīdzēt.VERB (Assistance) Visi leksiskās vienības lietojuma piemēri: 84 Freima elementi, to biežums un gramatiskā realizācija: <table><tr><th>Freima elementi</th><th>Piemēri</th><th>Realizācija</th></tr><tr><td>Helper</td><td>65</td><td>nsobj PRON (33) nsobj NOUN (22) nsobj PROPN (3) nsobj X (2) obj NOUN (1) nsobj VERB (1) nsobj-pass PRON (1)</td></tr><tr><td>Goal</td><td>56</td><td>comp VERB (53) dep VERB (1) obj VERB (1) acl VERB (1)</td></tr><tr><td>Benefitted party</td><td>45</td><td>obj NOUN (23) obj PRON (17) obj PROPN (2) obj NUM (1) obj VERB (1) comp VERB (1)</td></tr><tr><td>Focal entity</td><td>4</td><td>obj NOUN (3) obj NOUN (1)</td></tr><tr><td>Time</td><td>1</td><td>obj PRON (1)</td></tr></table>					Freima elementi	Piemēri	Realizācija	Helper	65	nsobj PRON (33) nsobj NOUN (22) nsobj PROPN (3) nsobj X (2) obj NOUN (1) nsobj VERB (1) nsobj-pass PRON (1)	Goal	56	comp VERB (53) dep VERB (1) obj VERB (1) acl VERB (1)	Benefitted party	45	obj NOUN (23) obj PRON (17) obj PROPN (2) obj NUM (1) obj VERB (1) comp VERB (1)	Focal entity	4	obj NOUN (3) obj NOUN (1)	Time	1	obj PRON (1)	Nodzīst teikumus X [a-p1039-p3s5] Varbūt tas ir vairāk ārēji - vainadzīgi, svečtes, bet galūms noteikti PALIDZ arī iekšāji cilvēkiem jūsties labāk. X [a-p159-p3s2] Kad Viņa atgriezās uz dzīvi šeit, daudz ziedoja dažādiem projektiem - lāpļa jumtus, PALIDZĒJA skolām iegādājot datorus un grāmatas. X [a-p2379-p8s2] Viņa informēja, ka Latvijā patlaban ir recepšu medikaments « Champix », kas efektīgi PALIDZ smēķētājam atmest kaitīgo ieradumu. X [a-p3344-p8s5] Bet manā uzdevē apkalpojošā sfērā strādājošajiem ir JAPALIDZ cilvēkiem! X [a-p3533-p3s4] Pupas satur daudz šķiedrvielu un pektīnu, kas PALIDZ izvadīt no organisma smago metālu sāļus. X [a-p580-p1s1] Tas PALIDZ sabalansēt ģimenes budžetu. X [a-p3315-p3s12] Ar mūsu mums ir labas attiecības, viņai ir divas lielas meitas, mēs PALIDZAM viena otra. X [a-p3753-p3s3] Viens no elementiem, kas PALIDZ noturēt speciālistus, ir projektu saturs, jo Latvija ir ceturtda valsts Eiropā interneta pieslēgumu ātruma ziņā, un ekspertiem ir vienreizēja iespēja strādāt ar šīm tehnoloģijām. X [a-p3116-p3s1] Tāpā veicina vielaipais intensitāti, tādējādi diētu laikā PALIDZ novāgt. X [a-p12740-p1s1] 4. maijā, atzīmējot Latvijas Republikas Neatkarības atjaunošanas dienu, daudzviet tika svinēti Baltā galdauta svētki, kas tiek iedibināti kā jauna tradīcija, lai pie kopēja galdā sanāktu tuvinieki, draugi, kaimiņi un nosvinētu savas valsts esamību, atcerētos tos, kuri PALIDZĒJA tas atdzimst.
Freima elementi	Piemēri	Realizācija																					
Helper	65	nsobj PRON (33) nsobj NOUN (22) nsobj PROPN (3) nsobj X (2) obj NOUN (1) nsobj VERB (1) nsobj-pass PRON (1)																					
Goal	56	comp VERB (53) dep VERB (1) obj VERB (1) acl VERB (1)																					
Benefitted party	45	obj NOUN (23) obj PRON (17) obj PROPN (2) obj NUM (1) obj VERB (1) comp VERB (1)																					
Focal entity	4	obj NOUN (3) obj NOUN (1)																					
Time	1	obj PRON (1)																					

↖ 3. attēls. Freima ASSISTANCE 'palīdzība' definīcija un centrālo FE uzskaitījums un skaidrojums BFN tīmekļvietnē.

↓ 4. attēls. FrameNet-LV tīmekļvietne: kreisajā pusē – leksiskā vienība *palīdzēt* (freims ASSISTANCE 'palīdzība'), marķēto FE pārskats; labajā pusē – lietojuma piemēri, kuros ar melnu atzīmēts marķētais verbs, ar citām krāsām – FE ar tā paplašinātājiem, kas vizuāli ierāmēti.

↗ 5. attēls. Leksiskās vienības freimā ASSISTANCE 'palīdzība' pie katras iekavās norādīts marķēto piemēru skaits.

↘ 6. attēls. Leksiskās vienības *palikt* dažādos freimos.

Freima elementi	Piemēri	Realizācija
Helper	65	nsubj PRON (33) nsubj NOUN (22) nsubj PROP (5) nsubj X (2) obj NOUN (1) nsubj VERB (1) nsubj:pass PRON (1)
Goal	56	ccomp VERB (53) dep VERB (1) iobj VERB (1) acl VERB (1)
Benefited party	45	iobj NOUN (23) iobj PRON (17) iobj PROP (2) iobj NUM (1) iobj VERB (1) ccomp VERB (1)
Focal entity	4	obl NOUN (3) iobj NOUN (1)
Time	1	obl PRON (1)

7. attēls. Leksiskās vienības *palīdzēt* (ASSISTANCE ‘palīdzība’) FE, to biežums un gramatiskā realizācija.

Piemēri	Valences šablons			
24	Goal	Helper		
11	ccomp VERB	nsubj PRON		
10	ccomp VERB	nsubj NOUN		
1	dep VERB	nsubj NOUN		
1	ccomp VERB	nsubj X		
1	ccomp VERB	nsubj:pass PRON		
20	Benefited party	Goal	Helper	
7	iobj NOUN	ccomp VERB	nsubj PRON	
5	iobj PRON	ccomp VERB	nsubj PRON	
3	iobj NOUN	ccomp VERB	nsubj NOUN	
1	iobj PROP	ccomp VERB	nsubj PRON	
1	iobj NOUN	iobj VERB	nsubj NOUN	
1	iobj NOUN	ccomp VERB	nsubj PROP	
1	iobj NOUN	acl VERB	nsubj PRON	
1	iobj PRON	ccomp VERB	nsubj X	

8. attēls. Leksiskās vienības *palīdzēt* (ASSISTANCE ‘palīdzība’) semantiskās un gramatiskās valences modeļu piemērs.

7. attēlā atspoguļots, kādi FE, cik reizu parādījušies marķētajos piemēros, kā arī norādīta realizēto FE leksēmu gramatika. Piem., FE *Helper* ‘palīgs’ realizēts kopumā 65 lietojuma piemēros, bet visbiežāk (33 reizes) tas izteikts ar vietniekvārdu teikuma priekšmeta funkcijā.

Lietojuma piemērus var atlasīt visus kopā vai kārtot, noņemot nost lieko, rādīt tikai teikumus, kuros ir konkrēti FE konkrētā gramatiskajā realizācijā, noteiktas FE kombinācijas u. tml. Piemēram, visi teikumi, kuros parādās FE *Helper* ‘palīgs’, kas gramatiski izteikts ar īpašvārdu teikuma priekšmeta funkcijā (sk. 9. att.). Turpinot piemēru atlasī, var nodzēst visus teikumus vai izvēlēties un pievienot vēl kādas kombinācijas.

Visas kombinācijas (3 piemēri), kur realizēti FE *Benefited party* ‘ieguvējs’ (sugasvārds netiešā papildinātāja funkcijā), *Goal* ‘mērķis’ (darbības vārds *ccomp* funkcijā (atbilstoši latviešu valodas sintakses teorijai – sekundāri predikatīvs komponents)), *Helper* ‘palīgs’ (sugasvārds teikuma priekšmeta funkcijā), redzamas 10. attēlā.

[X] [a-p661-p8s4] Saulkrastietes māsa **Arnita** un **Evita Čistjakovas** asinis nodod regulāri un ir priecīgas **PALIDZĒT** **nelaimē nokļuvušajiem** .

[X] [a-p7161-p5s1] **Ziedus izkārtot vāzēs** **dziedātāji** **PALIDZEJA** **aktrise Santa** Didžus , kura ņēma Lienei sadāvinātos ziedus un lika vāzēs , ļaujot jubilei sagaidīt viesus un neuztraukties par puķu vāzēm .

[X] [a-p1396-p10s1] — Vai **jums** **Eiropa** **PALIDZ ?**

[X] [a-p13056-p8s1] Jēkabpīlieši , **kuriem** savainojumu dēļ nevar **PALIDZĒT** **centra spēlētāji Jānis Pozņaks un Pāvels Veselovs** , piedzīvojuši jau ceturto zaudējumu pēc kārtas .

[X] [a-p12815-p7s4] **Malvinē** bija vecākā , auga paklausīga un strādīga , darīja visu , ko vecāki lika , **PALIDZEJA** **saimniecībā** un pieskatīja jaunākos bērnus .

[X] [a-p1039-p3s5] Varbūt tas ir vairāk ārēji - vainadzīni , svecītes , bet **gaišums** noteikti **PALIDZ** **arī iekšēji cilvēkiem** **justies labāk** .

[X] [a-c58-p41s11] Tos vislabāk **PALIDZĒS** **izvēlēsies** **katrai sievietei** **individuāli speciālists** , ņemot vērā pacientes vēlmes , gaudi , iespējas , krāsu sākotnējo izskatu , krāšu galu atrašanās vietu un augstumu , kā arī krāsu kurvja izmērus un auguma proporcijas .

[X] [a-p5561-p5s1] **Labu miegu** **izvaigņei** **PALIDZ** **saglabāt** **arī Harija Potera grāmatu lasīšana gultā** , jo aizraujošais stāsts palīdz nedomāt par savu dzīvi , bet pārnes sievieti grāmatas radītajā realitātē .

↑ 9. attēls. Ar īpašvārdu teikuma priekšmeta funkcijā izteikts FE *Helper* ‘palīgs’ leksiskās vienības *palīdzēt* apkaimē.

↓ 10. attēls. Viens gramatiskās valences modelis FE *Benefited_party* ‘ieguvējs’, *Goal* ‘mērķis’, *Helper* ‘palīgs’.

FN korpusa tīmekļvietne piedāvā datus aplūkot dažādos griezumos, piemēram, var salīdzināt semantisko lomu realizāciju pie konkrētiem vārdiem un deverbāliem lietvārdiem – freimā PROTECTING ‘aizsardzība’ ietilpst gan verbs *aizsargāt*, gan arī lietvārds *aizsardzība*. Šī freima centrālās semantiskās lomas ir *Protection* ‘aizsargs’, *Asset* ‘manta’ un *Danger* ‘apdraudējums’. Verbam tās tipiski realizējas kā teikuma priekšmets, tiešais papildinātājs un netiešais papildinātājs, piemēram, fragmentā *12 bruņumašīnas var kādu zemes pleķi aizsargāt*. Savukārt deverbālā lietvārda *aizsardzība* apkaimē parasti realizējas tikai lomas *Asset* ‘manta’ (*nodarbināto*) un *Danger* ‘apdraudējums’ (*pret risku*): *nodarbināto aizsardzība pret darba vides trokšņa radīto risku*.

Tāpat var aplūkot, kādas semantiskās lomas parādās leksiskās vienības apkaimē, kā tās realizētas sintaktiski, kāda ir to tipiskā secība un ar ko skaidrojamas atkāpes no tipiskās secības. Plašais marķēto verbu un to lietojuma piemēru klāsts ļauj pētīt arī kādas noteiktas semantiskās grupas verbus.

Nobeigums

Semantiski marķēto datu kopa joprojām tiek papildināta, un FN korpusa tīmekļvietne tiek uzlabota, bet jau šobrīd skaidri redzams, ka šādi dati un piekļuve tiem ir nepieciešami gan valodu pētniekiem, gan valodas tehnoloģijām.

Marķēšana starptautiski pazīstamā un vienotā datu formātā ļauj latviešu valodas resursus integrēt starptautiski atzītās daudzvalodu datu kopās, vienotais datu formāts nodrošina, ka latviešu valodas dati kļūst pieejami un izmantojami arī valodu pētniekiem, kuri neprot latviešu valodu. Visus FN pieejā balstītos resursus un uz to pamata izstrādātās lietotnes kopējā daudzvalodu tīklā apvieno iniciatīva *Global FN*, kuras mērķis ir kopīgi FN balstīti pētījumi¹⁰. Viena no tās darbībām ir daudzvalodu FN pieejā marķēts paralēlais korpus – viens teksts dažādās valodās (*TED Talk parallel corpus*; Torrent et al. 2018; Ohara 2020), kurā iekļauts arī marķēts teksts latviešu valodā.

Valodas pētniekiem šāds apjomīgs, ērti pieejams semantiski marķēts korpus ļauj analizēt vārdu nozīmes un to lietojumu tekstā, bet dabiskās valodas apstrādē tā ir nozīmīga semantiskās marķēšanas treniņdatu kopa, ko var izmantot mašintulkošanā, teksta sapratnē un tekstrādē.

10 <https://www.globalframenet.org/>

- Banarescu, Laura, Bonial, Claire, Cai, Su, Georgescu, Madalina, Griffitt, Kira, Hermjakob, Ulf, Knight, Kevin, Koehn, Philipp, Palmer, Martha, Schneider, Nathan (2013). Abstract Meaning Representation for Sembanking. *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pp. 178–186.
- de Marneffe, Marie-Catherine, Manning, Christopher, Nivre, Joakim, Zeman, Daniel (2021). Universal Dependencies. *Computational Linguistics*, No. 47(2), pp. 255–308.
- Fillmore, Charles J., Baker, Collin F. (2010). A Frames Approach to Semantic Analysis. Heine, B., Narrog, H. (eds.). *Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, pp. 313–341.
- Fillmore, Charles J., Johnson, Christopher R., Petruck, Miriam R. (2003). Background to FrameNet. *International Journal of Lexicography*, No. 16(3), pp. 235–250.
- Fillmore, Charles, J. (1976). The need for a frame semantics in linguistics. Karlgren, H. (ed.). *Statistical Methods in Linguistics*. Stockholm: Scriptor, pp. 5–29.
- Gilardi, Luca, Baker, Colin (2018). Learning to Align across Languages: Toward Multilingual FrameNet. *International FrameNet Workshop 2018: Multilingual FrameNets and Constructicons*, pp. 13–22.
- Gruzitis, Normunds, Dargis, Roberts, Rituma, Laura, Nešpore-Berzkalne, Gunta, Saulite, Baiba (2020). Deriving a PropBank Corpus from Parallel FrameNet and UD Corpora. *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pp. 63–69.
- Gruzitis, Normunds, Pretkalnina, Lauma, Saulite, Baiba, Rituma, Laura, Nešpore-Berzkalne, Gunta, Znotins, Arturs, Paikens, Peteris (2018a). Creation of a Balanced State-of-the-Art Multilayer Corpus for NLU. *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*, pp. 4506–4513.
- Gruzitis, Normunds, Nešpore-Berzkalne, Gunta, Saulite, Baiba (2018b). Creation of Latvian FrameNet based on Universal Dependencies. *Proceedings of the International FrameNet Workshop (IFNW)*, pp. 23–27.
- Levāne-Petrova, Kristīne (2019). Līdzsvarotais mūsdienu latviešu valodas tekstu korpus, tā nozīme gramatikas pētījumos. *Valoda: nozīme un forma*, 10, 131.–146. lpp.
- Nešpore-Bērzkalne, Gunta, Saulite, Baiba, Grūzitis, Normunds (2018). Latvian FrameNet: Cross-Lingual Issues. *Human Language Technologies – The Baltic Perspective*, IOS Press, Vol. 307, pp. 96–103.
- Ohara, Kyoko (2020). Finding Corresponding Constructions in English and Japanese in a TED Talk Parallel Corpus using Frames-and-Constructions Analysis. *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pp. 8–12.
- Palmer, Martha, Gildea, Daniel, Kingsbury, Paul (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, No. 31(1), pp. 71–106.
- Rituma, Laura, Saulite, Baiba, Nešpore-Bērzkalne, Gunta (2019). Latviešu valodas sintaktiski marķētā korpusa gramatikas modelis. *Valoda: nozīme un forma*, 10, 200.–216. lpp.
- Ruppenhofer, Josef, Ellsworth, Michael, Petruck, Miriam R. L., Johnson, Christopher R., Baker, Collin F., Scheffczyk, Jan (2016). *FrameNet II: Extended Theory and Practice*. Available: <https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf> [accessed 05.11.2021.].
- Torrent, Tiago Timponi, Ellsworth, Michael, Baker, Collin, Matos, Ely Edison da Silva (2018). The Multilingual FrameNet shared annotation task: A preliminary report. *Proceedings of the International FrameNet Workshop 2018: Multilingual FrameNets and Constructicons (IFNW)*, pp. 62–68.

Latvian *FrameNet*

Baiba Saulīte, Gunta Nešpore-Bērzkalne,
Laura Rituma, Viesturs Jūlijs Lasmanis,
Normunds Grūzītis

Keywords: semantic annotation, frame semantics, semantic role,
lexical unit, verb, deverbal derivative

This paper presents a FrameNet-annotated text corpus for Latvian language. Latvian FrameNet is a part of a FullStack corpus – medium-sized general-purpose multi-layered corpus, anchored in cross-lingual state-of-the-art syntactic and semantic representations: Universal Dependencies (UD), FrameNet and PropBank, as well as Abstract Meaning Representation (AMR). The FullStack has been designed considering the variety and balance of the corpus in terms of genres, domains, and lexical units.

For annotating the FrameNet layer in this corpus, we use the latest frame inventory of Berkeley FrameNet, while the annotation itself is done on top of the underlying UD layer. Thus, the annotation of frames and frame elements is guided by the dependency structure of a sentence, instead of the phrase structure. We strictly follow a corpus-driven approach, meaning that lexical units (verbs and deverbal derivatives) in Latvian FrameNet are created only based on the annotated corpus examples.

Currently, 570 Berkeley FrameNet frames have been used for semantic annotation of the Latvian FrameNet corpus, 2900 lexical units (average 5.1 lexical items per frame) and almost 26 000 usage examples (average 8.9 per lexical unit) have been tagged. To make this data available for linguistic research, the website of FrameNet-LV corpus has been created.

Vārdnīcas digitālā struktūra

Sanda Rapa

Publikācija tapusi Valsts pētījuma programmas "Humanitāro zinātņu digitālie resursi" projektā "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" (Nr. VPP-IZM-DH-2020/1-0001).

Katrai valodai ir savs (īpašs, no citām valodām atšķirīgs) digitālais nospiedums, kas balstās uz valodu atšķirībām, specifiskajām iezīmēm. Tāds ir arī fleksīvajai latviešu valodai, un tāpēc tās lingvistiskajai uzkrāšanai un eksponēšanai digitālajā vidē nepieciešama īpaša pieeja: proti, ņemot vērā latviešu valodas specifiskās iezīmes, latviešu vārdam leksikogrāfiskā avotā (jo īpaši digitālajā vidē) varētu būt jānorāda plašāks gramatiskais apraksts, jāņem vērā tā locījumu paradigma un diakritiskās zīmes utt. Saukdams šīs specifiskās iezīmes par “daudzslāņaino vārdnīcas arhitektūru” (Spohr 2011), to apliecina arī Deniss Spors (*Dennis Spohr*). Turklāt viņš piebilst: “[...] [I]r nepieciešams tāds leksikogrāfisks rīks, kas spēj tikt galā ne tikai ar dažādiem lietotāju tipiem un situācijām, bet arī tāds, kas nodrošina nepieciešamos mehānismus vārdnīcas satura individualizēšanai – individuālā lietotāja skatu pielāgošanai.” Elektronisko vārdnīcu leksikogrāfiskā uzbūve sasaucas ar drukāto izdevumu leksikogrāfisko sistēmu, taču dažos aspektos būtiski atšķiras. Šo kopīgo pazīmju un atšķirību identificēšana un definēšana ir šā raksta mērķis.

Valodu resursus digitālajā vidē mēdz iedalīt trijās grupās:

- dati (piemēram, datubāzes, korpusi, tēzauri, vārdnīcas, analītiskie rīki utt.);
- rīki (piemēram, tulkotāji, runas atpazīšanas un pareizrakstības rīki, teksta redaktora rīki utt.);
- metadati, kas ļauj informāciju ievadīt, izgūt un uztvert (Bird, Simons 2003: 376 u. c.).

Runājot par šo valodas digitālās pasaules kopumu, parasti tiek lietots termins “valodas digitalizācija” (*language digitization*), ar to saprotot “valodas dokumentēšanu un aprakstīšanu, kur rezultāti ir paredzēti mašīnlasišanai, ne cilvēku lietošanai” (Abney 2011: 1). Šajā rakstā tiek aplūkota neliela, bet būtiska valodas digitalizācijas daļa – elektronisku vārdnīcu un datubāzu sastādīšana, uz ko tiek attiecināts termins “elektroniskā leksikogrāfija” jeb “e-leksikogrāfija” (Logan 1991; Fuertes-Olivera, Bergenholtz 2013; Granger, Paquot 2012 u. c.) un kas ir gan rīku, gan metadatu pamatā.

Elektroniskajai leksikogrāfijai Latvijā ir pavisam nesena vēsture, taču pārdesmit gados ir paveikts daudz, un liela daļa no pašlaik pieejamajiem digitālajiem lingvistiskajiem leksikogrāfiskajiem avotiem ir LU Latviešu valodas institūta (LU LaVI) un LU Matemātikas un informātikas institūta (LU MII) sadarbības rezultāts: satura autors ir LU LaVI, savukārt elektroniskā apstrāde ir LU MII darba rezultāts.

S tarp šādiem avotiem ir gan digitālās vārdnīcas, gan datubāzes, kas kopumā aptver pārsimts tūkstošus valodas vienību, kur katrai pievienots lingvistisks apraksts, piemēri un reģistrācijas dati. Balstoties uz šo pieredzi, rakstā tiek aplūkota latviešu valodas vienību digitalizācijas vēsture, specifika un nozīme no digitālo rīku satura veidotāju viedokļa.

Ieskats Latvijas e-leksikogrāfijā

Elektroniskās leksikogrāfijas pirmsākumi meklējami jau 20. gadsimta 50.–60. gados, kad sāka domāt par drukāto vārdnīcu pārveidošanu mašīnlasāmā formā – taču galvenokārt datu klasificēšanai, kārtošānai, kodēšanai, lai atvieglotu vārdnīcu rakstīšanu (Logan, 1991: 351), nevis otrādi – lai uzreiz veidotu elektroniskas vārdnīcas. Latvijā šo pieeju pazīst jau 70.–80. gados, taču vēl 1982. gadā, kad Igaunijā jau tapušas vairākas elektroniskas vārdnīcas, leksikogrāfe Liene Roze raksta: “Pagaidām elektroniskās iekārtas vēl nav piemērotas masveida lietošanai, jo ir dārgas un darbs ar tām ir sarežģīts.” (Roze 1982: 112) Latvijā elektroniskas vārdnīcas sāk veidot 90. gadu beigās Matemātikas un informātikas institūta Mākslīgā intelekta laboratorijā (galvenokārt Andreja Spektora vadībā). Pirmās elektroniskās vārdnīcas top kā drukāto vārdnīcu elektronisks atspulgs – informācija viens pret vienu tiek ievadīta elektroniskā vidē.

LU Latviešu valodas institūtā (LU LaVI) izstrādāto vārdnīcu ceļš uz digitālo vidi sākas 2001. gadā, kad no drukātas vārdnīcas par elektronisku avotu kļūst LU LaVI (agrāk Latvijas Zinātņu akadēmijas Valodas un literatūras institūta) pētnieku sastādītā “Latviešu literārās valodas vārdnīca”. Tagad tā digitālajā vidē iekļaujas kā statisks resurss – tā ir LU LaVI sagatavoto un apgādā “Zinātne” no 1972. līdz 1996. gadam izdoto astoņu sējumu digitāla publikācija, kas strukturēta mašīnlasāmā formātā (LLVV, <https://llvv.tezaurs.lv>). Vārdnīcas digitalizēšana ļāva piemēru un nozīmju sistēmu padarīt uzskatāmāku, novērst gan pamanītās kļūdas, gan drukātam izdevuma diktētos ierobežojumus (piemēram, saīsinājumus rakstīt pilnos vārdos). LLVV ir pagaidām apjomīgākais digitālais leksikogrāfiskais avots (tajā iekļauti vairāk nekā 64 000 šķirklju), taču tā netiek mainīta un pielāgota mūsdienu valodas situācijai, paliekot kā sava veida statisks, nemainīgs liecinājums par 20. gadsimta 80.–90. gadu normēto valodu. Šāda vārdnīca ar laiku kļūs par vēsturisku liecību, kas noderēs arī valodas vēstures pētījumiem, taču, neatspoguļojot valodas attīstību, tā neseko lingvistiskās datubāzes principiem, no kuriem galvenais ir dinamika – ne tikai digitālā atjauninājuma, bet arī satura pilnveidošanas aspektā (Hunston 2002).

Tāpēc jau kopš 1997. gada LU Latviešu valodas institūtā tiek izstrādāts arī dinamiskās “Mūsdienu latviešu valodas vārdnīcas” (MLVV, <https://mlvv.tezaurs.lv>) saturs (2022. gadā vairāk nekā 45 000 šķirklju). Tā ir balstīta uz pēdējo gadu desmitu latviešu valodas materiāla, kas iegūts ne tikai no drukātiem izdevumiem, bet arī no interneta un mūsdienu latviešu valodas korpusa. MLVV ir pilna tipa vārdnīca, kurā tiek reģistrēti visi mūsdienu latviešu valodā dzirdētie vārdi. Turklāt tajā arvien tiek iekļauti arvien jauni vārdi, nevisairoties no jebkuras stilistiskas nokrāsas, un agrāk reģistrēto vārdu nozīmes tiek rediģētas, papildinātas atkarībā no

novērotā mūsdienu lietojuma, tādējādi atspoguļojot semantiskās pārmaiņas latviešu valodas pamatleksikā. Atšķirībā no LLVV “Mūsdienu latviešu valodas vārdnīca” aptver dažādus latviešu valodas slāņus, iekļaujot arī izplatītākos neliterārās leksikas vārdus, barbarismus u. c.

Par pagaidām dinamisku var saukt arī “Latviešu valodas vēsturisko vārdnīcu” (LVVV, <https://tezaurs.lv/lvvv/>), kuras izstrāde uzreiz digitālā formā tika aizsākta 2004. gadā, pamatojoties uz 2003. gadā izveidoto latviešu valodas seno tekstu korpusu (<http://senie.korpus.lv/>), kas tiek papildināts vēl līdz šim brīdim). Kaut arī LVVV aptver tikai 16.–17. gadsimta latviešu rakstu avotu leksiku, tā joprojām tiek turpināta un tiek papildināta ar vārdiem un piemēriem no jaunatklātiem avotiem. Patlaban vārdnīcā ir nepilni divi tūkstoši šķirkļu, taču pēc laika šai vārdnīcai varētu būt statiska avota statuss – ja izdosies aptvert un digitalizēt visus 16.–17. gadsimta avotus, tā būs pabeigta un tās saturs netiks būtiski mainīts.

Par dinamiskiem uzskatāmi arī LU Latviešu valodas institūtā izstrādātie pūlpakalpes rīki: 2017. gadā atklātā “Tautas vietvārdu datubāze” (TVD, www.vietvardi.lv) un 2022. gadā radītā “Apvidvārdu datubāze” (AD, www.apvidvardi.lu.lv). Kaut arī to saturu nav radījuši valodnieki, tie tomēr uzskatāmi par leksikogrāfiskiem avotiem, jo tajos atbilstoši elektroniskās leksikogrāfijas principiem izstrādāti vienas valodas vienības atspoguļošanai nepieciešamie lauki. Tos var aizpildīt jebkurš interesents (arī bez valodnieciskām zināšanām), taču iesniegtie dati pēc pārbaudes izmantojami tālāk valodnieciskos pētījumos un leksikogrāfiskos avotos. Šo rīku rašanos stimulēja vispārējā digitalizācija un tās iespējamo ieguvumu apjaušana, sabiedrības digitālo prasmju uzlabošanās un leksikogrāfiskās digitalizācijas iespējas. Ja statiskās vārdnīcas daļēji turpina drukāto izdevumu principus, tad pūlpakalpes izstrāde liek meklēt arvien jaunus vārdu reģistrācijas un vizualizācijas aspektus. Tieši pūlpakalpes vietnes, kurās darbojas ne tikai valodnieki, bet galvenokārt jebkurš ieinteresēts interneta lietotājs, liek plānot pēc iespējas arvien ērtāku un pārskatāmāku vārda digitālo ekspozīciju.

Leksikogrāfiskās digitalizācijas principi (salīdzinājumā ar drukātiem izdevumiem)

Mūsdienu valodas digitalizācija lielā mērā balstās uz gadu desmitos un pat simtos izstrādātām gramatikas un leksikogrāfijas atziņām. Īpaši tas sakāms par elektroniskajām vārdnīcām – tajās tiek ievēroti galvenie vārdnīcas mikrostruktūras (elementu izkārtojuma šķirklī), makrostruktūras (šķirkļu izkārtojuma vārdnīcā) un megastruktūras (ārpus vārdnīcas korpusa publicētās papildinformācijas) principi (par vārdnīcas struktūrām sk. Jērāne 2014). Joprojām ir jāpiekrist Roberta Vešlera (*Robert Weschler*) un Krisa Pitsa (*Chris Pitts*) 2000. gadā izteiktajai atziņai, ka “elektroniskās vārdnīcas vēl ir pamatos drukātas vārdnīcas, kas ievietotas mikročipā” (Weschler, Pitts 2000) – tām ir noteiktas ātrākas funkcijas, taču principi ir nemainīgi. Daudz vairāk digitālajā pasaulē mainās vārdnīcu struktūru svarīgums un samērs. Piemēram, makrostruktūra, ņemot vērā to, ka elektroniskā vārdnīcā vārdu var sameklēt, to ierakstot meklēšanas logā, nevis meklējot vārdnīcas šķirkļus kādā noteiktas sistēmas sarakstā, nav tik būtiska

un negrozāma. Arī megastruktūras apjoms un izvietojums nav reglamentēts. Tāpat būtiski atšķiras sinhronisku (mūsdienu vai kāda samērā neliela laikposma) vārdnīcu un diahronisku (vēsturisku vai reģionālu) vārdnīcu digitalizācijas principi. Sinhroniskajām vārdnīcām ir plašāks horizontālais (sinhroniskais) lauks: piemēram, LLVV un daļēji arī MLVV mikrostruktūru paplašina gramatiskā informācija, kas sniedz ziņas par vārda pamatformu (sk. 1. attēlu). Turpretim diahroniskajām vārdnīcām apjomīgs ir vertikālais (diahroniskais) lauks: piemēram, LVVV līdztekus mikrostruktūras pamatelementiem sniegtas ziņas par vārda attīstību laikā un vārda rakstību dažādos avotos (sk. 2. attēlu), bet TVD un AD par svarīgu papildkomponentu izvēlēta lokalizācija (vārda ģeoreferencēšana) (sk. 3. attēlu).

Mikrostruktūra

Elektroniskas vārdnīcas svarīgākais struktūras elements ir tās mikrostruktūra. Tāpat kā tradicionālajā leksikogrāfijā, arī leksikogrāfiskajā digitalizācijā to veido divi pamatelementi: vārds un tā nozīme. Pamatojoties uz vārda nozīmi un raksturu, tiek veidota elektroniskas vārdnīcas šķirklja arhitektūra. Tai jābūt paredzamai, striktai un vienkāršai, ar strukturētu sintaksi (Abel 2012: 84), lai lietotājs tajā varētu viegli orientēties. Tāda ir arī LU LaVI elektronisko vārdnīcu struktūra – to pamatos veido vārds un tā nozīme, kas tiek ilustrēta ar piemēriem (vārda un nozīmes lauks ir obligāti aizpildāmās vienības, bez kurām leksisko vienību vārdnīcā nevar saglabāt). Vārds un nozīme, ko ilustrē piemēri, ir izvēlēti par galvenajiem elementiem arī pūlpakalpes rīkos – valodas datu uzkrāšanas vietnēs.

Vārds ir pamatelements valodas sistēmā, un vārdam jābūt galvenajam gan šķirklja struktūras izkārtojumā, gan šķirklju sarakstā. Tas tiek ievērots arī digitālajā leksikogrāfijā – vārds ir centrālais elements. Atšķirībā no tradicionālās leksikogrāfijas, kur vārds tiek tehniski izcelts (parasti treknināts) šķirklja struktūrā, digitālajā leksikogrāfijā vārds tiek nošķirts no šķirklja struktūras – tas tiek eksponēts kā šķirklja virsraksts. Elektroniskā leksikogrāfija atšķirībā no tradicionālās ļauj vārdu nemanāmi saskaldīt vēl sīkākās daļās – morfēmās un fonēmās, kuras var anotēt un tālāk izmantot klasificēšanā, strukturēšanā, analizē, meklēšanā. Tas ir nenovērtējams ieguvums valodniekiem, kam agrāk šo visu funkciju nodrošināšanai vajadzēja izstrādāt atsevišķus leksikogrāfiskus avotus (piemēram, biežuma, inversās, sinonīmu utt. vārdnīcas).

Vārda lauka struktūra ir atšķirīga sinhroniskajā un diahroniskajā digitalizācijā. Sinhroniskajā digitalizācijā (mūsdienu vai kāda samērā neliela laikposma valodas digitālā atspoguļošanā) par šķirklja vārdu tiek norādīta viena vārdforma, kurpretim diahroniskajā digitalizācijā (vēsturiskā valodas attīstības vai reģionālistikas atspoguļošanā) šķirklja galvu var veidot vairāku vārdformu kopa (sal., piemēram, šķirkli *ābols* sinhroniskajā MLVV ar vienu pamatformu un diahroniskajā LVVV, kur šķirklja galvu veido 36 vārdformas, kas reģistrētas dažādos avotos un dažādos laikposmos).

Nozīmi daži pētnieki uzskata par svarīgāko elektroniskas vārdnīcas elementu, jo tas ir galvenais, ko elektroniskā vārdnīcā meklē lietotājs (Lew 2010: 291). Tāpēc tai vārdnīcas sastādīšanā

otrāds

otrāds [ùotrāds] -ais īpašības vārds

otrāda -ā īpašības vārds

otrādi apstākļa vārds

Tāds, kam ir cits, parasti pretējs, veids, stāvoklis; tāds, kas izpaužas citā, parasti pretējā, veidā.

✓ Piemēri *Izdarīt otrādi.*

naba

naba (4) s. nabba (4)

Nabel/ Nabba. Manc1638_L, 128A₂₄.

Nabba der Nabel. Fuer1650_70_1ms, 163₂.

Nabba, der Nabel. Fuer1650_70_2ms, 237₂.

naba

Nabel/ Nabba. Manc1638_L, 128A₂₄.

Nabba, der Nabel. Fuer1650_70_2ms, 237₂.

◆ : pr. *nabis* 'naba; ratu rumba'

[2016-08]

atsprēkliņ

otrādi, atmuguriski

Piemērs

Sīpolus netišām sastādīju atsprēkliņ

Piezīmes

Nav minēts

Vārda fiksēšanas vieta

Lejasciema pagasts ♡

↑ 1. attēls. “Mūsdienu latviešu valodas vārdnīcas” šķirklis.

↔ 2. attēls. “Latviešu valodas vēsturiskās vārdnīcas” šķirklis.

↓ 3. attēls. Apvidvārdu datubāzes šķirklis.

jāpievērš vislielākā vērība. Vārda nozīmes atspoguļošana elektroniskā vārdnīcā būtiski neatšķiras no tradicionālās leksikogrāfijas: tai tāpat jābūt precīzai, pietiekami īsai un noteiktai, pat neraugoties uz to, ka elektroniskā vidē nav nepieciešami tik lieli telpas un teksta ierobežojumi kā drukātos izdevumos. Nozīmi tāpat kā drukātos izdevumos var atspoguļot gan ar definīciju, gan ar ekvivalentu (sinonīmu), gan ar piemēru, taču elektroniskā leksikogrāfija ļauj bez šīm

tradicionālajām nozīmes ekspozīcijām izmantot arī citus risinājumus, piemēram, audiomateriālus, videomateriālus un neverbālus elementus (piemēram, attēlus, animācijas, neverbālas skaņas). Audiomateriāli ļauj daudz precīzāk uztvert valodas būtību, atspoguļo reālo valodu, savukārt neverbālie elementi palīdz izskaidrot grūti aprakstāmas parādības un priekšmetus (Lew 2010: 297ff.). Īpaši tas svarīgi reģionālajos valodas paveidos un izloksnēs, kur vārdu precīzā intonācijā var izrunāt tikai vietējais reģiona iedzīvotājs. Rēķinoties ar to, LU LaVI izstrādātajā “Apvīdvrādu datubāzē” ir iekļauta iespēja vārdu arī ierunāt.

Digitalizācija ļauj viegli risināt dažus drukātā izdevumā grūti risināmus leksikogrāfiskus trūkumus nozīmju skaidrošanā. Piemēram, cirkulārās nozīmes (proti, nozīmes, kurās vārds skaidrots ar šķirkļa vārda daļām, piemēram, putekļsūcējs ‘putekļu sūcējs’): aktīvu mijnorāžu izmantošana un vārdu daļu skaidrojuma ātra aplūkošana var būtiski atvieglot gan leksikogrāfa darbu, gan vārdnīcas lasītāja meklējumus. Tomēr no leksikogrāfiskā aspekta cirkulārās nozīmes arī leksikogrāfiskajā digitalizācijā nav vēlamas.

Nozīmes definēšanai gan drukātā, gan digitālā izdevumā vajadzētu izmantot proformas (Granger, Paquot 2012: 23): katrai semantiskajai kategorijai specifiskas definīcijas satura un struktūras veidnes, kurās viegli ievietot katra atšķirīgā vārda nozīmi vienotā formā, tādējādi uzlabojot vārdnīcas kopējo sistēmu. Elektroniskajā leksikogrāfijā proformas ir īpaši viegli pielāgojamas un izmantojamas gan šķirkļu veidošanā, gan to rediģēšanā, jo viegli atlasīt šādām veidnēm attiecīgās leksiskās ligzdas un pārskatīt nesakritības (Atkins, Rundell 2008: 123–124). Šādas proformas apzināti vai neapzināti ir atrodamas jebkurā vārdnīcā. Piemēram, LLVV dzīvnieka definēšanai lielākoties izmantota zoonīmiskā taksonomija apvienojumā ar tā uzvedības pazīmēm vai izmantošanas nolūku (piemēram, *kaķis* ‘neliels plēsēju kārtas mājdzīvnieks, kas medī peles, žurkas, putnus’; *suns* ‘suņu dzimtas mājdzīvnieks, kam ir saimnieciska vai dekoratīva nozīme’); Oksfordas vārdnīcā definīcijas veidnē iekļauts dzīvnieka lielums, apspalvojuma krāsa un taksonomija (turpat). Taču latviešu elektroniskajās vārdnīcās proformas ne vienmēr izmantotas – tas redzams citos šķirkļos, kur dzīvnieku definīcijā skaidrojuma struktūra ir citāda – nav pieminēta dzīvnieka zoonīmiskā taksonomija: piemēram, *zaķis* ‘vidēji liels savvaļas dzīvnieks ar garām ausīm, pagarinātām pakaklējām un īsu asti’ (LLVV, MLVV).

Diahroniskajās vārdnīcās obligāti mikrostruktūras elementi ir hronotopi – laika un telpas marķieri (ne velti diahronisko lingvistisko avotu joma dažreiz tiek dēvēta par telpisko humanitāro zinātni (*spatial humanities*)) (Won et al. 2018). Tā vietvārdu un apvīdvrādu datubāzē par obligātu lauku noteikta lokalizācija, savukārt LVVV vēlams visu formu hronoloģisks atspoguļojums.

Piemēri, kaut arī valodas skaidrojošajās vārdnīcās nav obligāti aizpildāmi lauki, LU LaVI izstrādātajos elektroniskajos leksikogrāfiskajos avotos ir pievienoti visiem vārdiem, jo vārdnīcas lietotājam, kurš gan LLVV, gan MLVV uzskata par normējošiem avotiem, tie dažkārt ir pat izšķiroši un palīdz uztvert vēlamā vai nevēlamā lietojuma nianses. LU LaVI vārdnīcās vienmēr pievienota arī vārda gramatiskā informācija (vārdšķira, vārda dzimte, stilistiskā norāde, dažādi ierobežojoši vai izņēmuma gadījumi).

Makrostruktūra un megastruktūra

Vārdnīcas makrostruktūras un megastruktūras loma elektroniskajā vidē ir būtiski pavājinājusies. Pavisam nenožīmīga ir kļuvusi makrostruktūra – šķirklju izkārtojums vārdnīcā. Elektroniskās leksikogrāfijas atbrīvošanos no alfabēta kārtības kā lielu ieguvumu un atslogu salīdzinājumā ar drukātajiem izdevumiem min visi e-leksikogrāfijas pētnieki (Granger, Paquot 2012; Lew 2010 u. c.). Īpaši to novērtēja un slavēja pirmo elektronisko vārdnīcu autori un lietotāji. Taču tagad, pēc vairākām aktīvām elektroniskās leksikogrāfijas desmitgadēm, lietotāji un vārdnīcu autori tomēr apjautuši alfabētiskas kārtības priekšrocības. Iespējams, šādu nepieciešamību radījusi valodas iekšējā paplašināšanās (daudzu atvasinājumu darināšana). Alfabētiskais saraksts ļauj viegli pārskatīt vārda apkaimei (visus vārda saknes atvasinājumus un līdzīgos vārdus). Nevienā no LU LaVI izstrādātajām elektroniskajām vārdnīcām šāda rādītāja nav. Lai kaut daļēji apmierinātu prasības pēc sistēmiska saraksta, MLVV tika izveidota sleja, kas rāda meklētā vārda apkaimei (piemēram, meklējot vārdu *skola*, vārdnīca dos ziņu, ka ir arī vārdi *skolasbērns*, *skolasnauda*, *skolēns*, *skolmeistars*, *skolniecīks*). Atvasinājumu reizēm ir tik daudz, ka tos visus nespēj parādīt pat apkaimes logs (tā MLVV *skolas* logā neredzēsīm *skolnieks*, *skolot*, *skoloties* u. c.), tāpēc gan no valodnieku, gan lietotāju puses izskanējuši lūgumi pēc kopēja alfabētiska saraksta vai cita kopēja sistēmiska atspoguļojuma. Taču elektroniskajā leksikogrāfijā tas joprojām nav atrisināts – liela apjoma datiem tas, kas tradicionālajā leksikogrāfijā tik pierasts, elektroniskos avotos var kļūt par pārbaudījumu.

Arī megastruktūras lomas pavājināšanās digitālajās humanitārajās zinātnēs ir pamanīta (Burdick et al. 2012). To redz arī jebkurš lietotājs. Atšķirībā no drukāta izdevuma, kur līdz vārdnīcas tekstam var nokļūt caur vārdnīcas nosaukumu, autoru sarakstu un vārdnīcas struktūras aprakstu, elektroniskajā leksikogrāfijā vienīgais redzamais elements ir saīsināts vārdnīcas nosaukums (visbiežāk abreviātūra) un vārdnīcas elektroniskās vietnes nosaukums. Gandrīz nemanāms kļūst vārdnīcas autors vai sastādītājs. Digitālajās humanitārajās zinātnēs autors vairs nav autonoma persona, jo vārdnīcas sastādīšana nebeidzas ar šķirklju sarakstīšanu. Tas, kas agrāk tika uzticēts izdevējiem (redigēšana, izkārtojums, izplatīšana, pielāgošana lietotāju vajadzībām), tagad ir digitālo humanitāro zinātņu priekšmets (Burdick et al. 2012: 83). Tāpēc arī vārdnīcas struktūras aprakstam digitālajā pasaulē tiek pievienots vārdnīcas elektroniskās ekspozīcijas skaidrojums.

Elektroniskās leksikogrāfijas ieguvumi

Kaut arī struktūru līmenī elektroniskā leksikogrāfija joprojām turpina tradicionālās leksikogrāfijas principus, tās ieguvumi gan no vārdnīcu lietotāju, gan sastādītāju puses tiek arvien vairāk novērtēti un arvien tiek paplašināti un pielāgoti. Apvienojot Silvianas Greindžeras (*Sylviane Granger*) (Granger, Paquot 2012: 2) atzinumus un LU LaVI pieredzi elektronisku vārdnīcu sastādīšanā, var izšķirt šādas elektroniskās leksikogrāfijas priekšrocības:

1. korpusa integrācija;
2. datu apjoma palielināšanās
un kvalitātes uzlabošanās;
3. efektīva piekļuve;
4. ērta pielāgošana;
5. hibrīdizācija;
6. vienota sistēma;
7. lietotāja līdzdalība.

Korpusu integrācija (valodas korpusu datu iepludināšana elektronisko vārdnīcu šķirklju izstrādē, nozīmes precizēšanā un piemēru bagātināšanā) kļūst arvien nozīmīgāka, un, palielinoties un uzlabojoties valodas resursiem, tā ir arvien efektīvāka. Elektroniskās leksikogrāfijas pētnieki uzskata, ka nākotnes vārdnīca pilnībā balstīsies uz kontekstu-alizāciju (vārdu nozīmes tiks izstrādātas, izvērtējot plašu to lietojumu kontekstu). Šķiet, nevienu mūsdienu vārdnīcu vairs nevarētu izveidot bez ātras un jaudīgas apjomīgu datu apstrādes, kas ļauj precīzi definēt vārdu nozīmes, savukārt lietotājiem dod iespēju iepazīt kontekstu. Šādas savstarpējas korpusa un vārdnīcas piemērs ir LVVV, kuras izstrādē tiek izmantots seno tekstu korpus "Senie" un no kuras šķirklja var nokļūt tieši pie izmantotās vārdformas konkrētā lietojuma vietā. Korpusu integrācija ļauj atrast un pievienot arvien vairāk piemēru un noteikt vārda lietojuma biežumu. Tas nozīmē, ka vienlaikus ar elektroniskas vārdnīcas izveidi jārūpējas arī par korpusa nodrošināšanu un paplašināšanu, jo, lai vārda nozīme būtu pēc iespējas precīzāka, tās izstrādes pamatā jābūt pēc iespējas lielākam korpusam.

Korpusu integrācija, valodas resursu uzkrāšana, uzglabāšana un apkopošana ir radījusi lielas iespējas arī leksikogrāfa darbā – tā rīcībā ir milzīgi valodu resursi, ko var izmantot leksikogrāfisko avotu izstrādes pamatā. Datu apjoma paplašināšanās paaugstina arī leksikogrāfisko avotu kvalitāti. Izmantojot korpusu un citu valodas resursu datus, valodnieks var precīzāk definēt vārda nozīmi un nozīmes nianšes, ja redz to vairākos piemēros vienkopus, turklāt vienlaikus var izvēlēties raksturīgākos piemērus, ko rādīt vārdnīcas mikrostrukturā. Salīdzinājumā ar agrākiem gadiem, kad piemēri tika meklēti drukātos izdevumos un katalogizēti kartotēkās, šāds datu ieguves ceļš daudzāk atvieglo valodnieku pētījumus un arī lietotāju meklējumus. Taču patlaban, kad datu apjoms arvien pieaug, jāatceras, ka pārāk daudz datu var apgrūtināt tā analizētāju, tāpēc digitālajām humanitārajām zinātnēm nepārtraukti jārūpējas par datu apstrādi un to digitālo ekspozīciju.

Vārdnīcu digitalizēšana gandrīz nepārtraukti paplašina lietotāja iespējas un ērtu piekļuvi – lasītājam vairs nav jāšķirsa drukāti izdevumi un jāpārzina alfabēts, mījnoražu atrašanai vien jāatver saite, kontekstu var iepazīt ne tikai sintagmu, teikumu vai rindkopu, bet pat diskursa apjomā. Elektroniskas vārdnīcas lietotājam vairs nav jāmeklē saīsinājumu saraksts, jo digitālo lauku telpa nav tik ļoti ierobežota kā drukātam izdevumam. Vārdnīcas vairāk nekā jebkad agrāk tiek pielāgotas lietotāja vajadzībām, un šo vajadzību apmierināšana tiek uzskatīta par nākotnes vārdnīcu mērķi. Lietotāju vēlmes konstatētas jau elektroniskās leksikogrāfijas aizsākumos: ir noskaidrots, ka vārdnīcas lasītājs galvenokārt

interesējas, vai šāds vārds vispār eksistē, kā to raksta un izrunā un kāda ir tā nozīme vai nozīmes nianse (Wallraff 2009). Par elektroniskas vārdnīcas pielāgošanu jādoma jau tās izveides sākumposmā, kad datu ievadišanas posms vēl nav sācies. Arvien vairāk elektronisko leksikogrāfu izvēlas radīt pēc iespējas vairāk valodas digitālās ekspozīcijas formu, lai vēlāk vārdnīcu varētu pielāgot vai pārveidot: tiek anotētas un kodētas valodas vienību daļas, sākot no plašākiem tekstiem līdz pat fonēmām; tiek pievienota izruna audioierakstā vai starptautiskā fonētiskā sistēmā; tiek veidotas pilnas vārda gramatiskās paradigmas utt. Šāda elektronisko vārdnīcu arhitektūras izstrāde tālāk var rosināt valodas resursu hibridizāciju, kurā, kā plāno digitālo humanitāro zinātņu pētnieki, robežas starp vārdnīcu, rīku un korpusu palēnām vājināsies vai pat izzudīs (Granger, Paquot 2012: 5) un tie visi taps par vienotu valodas resursu, proti, vienā resursā varēs atrast visu par konkrētu valodas vienību. Latvijas elektroniskajā leksikogrāfijā dažādu valodas resursu robežas vēl skaidri iezīmējas, taču, piemēram, LVVV ciešā sasaiste ar korpusu un vietnes *tezaurus.lv* datu akumulēšana no visām elektroniskajām vārdnīcām jau liecina par daļēju hibridizāciju arī latviešu humanitārajās zinātnēs.

Hibridizāciju arvien var uzlabot digitālo resursu sistēmiskums. Vienota sistēma un sistēmiskums ir visu digitālo humanitāro zinātņu pamatā, un tas ir galvenais digitālo humanitāro zinātņu devums humanitārajām zinātnēm, arī elektroniskajai leksikogrāfijai. Digitālā vide prasa proformas, ko pielāgot lieliem valodas resursiem, un īpaši elektroniskajā leksikogrāfijā tāas tiek izstrādāti un izmantotas ikviena šķirķļa izveidē. Līdz ar to pārdomāti izstrādāta digitālā vide atšķirībā no drukāta izdevuma nekad neļaus pazaudēt kādu no sistēmas pamatelementiem, neļaus aizmirst nevienu obligāto informācijas vienības aspektu (piemēram, TVD vietvārdu nevar ievadīt, ja nav norādīta precīza tā atrašanās vieta administratīvā teritorijā, AD apvidvārdu nevar pievienot, ja nav norādīta tā nozīme un/vai lokalizācija).

Šādas obligāto elementu sistēmas ir ļoti svarīgas pūlpakalpes rīkos, kur leksikogrāfiskos laukus var aizpildīt jebkurš lietotājs. Latvijā lietotāju līdzdalība leksikogrāfisku avotu izstrādē tiek īpaši veicināta. Digitālais laikmets nevar pastāvēt bez nemītīgas sasaistes starp lietotāju un izstrādātāju. Elektronisku vārdnīcu izstrādātājiem nemītīgi jāinteresējas par lietotāju vēlmēm un palaikam jāatļauj lingvistisko leksikogrāfiju veidot arī pašiem. Sabiedrība tiek iesaistīta ne tikai elektronisko vārdnīcu šķirķļu rediģēšanā (piemēram, MLVV un LLVV ir iespējams ziņot par kādu neprecizitāti valodnieku izstrādāto šķirķļu elementos), bet arī vārdnīcas vai datubāzes satura veidošanā. LU Latviešu valodas institūtā ir izstrādāti divi pūlpakalpes rīki – leksikogrāfiski avoti, kas balstās tikai uz lietotāju iesūtītās informācijas datiem (“Tautas vietvārdu datubāze” un “Apvidvārdu datubāze”).

Šādi elektroniski leksikogrāfiski rīki sniedz ne tikai citādi grūti iegūstamus valodas datus, bet arī rāda, ka sabiedrība aptver savas valodas faktu vērtību un apzinās, ka tie jāsgādā un jādara zināmi citiem.

Digitalizācija nodrošina ne tikai datu pieejamību, pārskatāmību, bet arī dabas saudzēšanu, ko uzsver ikviens digitālā laikmeta pētnieks.

Secinājumi

LU LaVI pētnieku darbs veido trīs lielāko latviešu valodas elektronisko vārdnīcu pamatu: LLVV, MLVV, LVVV. Tās aptver latviešu valodas pamatkorpusu gan diahroniski, gan sinhroniski: LVVV tiecas kļūt par seno rakstu avotu tēzauru, LLVV aptver literārās valodas vārdu kopumu, savukārt MLVV pamazām tiek publicēti visi mūsdienās lietotie un arvien no jauna radušies vārdi un vārdu savienojumi. Pēdējos gados LU LaVI izstrādātie elektroniskās leksikogrāfijas avoti tiecas aptvert arī Latvijas reģionos lietoto valodu: AD un TVD mērķis ir uzkrāt literārajā valodā neiekļautos, izloksnēs lietotos vārdus.

Iepazīstot digitālo humanitāro zinātņu pētnieku atziņas, pieredzi un paraugus, secināms, ka latviešu elektroniskajā leksikogrāfijā vēl nepieciešami uzlabojumi. Pirmkārt, latviešu elektronisko vārdnīcu šķirkļu izstrādē trūkst proformu (šķirkļu veidnes), kas atvieglotu gan vārdnīcu sastādītāju darbu, atgādinot par obligātām formulējumu struktūras un semantikas sadaļām, gan ļautu lietotājam vieglāk uztvert vārdu nozīmes un gūt pilnvērtīgāku vārda nozīmes skaidrojumu. Otrkārt, lai gan meklēšanas iespējas elektroniskajās vārdnīcās ir gandrīz neierobežotas, visām vārdnīcām būtu nepieciešams leksēmu alfabētisks rādītājs, kas lietotājam ļautu atrast vārdu, kura formu viņš precīzi nezina, turklāt radītu pilnīgāku priekšstatu par valodas sistēmu. Treškārt, lai atspoguļotu reālo valodu un valodas būtību, elektroniskajām vārdnīcām būtu nepieciešams pievienot audiomateriālus (vārda reālo izrunu) – apjaušot to nozīmību, iespēja ierunāt vārdu ir iekļauta “Apvidvārdu datubāzē”. Ceturtkārt, nepieciešams arvien paplašināt valodu korpusus un tos integrēt elektronisko vārdnīcu izstrādē, lai nodrošinātu precīzāku šķirkļu izstrādi.

- Abel, Andrea (2012). Dictionary writing systems and beyond. Granger, Sylviane; Paquot, Magali (eds.). *Electronic Lexicography*. Oxford: Oxford University Press, pp. 83–106.
- Abney, Steven (2011). *Language Digitization*. Write-up of a talk given at the University of Michigan. Available: <http://www.vinartus.net/spa/p102-v2.pdf> [accessed 01.11.2011.].
- AD – *Apvidvārdu datubāze*. Pieejams: <https://apvidvardi.lu.lv> [skatīts 17.05.2022.].
- Atkins, Beryl T. Sue, Rundell, Michael (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Beal, John, Corrigan, Karen, Moisl, Hermann (eds.) (2007). *Creating and digitizing language corpora*, Vol. 1. *Synchronic databases*. Hampshire&New York: Palgrave Macmillan.
- Bird, Steven, Simons, Gary (2003). Extending Dublin Core Metadata to Support the Description and Discovery of Language Resources. *Computers and the Humanities*, 37 (4), pp. 375–388.
- Burdick, Anne, Drucker, Johanna, Lunenfeld, Peter, Presner, Todd, Schnapp, Jeffrey. *Digital Humanities*. Cambridge, London: MIT Press, 2012.
- Fuertes-Olivera, Pedro A., Bergenholz, Henning (2013). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London: Continuum.
- Jērāne, Santa (2014). Vietvārdu vārdnīcu megastruktūra. *Linguistica Lettica*, Nr. 22. Rīga: LU Latviešu valodas institūts, 185.–204. lpp.
- Jērāne, Santa (2021). Cirkulāri skaidrojumi “Mūsdienu latviešu valodas vārdnīcā”: dažas teorētiskas un praktiskas problēmas. *Linguistica Lettica*, Nr. 28. Rīga: LU Latviešu valodas institūts, 164.–173. lpp.
- Lew, Robert (2010). Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos*, No. 20, pp. 290–306.
- Logan, Harry M. (1991). Electronic Lexicography. *Computers and the Humanities*, No. 25(6), pp. 351–361.
- LLVV = *Latviešu literārās valodas vārdnīca* [elektroniska vārdnīca]. Pieejams: <https://llvv.tezaurs.lv/> [skatīts 17.05.2022.].
- LVVV = *Latviešu valodas vēsturiskā vārdnīca* [elektroniska vārdnīca]. Pieejams: <https://tezaurs.lv/lvvv/> [skatīts 17.05.2022.].
- MLVV = *Mūsdienu latviešu valodas vārdnīca* [elektroniska vārdnīca]. Pieejams: <https://mlvv.tezaurs.lv/> [skatīts 17.05.2022.].
- Roze, Liene (1982). *Pasaule vārdnīcas skatījumā*. Rīga: Zinātne.
- Granger, Sylviane; Paquot, Magali (2012). *Electronic Lexicography*. Oxford: Oxford University Press.
- Spohr, Dennis (2011). A multi-layer architecture for ‘pluri-monofunctional’ dictionaries. *e-Lexicography – The Internet, Digital Initiatives and Lexicography*. London and New York: Continuum, pp. 103–120.
- TVD – *Tautas vietvārdu datubāze*. Pieejams: www.vietvardi.lv [skatīts 17.05.2022.].
- Wallraff, Barbara (2009). The uncertain future of dictionaries. *The Atlantic*, 12 January. Available: http://barbarawallraff.theatlantic.com/archives/2009/01/the-uncertain_future_of_dictionaries.php [accessed 12.12.2021.].
- Won, Miguel. Murrieta-Flores, Patricia, Martins, Bruno (2018). Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora. *Frontiers in Digital Humanities*, Vol. 5:2. Available: <https://www.frontiersin.org/articles/410.3389/fdigh.2018.00002/full> [accessed 30.10.2021.].
- Weschler, Robert, Pitts, Chris (2000). An experiment using electronic dictionaries with EFL students. *The Internet TESL Journal*, No. 6. Available: www.iteslj.org/Articles/Weschler-ElectroDict.html [accessed 28.10.2021.].

The Digital Structure of Dictionary

Sanda Rapa

Keywords: e-lexicography, electronic dictionary, microstructure of e-dictionary, macrostructure of e-dictionary, history of Latvian e-lexicography

E-lexicography has not got long history in Latvia: the first electronic dictionaries appeared in the late 1990s. However, many achievements can be traced over the past twenty years: electronic dictionaries of Early Latvian (<https://tezaurs.lv/lvvv>), Standard Latvian (<https://llvv.tezaurs.lv/>) and Modern Latvian (<https://mlvv.tezaurs.lv/>) have been published. Also, databases for crowdsourcing of Latvian place names (www.vietvardi.lv) and regional words (<https://apvidvardi.lu.lv>) have been created. These resources comprise more than 100 thousand words each of which is accompanied by a linguistic description, examples and data of registration. All of the e-lexicographic sources in Latvian are elaborated in cooperation between the Latvian Language Institute and the Institute of Mathematics and Computer Sciences of the University of Latvia. Based on this experience, the article deals with the history, specificity and importance of the digitization of Latvian language from the point of view of content creators of digital tools.

The article is structured into three subchapters. The first part provides an insight into the history of Latvian e-lexicography, the second part analyses the principles of compiling the electronic lexicographic sources in Latvia (including the analysis of microstructure and macrostructure of published electronic dictionaries). The third part deals with the pros and cons of Latvian lexicographic sources as opposed to printed sources (the analysis of corpus integration, data sufficiency and unification, efficiency of access, customization, hybridization, user input is provided). The aim of the article is to outline priorities for the future of Latvian electronic lexicography.

The analysis of Latvian electronic dictionaries reveals four areas where Latvian e-lexicography still requires an improvement. Firstly, there is a lack of proformas in all electronic dictionaries of Latvian language which could help to unify the structure of definition and meaning of the entries. Secondly, not all Latvian electronic dictionaries comprise an alphabetical list of lemmas which could help to find an entry without knowing the spelling of the word. Thirdly, audio pronunciation for each entry is necessary in order to provide real (standardized) pronunciation. Fourthly, corpus integration is essential to produce rich lexical entries and to define the meaning of lemmas more precisely.

CLARIN valodas resursu un rīku pētniecības infrastruktūra humanitārajām un sociālajām zinātnēm

Inguna Skadiņa, Ilze Auziņa,
Roberts Darģis, Arnis Voitekāns

CLARIN Latvija nacionālā mezgļpunkta izveidi un uzturēšanu atbalsta Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi” projekts “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001) un Eiropas Reģionālās attīstības fonda projekts “Latvijas Universitāte un institūti Eiropas pētniecības telpā – ekselence, aktivitāte, mobilitāte, kapacitāte” (Nr. 1.1.1.5/18/I/016).

Ievads

Lai arī digitālu valodas resursu un rīku izveide uzsākta drīz pēc pirmo datoru parādīšanās, nepieciešamība veidot digitālas valodas resursu krātuves, kas būtu pieejamas ikvienam pētniekam neatkarīgi no to atrašanās vietas, aktualizējās gadsimtu mijā līdz ar straujo interneta uzplaukumu un digitālos avotos balstītas pētniecības attīstību.

Iemesli, kāpēc nepieciešams veidot valodas resursu krātuves un pētniecības infrastruktūras, ir vairāki. No vienas puses, digitāla valodas resursu un rīku pētniecības infrastruktūra ir pamatnosacījums digitālo humanitāro zinātņu pētījumiem. No otras puses, uzkrājoties arvien vairāk digitālajiem valodas resursiem, rodas nepieciešamība šos pētniecības datus padarīt pieejamus plašam pētnieku lokam un nodrošināt to saglabāšanu ilgtermiņā (t. i., pasargāt no pazušanas pēc kāda konkrēta projekta vai iniciatīvas beigām). Visbeidzot, jāpiemin otrais Gūtenberga efekts (valodas, kas nebūs vai būs nepietiekami pārstāvētas digitālajā vidē, ir apdraudētas un var ar laiku izzust), kas kā iespējams apdraudējums valodas pastāvēšanai tika aktualizēts jau 20. gadsimta 90. gados.

Apzinoties nepieciešamību veidot valodas resursu un rīku krātuves, novērst digitālo valodas resursu sadrumstalotību un veicināt to saglabāšanu ilgtermiņā, 2006. gadā vairāku valstu pētnieku grupa nāca klajā ar iniciatīvu veidot Vienotu valodas resursu un tehnoloģiju infrastruktūru (CLARIN – *Common Language Resources and Technology Infrastructure*; Vāradi et al. 2008). Šī iniciatīva aizrāva daudzus Eiropas zinātniekus – apvienojot spēkus, tika veidoti valodas resursu un rīku pārskati, izstrādāta tehniskā specifikācija Eiropas līmeņa repozitoriju sistēmas izveidei, apzināti un pētīti licencēšanas modeļi, kā arī definēts Eiropas pētnieciskās infrastruktūras juridiskais pamats un pārvaldības modeļi. 2012. gada 29. februārī, pēc tam, kad Eiropas Savienība (ES) bija izveidojusi Eiropas pētniecības infrastruktūru (ERIC) juridisko pamatu, deviņas ES valstis nodibināja CLARIN ERIC. Latvija CLARIN ERIC iestājās 2016. gada vasarā (Skadiņa et al. 2020).

Pašlaik CLARIN apvieno 25 valstis: 22 valstis ir pilntiesīgas ERIC locekles, bet trīs valstīm ir novērotājas statuss. CLARIN mērķi – vienotu piekļuvi valodas resursiem un rīkiem – infrastruktūra realizē ar vairāk nekā 60 centriem, kuri apvienoti virtuālā tīklā un ļauj zinātniekiem, izmantojot vienotu pierakstīšanos, piekļūt valodas resursiem un rīkiem neatkarīgi no viņu atrašanās vietas. CLARIN virtuālajā

valodas resursu krātuvē (VLO – *Virtual Language Observatory*)¹ pašlaik reģistrēti vairāk nekā 1,25 miljoni valodas resursu un rīku.

CLARIN piedāvā ne tikai ilgtermiņa risinājumus un tehnoloģijas digitālo valodas datu un rīku izvietošanai, bet arī konsultē un atbalsta pētniekus, kas veic valodas datus balstītus pētījumus.

Šī raksta mērķis ir iepazīstināt digitālo humanitāro zinātņu pētniekus ar CLARIN ERIC un CLARIN Latvijas mezglpunkta (CLARIN-LV) piedāvātajām iespējām, sniedzot ieskatu gan pētniecības infrastruktūrā apkopotajos valodas resursos Latvijā un Eiropā, gan dažādu valodas resursu un rīku kopās (*Resource Families*²). Rakstā arī sniegti praktiski ieteikumi valodas resursu deponēšanai un licencēšanai, kā arī repozitorijā ievietoto valodas datu citēšanai.

CLARIN ERIC

CLARIN ir digitāla infrastruktūra, kas piedāvā datus, rīkus un pakalpojumus valodas resursos balstītas pētniecības atbalstam. CLARIN piedāvā:

- ērti lietojamus valodas resursus un rīkus (skat. nodaļas “Vienotā valodas resursu un rīku krātuve” un “Valodas resursu saimes”) – piedāvājot rīkus valodas datu izpētei, analīzei un saglabāšanai, CLARIN nodrošina infrastruktūru pētījumiem, kuru pamatā ir digitālie valodu resursi,
- zināšanu infrastruktūru (skat. nodaļu “CLARIN-LV zināšanu infrastruktūra”) – zināšanu infrastruktūras mērķis ir nodrošināt, lai CLARIN konsorcijs pieejamās zināšanas būtu sasniedzamas,
- finansējumu (skat. nodaļu “CLARIN atbalsts sadarbībai valodas resursu izveidē”) – CLARIN veicina pētniecību dažādās jomās, regulāri izsludinot konkursus projektiem, kas sekmē CLARIN infrastruktūras resursu izmantošanu un paplašināšanu.

CLARIN infrastruktūras un centru darbību nodrošina CLARIN dalībvalstu nacionālie konsorcijs. Nacionālos konsorcijs parasti veido universitātes, institūti, bibliotēkas un publiskie arhīvi. Katrā dalībvalstī ir vismaz viens centrs (repozitorijs), kurā tiek apkopoti valodas resursi un rīki.

1 CLARIN *Virtual Language Observatory* (VLO), <https://vlo.clarin.eu/>

2 CLARIN *Resource families*, <https://www.clarin.eu/resource-families>

Vienotā valodas resursu un rīku krātuve

CLARIN virtuālā valodas resursu krātuve ir katalogs, kurā apkopoti CLARIN dalībvalstu centru dokumentēto valodas resursu un rīku (VRR) metadati. VLO nodrošina plašas meklēšanas iespējas, ļaujot atrast nepieciešamo starp vairāk nekā 1,25 miljoniem valodas datu ierakstu.

Kad nepieciešamais VRR ir atrasts, lietotājs tiek novirzīts uz to CLARIN repozitoriju, kurā tas noglabāts. Repozitorijs nodrošina piekļuvi VRR atbilstoši tā licencēšanas nosacījumiem, t. i., brīvpieejas resursi ir lejupielādējami vai pārlūkojami, bet ar licenci aizsargātiem resursiem var piekļūt, izmantojot vienotu pierakstīšanos.

VLO ir apkopota informācija ne tikai par CLARIN centros reģistrētajiem valodas resursiem un rīkiem, bet arī par citu iniciatīvu valodas resursiem. Piemēram, nozīmīgu daļu latviešu valodas ierakstu veido vairākas Europeana³ Latvijas Nacionālās bibliotēkas kolekcijas (*Europeana – National Library of Latvia*). Pašlaik VLO apkopota informācija par 2555 latviešu valodas resursiem un rīkiem no 32 dažādiem repozitorijiem un kolekcijām⁴. VLO saturs tiek atjaunots divas reizes nedēļā.

Valodas resursu kopas

Digitālo humanitāro, humanitāro un sociālo zinātņu pētniekiem noderīgi ir CLARIN resursu kopu (*Resource Families*) iniciatīvas sagatavotie pārskati par CLARIN infrastruktūrā pieejamajām valodas resursu un rīku saimēm. Pašlaik ir sagatavoti pārskati par 12 korpusu saimēm (datorizētas saziņas korpusi, akadēmisko tekstu korpusi, vēsturisko tekstu korpusi, valodas apguvēju korpusi, literārie korpusi, manuāli marķētie korpusi, multimodālie korpusi, laikrakstu korpusi, paralēlie korpusi, parlamentāro debašu korpusi, atsaucēs korpusi (*Reference corpora*), runas korpusi), piecām leksisko resursu saimēm un četrām rīku saimēm (normalizācijas rīki, nosaukto entitāšu atpazīšanas rīki, morfoloģiskās analīzes un lemmatizācijas rīki, noskaņojuma analīzes rīki).

CLARIN atbalsts sadarbībai valodas resursu izveidē

CLARIN ERIC finansiāli atbalsta dažādas iniciatīvas, kuru mērķis ir sekmēt sadarbību CLARIN valstu konsorcijs starpā CLARIN statūtos minēto uzdevumu veikšanai.

Viena no iniciatīvām ir jaunu valodas resursu izveide, kas sekmīgi realizēta projektā *ParlaMint* (Erjavec et al. 2022). Projekta mērķis bija izveidot daudzvalodu parlamentāro debašu korpusu ar standartizētiem metadatiem, vienotu datu un marķējuma formātu. Korpusā (Erjavec et al. 2021) iekļauti 17 valstu parlamentu dati, tajā skaitā arī Latvijas Republikas

3 <https://www.europeana.eu/en/item/97/418759>

4 VLO skatīts 05.05.2022.

Saeimas sēžu stenogrammas. Visi dati ir automātiski morfoloģiski un sintaktiski marķēti, pievienots arī nosaukto entitāšu marķējums. Dati ir sagatavoti TEI (*Text Encoding Initiative*) formātā un ir pieejami gan meklēšanai *noSketchEngine* platformā⁵, gan lejupielādei no CLARIN.SI repozitorijā ar lingvistisko marķējumu⁶ un bez tā⁷.

CLARIN-LV valodas resursu un rīku krātuve kā pētniecības datu repozitorijs

CLARIN-LV repozitorijs ir viens no vairāk nekā 60 virtuālajiem valodas resursu centriem. Tā mērķis ir sekmēt digitālo humanitāro zinātņu pētījumus, apkopojot latviešu un citu Latvijā pētītu valodu resursus un rīkus, nodrošinot VRR pieejamību ilgtermiņā un rūpējoties par to izmantošanu atbilstoši licences nosacījumiem.

Valodas resursi

Valodas resursu uzkrāšana un dokumentēšana CLARIN-LV repozitorijā tika sākta 2020. gada vasarā, pēc CLARIN-LV repozitorija izveides un stabilizēšanas un sākotnējās valodas resursu un rīku apzināšanas. Pašlaik (t. i., 2022. gada maijā) repozitorijā iekļauti 35 valodas resursi – 22 korpusi, 11 leksikoni un 2 rīki. Vairāku valodas resursu metadati ir sagatavošanā vai iesniegti izvērtēšanai. Repozitorijā ir gan monolingvālas, gan multilingvālas datu kopas. Lai arī lielākā daļa pašlaik CLARIN-LV repozitorijā iekļauto valodas resursu veidoti LU MII Mākslīgā intelekta laboratorijā, krātuvei pievienoti pirmie citu institūciju – LU Literatūras, folkloras un mākslas institūta, LU Latviešu valodas institūta, Rēzeknes Tehnoloģiju akadēmijas, Ventspils Augstskolas un Latviešu valodas aģentūras – izstrādātie resursi. Šobrīd visi CLARINLV repozitorijā iekļautie valodas resursi un rīki ir brīvpieejši – tos var izmantot pētniecībā caur pārlūkprogrammu, korpusu pārvaldības rīku *noSketchEngine* vai kādu citu platformu, vairāk nekā 10 datu kopas ir lejupielādējamas un brīvi izmantojamas atbilstoši licencei.

Citēšana

Būtiski pieaugot digitālo datu daudzumam un to lietojumam, kvalitatīvu datu kopu, it īpaši valodas resursu, izveide ir nozīmīgs pētniecības rezultāts, kuru, līdzīgi kā publikāciju, ir

5 <http://www.clarin.si/noske/>

6 <https://www.clarin.si/repository/xmlui/handle/11356/1432>

7 <https://www.clarin.si/repository/xmlui/handle/11356/1431>

Corpus of Latvian Pandemic Diaries 2020–2021



Please use the following text to cite this item or export to a predefined format:

BIBTEX

CMDI

Reinsone, Sanita; Ļaksa-Timinska, Ilze and Jaudzema, Justine, 2021, *Corpus of Latvian Pandemic Diaries 2020–2021*, CLARIN-LV digital library at IMCS, University of Latvia, <http://hdl.handle.net/20.500.12574/48>.



bibtex

```
@misc{20.500.12574/48,
  title = {Corpus of Latvian Pandemic Diaries 2020–2021},
  author = {Reinsone, Sanita and Ļaksa-Timinska, Ilze and Jaudzema, Justine},
  url = {http://hdl.handle.net/20.500.12574/48},
  note = {{CLARIN}-LV digital library at {IMCS}, University of Latvia},
  copyright = {Creative Commons - Attribution-ShareAlike 4.0 International ({CC} {BY}-{SA} 4.0)},
  year = {2021} }
```



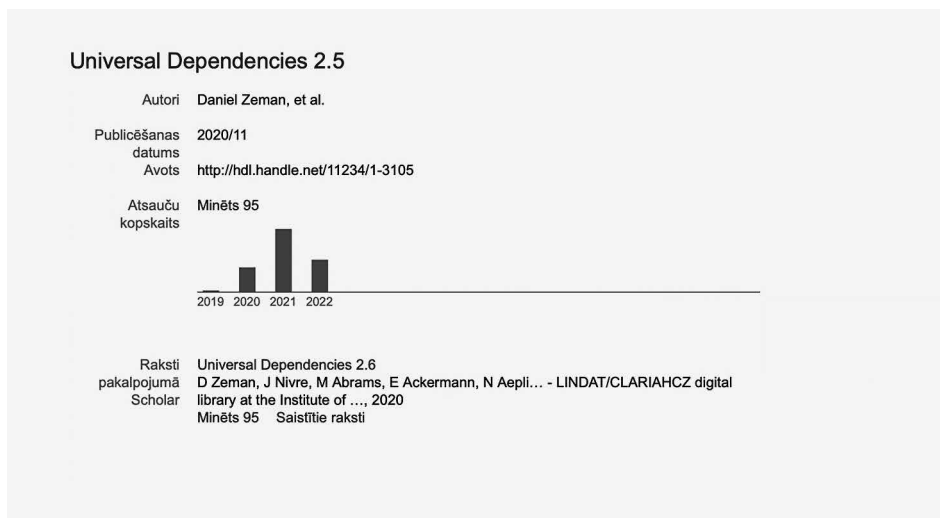
1. attēls. CLARIN-LV sagatavotās Pandēmijas dienasgrāmatu datu kopas (Reinsone et al. 2021) atsauce tekstualā un *bibtex* formātā. Attēlā redzams arī ieraksta unikālais identifikators <http://hdl.handle.net/20.500.12574/48>.

nepieciešams citēt. Datu pieejamību un to citēšanu sekmē atvērtās zinātnes (*Open Science*) un FAIR principu⁸ ieviešana Eiropā un Latvijā.

Iespēju citēt datu kopas piedāvā arī CLARIN repozitoriji. Katrai repozitorijā dokumentētai datu kopai vai rīkam tiek piekārtots unikāls identifikators (PID – *persistent identifier*), kas ļauj to viennozīmīgi identificēt un citēt (skat. 1. attēlu). Citēšanai nepieciešamo atsauci repozitorijš ģenerē automātiski no metadatiem, kurus sagatavojis valodas resursa iesniedzējs (deponētājs). Tas ne tikai atvieglo citēšanas procesu resursa izmantotājam, bet arī palīdz iegūt informāciju, cik pētnieku šo resursu ir citējuši (izmantojuši), kas bieži ir būtiski finansējuma iegūšanai resursa uzturēšanai un attīstīšanai. Repozitorijš arī uzskaita resursa skatījumu un lejupielāžu skaitu (skat. nodaļu “Populārākie valodas resursi”).

Automātiski veidotie citējumi, kā arī valodas resursu ieraksti CLARIN katalogā sekmē datu kopas pamanāmību un lietojumu ne tikai Latvijā, bet arī Eiropā un citur pasaulē. Piemēram, *Google Scholar* atsauces uz CLARIN-LV un citos CLARIN repozitorijos apkopotajiem valodas resursiem iekļauj autoru profilos. *Google Scholar* atsauce uz Universālo atkarību datu kopu (*Universal Dependencies 2.5*), kas iekļauta LINDAT/CLARIAH-CZ repozitorijā un kurā ietverts arī sintaktiski marķēts latviešu valodas korpuss, ietver gan repozitorija nosaukumu, gan unikālo identifikatoru, gan publicēšanas gadu un citējumu skaitu (skat. 2. attēlu).

8 FAIR – *Findability, Accessibility, Interoperability, and Reuse of digital assets*, skat. <https://www.go-fair.org/fair-principles/>



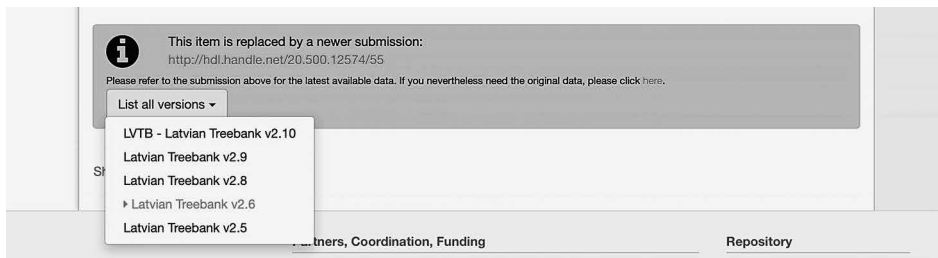
2. attēls. *Google Scholar* atsauce uz Universālo atkarību datu kopām.

Deponēšana

Viens no CLARIN pētniecības infrastruktūras pamatuzdevumiem ir valodas resursu un rīku uzglabāšana ilgtermiņā. Atbilstoši CLARIN ERIC statūtiem katrā dalībvalstī jābūt vismaz vienam sertificētam datu centram, kas tehniski var nodrošināt drošu datu uzglabāšanu un ar to uzturēšanu saistītus pakalpojumus.

CLARIN repozitorijos tiek uzglabāti valodas resursi, kurus veidojuši zinātnieki dažādu projektu laikā. Svarīgi, ka šie resursi ir uzticami, t. i., tos pēc projekta beigām var izmantot citi pētnieki. Tāpēc CLARIN (un CLARIN-LV kā dalībnieks) piedāvā zinātniekiem valodas resursus deponēt, t. i., nodot ilgtermiņa uzglabāšanā.

Datu iesniegšanu parasti veic zinātnieks (kāds no autoriem) vai zinātniskās institūcijas nozīmēts datu pārvaldnieks. Lai iesniegums nebūtu anonīms un dati būtu uzticami, iesniedzējam jāpieslēdzas CLARIN repozitorijam, izmantojot savu akadēmiskā lietotāja kontu (skat. arī nodaļu “Licencēšana un vienotā pierakstīšanās”). Pēc pieslēgšanās CLARIN repozitorijam iesniedzējs sagatavo valodas resursa vai rīka metadatus (informāciju par valodas resursu vai rīku), aprakstot iesniedzamo datu kopu. Viens no svarīgākajiem metadatu laukiem ir VRR autori (veidotāji), šis lauks ļauj atsaukties uz datu kopu un tās autoriem līdzīgi kā publikācijas gadījumā. Tāpat iespējams norādīt publikāciju, kurā datu kopa vai rīks aprakstīts. Ne mazāk būtiska informācija ir finansējuma avots. Norādītie metadati var tikt izmantoti meklēšanai un valodas resursu atlasīšanai.



3. attēls. Latviešu valodas sintaktiski marķētu tekstu korpusu (*Latvian Treebank*) versijas CLARIN-LV repozitorijā.

Pēc metadatu sagatavošanas deponējamo valodas resursu vai rīku iespējams augšupielādēt CLARIN repozitorijā, norādot tā izmantošanas nosacījumus (licenci). Iesniedzot datus CLARIN repozitorijam, iesniedzējs arī dod tiesības CLARIN-LV repozitorijam izplatīt datus atbilstoši licences nosacījumiem⁹.

Sagatavotie metadati un pievienotā datu kopa tiek publicēta tikai pēc tam, kad to pārskatījis CLARIN-LV repozitorija datu pārvaldnieks, kura uzdevums ir pārliecināties par iesniegto metadatu un pievienoto datu kvalitāti un ticamību.

Laika gaitā datu kopas var tikt mainītas un papildinātas. CLARIN repozitoriji neatbalsta izmaiņas iesniegtajos datos (izņēmums var būt kļūdas un neprecizitātes datu kopas metadatos), bet piedāvā iespēju veidot vairākas datu kopas versijas. Tas ļauj zinātniekiem izvēlēties pētījumam piemērotāko versiju. Piemēram, ja nepieciešams kādu pētījumu atkārtot vai veikt rezultātu salīdzināšanu, piemērotāka var izrādīties kāda vecāka datu kopas versija (skat. 3. attēlu).

Tā kā valodas resursu krātuves un deponēšana ir jaunums Latvijā, 2021. gada sākumā tika rīkots praktiskais seminārs par valodas resursu deponēšanu, kura dalībnieki praksē iepazinās ar deponēšanas gaitu, kā arī tika izveidota deponēšanas instrukcija¹⁰, kuru var izmantot ikviens pētnieks, kas vēlas iesniegt pētniecības datu kopas ilgtermiņa uzglabāšanai CLARIN-LV repozitorijā.

Licencēšana un vienotā pierakstīšanās

Līdz ar atvērtās zinātnes kustības izplatību Latvijā un pasaulē arvien aktuālāka kļūst datu kopu autorība, to licencēšana un tajā iekļauto datu autortiesību un privātuma jautājumi. Lai šos jautājumus risinātu, CLARIN ir izveidota Juridisko jautājumu komiteja, kas izstrādā vadlīnijas un ieteikumus ar datu pārvaldību saistītos jautājumos un seko līdzi aktualitātēm.

⁹ Deponēšanas process detalizēti aprakstīts repozitorija lapā: <https://repository.clarin.lv/repository/xmlui/page/deposit>

¹⁰ https://www.clarin.lv/attachments/CLARIN%202020_resursu_iesniegsana.pdf

Prague Czech-English Dependency Treebank 2.0

Please use the following text to cite this item or export to a predefined format: **BIBTEX** **CMDI**

Hajič, Jan; et al., 2012, *Prague Czech-English Dependency Treebank 2.0*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11858/00-097C-0000-0015-8DAF-4>.

This resource is also integrated in following services: **PML-TQ** **KonText**

Share: **f** **t**

LINDAT / CLARIAH-CZ

This item is **Restricted Use** and licensed under:
CC-BY-NC-SA + LDC99T42

Sign in to **LINDAT/CLARIAH-CZ Repository**
Login via Your home institution (e.g. university)

- Univerzita Karlova v Praze
Czech Republic 998 km
- Institute of Mathematics and Computer Science, University of Latvia
Latvia
- Ecole Estienne

4. attēls. Vienotās pierakstīšanās piemērs piekļuvei
LINDAT/CLARIAH-CZ repozitorijā *Prague Czech-
English Dependency Treebank 2.0*.

Katrai CLARIN datu kopai, kas iesniegta repozitorijā, tiek norādīta licence, ar kādu zinātnieki un interesenti var to izmantot. Licenci datu kopai izvēlas datu kopas iesniedzējs (parasti zinātnieks vai zinātnieku grupa, kas šo valodas resursu veidojusi). Tā kā zinātnieki ne vienmēr pārzina dažādu licenču specifiku, ir izveidots CLARIN licenču ieteicējs, kas, lietotājam atbildot uz vairākiem jautājumiem, palīdz izvēlēties piemērotāko licenci.

Tā kā CLARIN atbalsta atvērto zināt¹¹, tad datu kopām ieteikts izmantot *Creative Commons* licences. Šai licenču grupai pievienotās pazīmes apraksta licencēšanas nosacījumus. Biežāk lietotās pazīmes (atribūti) ir: BY – prasa norādīt datu kopas autorību un ļauj to brīvi izmantot, SA – nosaka, ka atvasinātās datu kopas jāizplata ar tādu pašu licenci kā oriģināls, NC – liedz komerciālu lietojumu, ND – liedz veidot atvasinātus un adaptētus darbus. Vairums latviešu valodas resursu un rīku, kuri katalogizēti CLARIN VLO, ir publiski pieejami.

11 <https://www.clarin.eu/eosc>

Items Viewed

(more than 20 times)

Item/Handle	Number of views
Tēzauris.lv 2020 (Spektors, Andrejs et al) (20.500.12574/9)	185 [+bots: 188]
LVBERT - Latvian BERT (Znotiņš, Artūrs) (20.500.12574/43)	125 [+bots: 127]
Latvian Treebank v2.8 (Rituma, Laura et al) (20.500.12574/55)	117 [+bots: 117]
Latvian AMR Sembank (Znotiņš, Artūrs et al) (20.500.12574/40)	112 [+bots: 113]
LUIS: data collection for task oriented dialogue system creation (Gunta, Nešpore-Bērzkalne et al)...	104 [+bots: 104]

5. attēls. Pieci visvairāk skatītie valodas resursi
2022. gada aprīlī.

Lai gan *Creative Commons* licence ir ļoti populāra, jo ļauj datus brīvi izmantot gan pētniecībā, gan vairumā gadījumu arī jaunu darbu izveidei, ne vienmēr ar datu kopām var tik brīvi rīkoties. Lai kontrolētu šādu datu kopu lietojumu, CLARIN repozitoriji izmanto vienotu pierakstīšanas ierobežota lietojuma datu kopu lietojuma kontroli. Vienotā pierakstīšanās (Latvijā tiek realizēta, izmantojot LAIFE identitāšu federāciju) nodrošina gan datu kopu izmantošanu atbilstoši licencei (to lietotājs digitāli apstiprina pirms lietošanas uzsākšanas), gan arī iespēju izmantot ierobežotas piekļuves datu kopas kādā no CLARIN ERIC repozitorijiem. Piemēram, LINDAT CLARIN repozitorijā Prāgā ir apkopoti daudzu valodu, t. sk. latviešu valodas, sintaktiski marķēti korpusi. Vairums (84 no 109) resursu ir izmantojami ar CC-BY-NC-SA licenci, bet, piemēram, Prāgas čehu-angļu atkarību sintaktiski marķētā korpusa 2.0 versija¹² tiek izplatīta ar papildu ierobežojumiem, tāpēc pirms lejupielādes ar vienoto pierakstīšanas tiek identificēts datu lietotājs (skat. 4. attēlu).

Populārākie valodas resursi un rīki

Valodas resursu uzkrāšana un dokumentēšana CLARIN-LV repozitorijā tika sākta 2020. gada vasarā, vispirms iekļaujot LU MII Mākslīgā intelekta laboratorijas populārākos valodas resursus. Pēc praktiskā semināra 2021. gada ziemā valodas resursus sāka reģistrēt arī citas zinātniskās organizācijas. 2021. gada vasaras sākumā repozitorijs tika papildināts ar populārākajiem valodas korpusiem, bet vēlāk ar leksikoniem.

Trīs visvairāk skatītie valodas resursi ir *Tēzauris.lv* (Spektors et al. 2019), “Līdzsvarotais latviešu valodas korpus” (Levāne-Petrova un Darģis 2018) un “Sintaktiski marķētais latviešu valodas korpus v2.5” (Rituma et al. 2020), kuri visi ir reģistrēti repozitorijā izveides pirmsākumos.

12 <http://hdl.handle.net/11858/00-097C-0000-0015-8DAF-4>

Tā kā dažādi valodas resursi un rīki repozitorijā iekļauti atšķirīgā laikā, precīzāku informāciju par populārākajiem resursiem sniedz statistika griezumā (skat. 5. attēlu). Redzams, ka lietotājus interesē gan valodas korpusi, gan leksikoni. Tāpat vērojama interese par mašīnlasāmām datu kopām un modeļiem. Tomēr interese par valodas apstrādes rīkiem ir mazāka, ko varētu skaidrot ar to, ka CLARIN lietotāji galvenokārt ir humanitāro zinātņu pētnieki, kurus vairāk interesē valodas resursi, kas ir ērti izmantojami, nevis iespēja veidot savus risinājumus, izmantojot repozitorijā iekļautos datus.

Analizējot šo statistiku, ir būtiski atcerēties, ka tā norāda tikai valodas resursa pamanāmību CLARIN-LV repozitorijā, bet neatspoguļo kopējo datu kopas lietojumu (piem., korpusa lietojumu, izmantojot *SketchEngine* platformu).

CLARIN-LV zināšanu infrastruktūra

CLARIN pētniecības infrastruktūra nav tikai valodas resursu un rīku repozitorijs, bet arī zināšanas, ko uzkrājuši daudzie CLARIN centri un nacionālie konsorcijs. Zināšanu apmaiņa ietver darbības, kas saistītas ar lietotāju iesaistīšanu un izglītošanu, piemēram, mācības (darbsemināri, lekcijas), diskusijas un zināšanu izplatīšanas pasākumi (konferences, semināri), kā arī ikdienas atbalsts, izmantojot zināšanu centrus.

Pakalpojumi humanitāro zinātņu pētnieku atbalstam – lietotāju iesaiste, izglītība un konsultācijas

Lai iepazīstinātu Latvijas humanitāro un sociālo zinātņu pārstāvjus ar CLARIN resursiem un rīkiem, mācītu, kā izmantot latviešu valodas resursus, pēdējo gadu laikā ir organizēti vairāki semināri.

2018. gada sākumā seminārs “Valodas resursi un rīki digitālajām humanitārajām zinātnēm” tika organizēts, lai demonstrētu LU MII Mākslīgā intelekta laboratorijā izstrādātos valodas rīkus un resursus. Seminārā plašāka publika pirmo reizi pēc Latvijas pievienošanās CLARIN ERIC tika iepazīstināta ar infrastruktūras nacionālajiem un starptautiskajiem mērķiem, un semināra dalībnieki tika aicināti aktīvi piedalīties CLARIN ekspertu tīkla izveidošanā Latvijā. Seminārā pulcējās humanitāro un sociālo zinātņu pārstāvji, tostarp filologi, žurnālisti, politiologi, tulki, bibliotekāri, vēsturnieki.

Papildus ir organizēti vairāki praktiski semināri, kuros pētnieki iepazīstināti ar vairākiem latviešu valodas korpusiem – “Līdzsvaroto mūsdienu latviešu valodas tekstu korpusu” (LVK2018), “Latviešu valodas sintaktiski marķēto korpusu”, “Latviešu valodas apguvēju korpusu” (LaVA).

Informācija par CLARIN resursiem un rīkiem ir integrēta Latvijas Universitātes Humanitāro zinātņu fakultātes maģistra studiju programmas kursā “Ievads datorlingvistikā”. Kursa laikā studenti tiek iepazīstināti ar pētniecības infrastruktūru, tiek aicināti izpētīt CLARIN repozitoriju – atrast saviem pētījumiem noderīgus resursus un rīkus – un pastāstīt

par saviem atklājumiem seminārā. Savukārt topošajiem skolotājiem Latvijas Universitātes Pedagoģijas, psiholoģijas un mākslas fakultātē tiek piedāvāts kurss “Valodas korpuss izglītības procesā”. Par valodas resursiem ir stāstīts arī demonstrāciju seminārā “Digitālie resursi novadu savdabības atklāšanai, izglītībai un kultūrai”.

Zināšanu infrastruktūra

Lai palīdzētu zinātniekiem neapjukt milzīgajā CLARIN valodas resursu daudzveidībā un sekmētu zināšanu apmaiņu un pārnesi, CLARIN ir izveidoti vairāk nekā 20 zināšanu centri¹³ (*knowledge centres*, *K-Centres*, K-centri). K-centri konsultē ne tikai par kādu konkrētu valodu vai modalitāti (runātā valoda, zīmju valoda u. c.), bet arī par valodas apstrādes rīkiem, dažādiem datu kopu veidiem u. c. ar valodas pētniecību saistītiem jautājumiem.

CLARIN-LV kopā ar Helsinku Universitāti (FIN-CLARIN), Trumses Universitāti (CLARINO) un Vītauta Dižā Universitāti (CLARIN-LT) apvienojušies zināšanu centrā SAFMORIL¹⁴, kas sniedz atbalstu digitālo humanitāro zinātņu pētniekiem, valodniekiem un valodu tehnoloģiju izstrādātājiem, kuri analizē un apstrādā morfoloģiski bagātu valodu datus. Pašlaik CLARIN-LV ar partneriem gatavo pieteikumu leksikogrāfijas zināšanu centra ELEXIS izveidei.

CLARIN-LV sniedz konsultācijas par latviešu valodas resursiem un rīkiem. Līdz šim biežāk uzdotie jautājumi saistīti ar valodas korpusiem, tajos lietoto marķējumu un meklēšanas iespējām. Tāpat vairākiem pētniecības centriem sniegti padomi un praktiska palīdzība valodas korpusu izveidē.

Veiksmīga sadarbība izveidojusies ar Latvijas Nacionālo bibliotēku, palīdzot uzstādīt *noSketchEngine* platformu un apmācot resursu sagatavošanā un ievietošanā. Gan LNB, gan citas zinātniskās organizācijas kā servisu aktīvi izmanto latviešu valodas apstrādes rīkkopu NLP-PIPE¹⁵ (Znotiņš 2015), kas piedāvā tekstu morfoloģisko marķēšanu, sintaktisko parsēšanu un nosaukto entitāšu atpazīšanu.

Iedvesmojoties no CLARIN vienotā (*federated*) satura meklēšanas servisa CLARIN-FCS¹⁶, kas ļauj meklēt valodas korpusos neatkarīgi no tā darbināšanas vietas, CLARIN-LV ir izveidojis vienotu meklēšanu Latvijas korpusos *noSketchEngine* platformā¹⁷ (skat. 6. attēlu). Šobrīd meklēšana tiek veikta tikai LU MII korpusos, bet tiek strādāts arī pie LNB korpusu pievienošanas meklēšanas sistēmai.

13 <https://www.clarin.eu/content/knowledge-centres>

14 <https://www.clarin.eu/blog/tour-de-clarin-clarin-knowledge-centre-systems-and-frameworks-morphologically-rich-languages>

15 <http://nlp.aillab.lv/>

16 <https://www.clarin.eu/content/federated-content-search-clarin-fcs>

17 <http://www.korpuss.lv/search>

Nacionālā korpusu kolekcija

MeklēšanaMateriāliPar NKK

LV

valoda

Meklēt

Vaicājumam **valoda** atrasti 736 314 pieminējumi 23 no 24 korpusiem

Korpus	Relatīvais biežums uz 1 miljonu	Absolūtais biežums	Par korpusu
LaVA Latviešu valodas apguvēju korpus	3099	747	Vairāk: korpus.lv/id/LaVA Izstrādātāji: LU MIL Mezglis: nosketch.korpus.lv
Disertācijas Promocijas darbu korpus	1186	27 687	Vairāk: korpus.lv/id/Disertācijas Izstrādātāji: LU MIL Mezglis: nosketch.korpus.lv
Vikipēdija Vikipēdijas korpus	898	24 888	Vairāk: korpus.lv/id/Vikipēdija Izstrādātāji: LU MIL Mezglis: nosketch.korpus.lv
Saeima LR 5.–12. Saeimas sēžu stenogramu korpus	873	21 029	Vairāk: korpus.lv/id/Saeima Izstrādātāji: LU MIL, RSU Mezglis: nosketch.korpus.lv

6. attēls. Vienotā meklēšana latviešu valodas korpusos.

CLARIN-LV nākotnes uzdevumi un loma Latvijas digitālo humanitāro zinātņu attīstībā

Pēdējo piecu gadu laikā CLARIN-LV ir kļuvis par stabilu un atpazīstamu valodas resursu krātuvi Latvijā un Eiropā. CLARIN-LV ļauj zinātniekiem izmantot repozitorijā reģistrētos resursus, sniedz konsultācijas par to lietojumu. CLARIN-LV mērķis ir kļūt par tādu starptautiski integrētu nacionālā līmeņa valodas resursu un rīku infrastruktūru, kas pilnvērtīgi nodrošina digitālo humanitāro zinātņu, datorlingvistikas un valodu tehnoloģiju pētnieku kopienų vajadzības un sekmē pilnvērtīgu latviešu valodas dzīvi digitālajā laikmetā.

Šī mērķa sasniegšanai nepieciešams turpināt darbu pie repozitorija satura, papildinot to ar Latvijā pētniecības projektos tapušiem valodas resursiem un rīkiem. Svarīga ir arī valodas resursu un rīku pielāgošana pētnieku vajadzībām. Jau pašlaik citos CLARIN centros, piemēram, Čehijā, Islandē, Polijā un Slovēnijā, CLARIN nav tikai repozitorijs un kompetences centrs, bet arī digitālajām humanitārajām zinātnēm pielāgoti pakalpojumi un vide pētniecībai. Pirmie pielāgotie risinājumi tiek veidoti Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi” projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001), kur latviešu valodas rīkkopa NLP-PIPE tiek pielāgota izmantošanai Latvijas Nacionālās bibliotēkas tekstu apstrādē.

Nākotnē ir svarīgi turpināt ciešu DHZ pārstāvju, datorlingvistu un datorzinātnieku sadarbību jaunos pētniecības projektos un Valsts pētījumu programmās, tādējādi nodrošinot gan DHZ tālāku uzplaukumu, gan sekmējot valodas resursu izveidi un pilnveidi. Nepieciešams

sistemātisks un mērķtiecīgs valsts atbalsts ilgtspējīgiem pētījumiem valodas resursu un rīku izveidē un uzturēšanā, kas nodrošinātu valodu līdztiesību ikdienas lietojumā un digitālajā vidē.

Vēl arvien novērojama būtiska zināšanu un izpratnes plaisa starp valodas resursu un rīku veidotājiem un to lietotājiem – humanitāro un sociālo zinātņu pētniekiem –, kas kavē CLARIN pētniecības infrastruktūras pilnvērtīgu izmantojumu. Viens no soļiem, kas ļautu palielināt savstarpēju sapratni un sadarbību, ir sekmīga Latvijas Atveseļošanas un noturības mehānisma plāna ieviešana¹⁸, kurš paredz augsta līmeņa digitālo prasmju apguvi valodu tehnoloģiju jomā. Sekmējot Latvijas atvērtās zinātnes stratēģijas ieviešanu, jāturpina darbs pie licencēšanas, īpaši atvērtās piekļuves, un FAIR principu popularizēšanas zinātnieku vidū un atbalsta CLARIN-LV infrastruktūrā.

18 <https://likumi.lv/ta/id/322858-par-latvijas-atveselosanas-un-noturibas-mehanismu>

- Erjavec, Tomaž, Ogrodniczuk, Maciej, Osenova, Petya, Ljubešič, Nikola, Simov, Kiril, Pančur, Andrej, Rudolf, Michal, Kopp, Matyáš, Barkarson, Starkaður, Steingrímsson, Steinþór, Çöltekin, Çağrı, de Does, Jesse, Depuydt, Katrien, Agnoloni, Tommaso, Venturi, Giulia, Calzada Pérez, D de Macedo, Luciana, María, Navarretta, Costanza, Luxardo, Giancarlo, Coole, Matthew, Rayson, Paul, Morkevičius, Vaidas, Krilavičius, Tomas, Dargis, Roberts, Ring, Orsolya, van Heusden, Ruben, Marx, Maarten, Fišer, Darja, (2022). The ParlaMint corpora of parliamentary proceedings. *Language Resources and Evaluation*.
- Erjavec, Tomaž et al. (2021). *Multilingual comparable corpora of parliamentary debates ParlaMint 2.1.*, Slovenian language resource repository CLARIN.SI. Available: <http://hdl.handle.net/11356/1432> [accessed 09.05.2022.].
- Levāne-Petrova, Kristīne, Dargis, Roberts (2018). *Balanced Corpus of Modern Latvian (LVK2018)*. CLARIN-LV digital library at IMCS, University of Latvia. Available: <http://hdl.handle.net/20.500.12574/11> [accessed 09.05.2022.].
- Rituma, Laura, Pretkalniņa, Lauma, Saulīte, Baiba, Nešpore-Bērzkalne, Gunta, Grūzītis, Normunds (2020). *Latvian Treebank v2.5*. CLARIN-LV digital library at IMCS, University of Latvia. Available: <http://hdl.handle.net/20.500.12574/10> [accessed 09.05.2022.].
- Reinsone, Sanita, Ļaksa-Timinska, Ilze, Jaudzema, Justīne (2021). *Corpus of Latvian Pandemic Diaries 2020–2021*. CLARIN-LV digital library at IMCS, University of Latvia. Available: <http://hdl.handle.net/20.500.12574/48> [accessed 09.05.2022.].
- Skadina, Inguna, Auzina, Ilze, Gruzitis, Normunds, Znotins, Arturs (2020). *Clarín in Latvia: From the preparatory phase to the construction phase and operation. Proceedings of the 5th Conference on Digital Humanities in the Nordic Countries (DHN)*.
- Spektors, Andrejs, Pretkalniņa, Lauma, Grūzītis, Normunds, Paikens, Pēteris, Rituma, Laura, Saulīte, Baiba (2019). *Tēzauris lv 2020*. CLARIN-LV digital library at IMCS, University of Latvia. Available: <http://hdl.handle.net/20.500.12574/9> [accessed 09.05.2022.].
- Váradí, Tamás, Krauwer, Steven, Wittenburg, Peter, Wynne, Martin, Koskenniemi, Kimmo (2008). CLARIN: Common Language Resources and Technology Infrastructure. *Proceedings of the Sixth International Conference on Language Resources and Evaluation*.
- Znotiņš, Artūrs (2015). *NLP-PIPE: Latvian NLP Tool Pipeline*. CLARIN-LV digital library at IMCS, University of Latvia. Available: <http://hdl.handle.net/20.500.12574/4> [accessed 09.05.2022.].

CLARIN Language Resources and Technology Infrastructure for the Humanities and Social Sciences

Inguna Skadiņa, Ilze Auziņa,
Roberts Darģis, Arnis Voitekāns

Keywords: research infrastructures, digital humanities, language resources, Latvian language, knowledge infrastructure, open science, FAIR

Established in 2012, CLARIN research infrastructure aims to maintain an infrastructure to support the sharing, use and sustainability of the language data and tools for research in the social sciences and humanities (SSH). In Latvia, after joining CLARIN ERIC in 2016, work on creation of CLARIN research infrastructure started in 2018. This paper aims to provide overview of CLARIN infrastructure and introduce the researchers of Latvian digital humanities to the opportunities offered by CLARIN ERIC and in particular CLARIN Latvia (CLARIN-LV) node.

At first, we introduce to the fundamental elements of CLARIN ERIC – Virtual Language Observatory, Language Resource Families and funding and cooperation mechanisms. Then, we provide an overview of CLARIN-LV repository and language resources and tools documented. We explain citation mechanisms and provide practical recommendations for depositing and licensing. The core values of CLARIN are very closely aligned with the FAIR data principles, therefore we stress the importance of long-term preservation of research outcomes (including language resources and tools) and explain CLARIN role in supporting open science and FAIR principles. Furthermore, mechanisms for the knowledge sharing and user support are presented. It includes CLARIN knowledge centres, targeted seminars, university level courses and individual consultations. Finally, paper outlines future goals of CLARIN-LV and the next steps to implement them. The main directions include extension and adaptation of repository content, long-term cooperation with SSH researchers, support for higher level education in language technologies and support for implementation of open science principles.

Autori / Authors

Ančevska, Ieva – *Dr. philol.*,
folkloriste. Liepājas Universitātes
Kurzemes Humanitārais institūts.
ieva.ancvska1@gmail.com

Andronova, Everita – *Mg. philol.*,
valodniece. Latvijas Universitātes
Matemātikas un informātikas in-
stitūts. *everita.andronova@lumii.lv*

Apenīte, Mārite – *Mg. philol.*,
Mg. sc. soc., informācijas zinātniece.
Latvijas Nacionālā bibliotēka.
marite.apenite@lnb.lv

Auziņa, Ilze – *Dr. philol.*,
valodniece. Latvijas Universitātes
Matemātikas un informātikas
institūts. *ilze.auzina@lumii.lv*

Baklāne, Anda – *Mg. phil.*, digitālās
pētniecības pakalpojumu vadītāja.
Latvijas Nacionālā bibliotēka.
anda.baklane@lnb.lv

Bojārs, Uldis – *Dr. sc. comp.*,
datorzinātnieks. Latvijas Nacionālā
bibliotēka. *uldis.bojars@lnb.lv*

Briška, Anna – inženiere. Rēzeknes
Tehnoloģiju akadēmija.
anna.briska@rta.lv

Darģis, Roberts – *Mg. sc. comp.*,
datorzinātnieks. Latvijas Universi-
tātes Matemātikas un informātikas
institūts. *roberts.dargis@lumii.lv*

Daugavietis, Jānis – *Dr. sc. soc.*,
sociologs. Latvijas Universitātes
Literatūras, folkloras un mākslas in-
stitūts. *janis.daugavietis@lulfmi.lv*

Eglāja-Kristsone, Eva – *Dr. philol.*,
literatūrzinātniece. Latvijas
Universitātes Literatūras, folkloras
un mākslas institūts.
eva.eglaja@lulfmi.lv

Eversone, Madara – *Mg. philol.*,
literatūrzinātniece. Latvijas Uni-
versitātes Literatūras, folkloras un

mākslas institūts.
madara.eversone@lulfmi.lv

Fridenberga, Anna – *Dr. philol.*,
valodniece. Latvijas Universitātes
Latviešu valodas institūts.
anna.fridenberga@lu.lv

Garda, Ansis – komunikācijas
zinātnieks. Latvijas Nacionālā
bibliotēka. *ansis.garda@lnb.lv*

Goldberga, Anita – *Mg. philol.*,
Mg. sc. soc., informācijas zinātniece.
Latvijas Nacionālā bibliotēka.
anita.goldberga@lnb.lv

Grasmanis, Mikus – *Mg. sc. comp.*,
datorzinātnieks. Latvijas Universi-
tātes Matemātikas un informātikas
institūts. *mikus.grasmanis@lumii.lv*

Grūzītis, Normunds – *Dr. sc. comp.*,
datorzinātnieks. Latvijas Universitātes
Matemātikas un informātikas institūts.
normunds.gruzitis@lumii.lv

Juško-Štekele, Angelika – *Dr.*
philol., folkloriste. Rēzeknes Teh-
noloģiju akadēmija.
angelika.jusko-stekele@rta.lv

Kaija, Inga – *Dr. philol.*, valodniece.
Rīgas Stradiņa universitāte.
inga.kaija@rsu.lv

Karlsona, Agnese – *Mg. art.*, socio-
loģe. Latvijas Kultūras akadēmijas
Kultūras un mākslu institūts.
agnese.karlsona@lka.edu.lv

Kārkla, Zita – *Dr. philol.*, literatūr-
zinātniece. Latvijas Universitātes
Literatūras, folkloras un mākslas
institūts. *zita.karkla@lulfmi.lv*

Kļavinska, Antra – *Dr. philol.*,
valodniece. Rēzeknes Tehnoloģiju
akadēmija. *antra.klavinska@rta.lv*

Kreislere, Maira – *Mg. sc. soc.*,
informācijas zinātniece. Latvijas

Nacionālā bibliotēka.
maira.kreislere@lnb.lv

Kristāla, Ance – *Mg. sc. soc.*, socio-
loģe. Latvijas Kultūras akadēmija,
Kultūras un mākslu institūts.
ance.kristala@lka.edu.lv

Kunda, Ilona – *Dr. sc. soc.*, socio-
loģe. Latvijas Kultūras akadēmija,
Kultūras un mākslu institūts.
ilona.kunda@lka.edu.lv

Laime, Sandis – *Dr. philol.*,
folklorists. Latvijas Universitātes
Literatūras, folkloras un mākslas
institūts. *sandis.laime@lulfmi.lv*

Lasmanis, Viesturs Jūlijs – dator-
zinātņu students. Latvijas Universi-
tātes Matemātikas un informātikas
institūts. *viesturs.lasmanis@lumii.lv*

Levāne-Petrova, Kristīne –
Mg. philol., valodniece. Latvijas
Universitātes Matemātikas
un informātikas institūts.
kristine.levane-petrova@lumii.lv

Ļaksa-Timinska, Ilze – *Mg. philol.*,
literatūrzinātniece. Latvijas
Universitātes Literatūras,
folkloras un mākslas institūts.
ilze.laksa-timinska@lulfmi.lv

Martena, Sanita – *Dr. philol.*,
valodniece. Rēzeknes Tehnoloģiju
akadēmija. *sanita.martena@rta.lv*

Matulis, Haralds – *Mg. phil.*,
filozofs. Latvijas Universitātes Lite-
ratūras, folkloras un mākslas
institūts.
haralds.matulis@lulfmi.lv

Naua, Nikole (*Nicole Nau*) –
Dr. habil. phil., valodniece.
Poznaņas Adama Miceviča
Universitāte, Moderno valodu un
literatūru fakultāte un Rēzeknes
Tehnoloģiju akadēmija
nicole.nau@rta.lv

Nešpore-Bērzkalne, Gunta – *Mg. philol.*, valodniece. Latvijas Universitātes Matemātikas un informātikas institūts.
gunta.nespore@lumii.lv

Oga, Jānis – *Dr. philol.*, literatūrzinātnieks. Latvijas Universitātes Literatūras, folkloras un mākslas institūts. *janis.oga@lulfmi.lv*

Pokratniece, Kristīne – *Mg. philol.*, folkloriste. Latvijas Universitātes Matemātikas un informātikas institūts.
kristine.pokratniece@lumii.lv

Rapa, Sanda – *Dr. philol.*, valodniece. Latvijas Universitātes Latviešu valodas institūts.
sanda.rapa@lu.lv

Rašmane, Anita – *Mg. math.*, *Mg. sc. soc.*, matemātiķe, informācijas zinātniece. Latvijas Nacionālā bibliotēka. *anita.rasmane@lnb.lv*

Reinsone, Sanita – *Dr. philol.*, folkloriste. Latvijas Universitātes Literatūras, folkloras un mākslas institūts. *sanita.reinsone@lulfmi.lv*

Rituma, Laura – *Mg. philol.*, valodniece. Latvijas Universitātes Matemātikas un informātikas institūts. *laura.rituma@lumii.lv*

Saulespurēns, Valdis – *Mg. sc. comp.*, pētniecības tehnoloģiju analītiķis. Latvijas Nacionālā bibliotēka.
valdis.saulespurens@lnb.lv

Saulīte, Baiba – *Dr. philol.*, valodniece. Latvijas Universitātes Matemātikas un informātikas institūts.
baiba.valkovska@lumii.lv

Siliņa-Piņķe, Renāte – *Dr. philol.*, valodniece. Latvijas Universitātes Latviešu valodas institūts.
renate.silina-pinke@lu.lv

Skadiņa, Inguna – *Dr. sc. comp.*, datorzinātniece. Latvijas Universitātes Matemātikas un informātikas institūts. *inguna.skadina@lumii.lv*

Skrūzmane, Elga – *Mg. philol.*, valodniece. Latvijas Universitātes Latviešu valodas institūts.
elga.skruzmane@lu.lv

Spektors, Andrejs – *Dr. phys.*, matemātiķis. Latvijas Universitātes Matemātikas un informātikas institūts.
andrijs.spektors@lumii.lv

Stepanovs, Kaspars – *Dr.*, radiologs. Rīgas Austrumu klīniskā universitātes slimnīca.
kstepanovs@gmail.com

Trumpa, Anta – *Dr. philol.*, valodniece. Latvijas Universitātes Latviešu valodas institūts.
anta.trumpa@lu.lv

Ungure, Elza – *Mg. hum.*, antropoloģe. Latvijas Nacionālā bibliotēka.
elza.ungure@lnb.lv

Vanags, Pēteris – *Dr. habil. philol.*, valodnieks. Latvijas Universitātes Latviešu valodas institūts.
peteris.vanags@lu.lv

Voitkāns, Arnis – *Mg. phys.*, fiziķis, datorzinātnieks. Latvijas Universitātes Matemātikas un informātikas institūts. *arnis.voitkans@lu.lv*

Authors

Ančevska, Ieva – *Dr. philol.*, folkloriste. Kurzeme Institute of Humanities of the University of Liepāja.
ieva.ancevska1@gmail.com

Andronova, Everita – *Mg. philol.*, linguist. Institute of Mathematics and Computer Science of the

University of Latvia.
everita.andronova@lumii.lv

Apenīte, Mārite – *Mg. philol.*, *Mg. sc. soc.*, National Thesaurus expert. National Library of Latvia.
marite.apenite@lnb.lv

Aužiņa, Ilze – *Dr. philol.*, linguist. Institute of Mathematics and Computer Science of the University of Latvia. *ilze.auzina@lumii.lv*

Baklāne, Anda – *Mg. phil.*, head of digital research services. National Library of Latvia.
anda.baklane@lnb.lv

Bojārs, Uldis – *Dr. sc. comp.*, computer scientist. National Library of Latvia. *uldis.bojars@lnb.lv*

Brīška, Anna – engineer. Rezekne Academy of Technologies.
anna.briska@rta.lv

Darģis, Roberts – *Mg. sc. comp.*, computer scientist. Institute of Mathematics and Computer Science of the University of Latvia.
roberts.dargis@lumii.lv

Daugavietis, Jānis – *Dr. sc. soc.*, sociologist. Institute of Literature, Folklore and Art of the University of Latvia. *janis.daugavietis@lulfmi.lv*

Eglāja-Kristsone, Eva – *Dr. philol.*, literary scholar. Institute of Literature, Folklore and Art of the University of Latvia.
eva.eglaja@lulfmi.lv

Eversone, Madara – *Mg. philol.*, literary scholar. Institute of Literature, Folklore and Art of the University of Latvia. *madara.eversone@lulfmi.lv*

Fridenberga, Anna – *Dr. philol.*, linguist. Latvian Language Institute of the University of Latvia.
anna.fridenberga@lu.lv

Autori / Authors

Garda, Ansis – communication scientist. National Library of Latvia. *ansis.garda@lnb.lv*

Goldberga, Anita – *Mg. philol., Mg. sc. soc.*, information scientist. National Library of Latvia. *anita.goldberga@lnb.lv*

Grasmanis, Mikus – *Mg. sc. comp.*, computer scientist. Institute of Mathematics and Computer Science of the University of Latvia. *mikus.grasmanis@lumii.lv*

Grūzītis, Normunds – *Dr. sc. comp.*, computer scientist. Institute of Mathematics and Computer Science of the University of Latvia. *normunds.gruzitis@lumii.lv*

Juško-Štekele, Angelika – *Dr. philol.*, folklorist. Rezekne Academy of Technologies. *angelika.jusko-stekele@rta.lv*

Kaija, Inga – *Dr. philol.*, linguist. Rīga Stradiņš University. *inga.kaija@rsu.lv*

Karlsona, Agnese – *Mg. art.*, sociologist. Institute of Arts and Cultural Studies of the Latvian Academy of Culture. *agnese.karlsona@lka.edu.lv*

Kārkla, Zita – *Dr. philol.*, literary scholar. Institute of Literature, Folklore and Art of the University of Latvia. *zita.karkla@ulfmi.lv*

Kļavinska, Antra – *Dr. philol.*, linguist. Rezekne Academy of Technologies. *antra.klavinska@rta.lv*

Kreislere, Maira – *Mg. sc. soc.*, information scientist. National Library of Latvia. *maira.kreislere@lnb.lv*

Kristāla, Ance – *Mg. sc. soc.*, sociologist. Latvian Academy of Culture,

Institute of Arts and Cultural Studies. *ance.kristala@lka.edu.lv*

Kunda, Ilona – *Dr. sc. soc.*, sociologist. Latvian Academy of Culture, Institute of Arts and Cultural Studies. *ilona.kunda@lka.edu.lv*

Laime, Sandis – *Dr. philol.*, folklorist. Institute of Literature, Folklore and Art of the University of Latvia. *sandis.laime@ulfmi.lv*

Lasmanis, Viesturs Jūlijs – computer science student. Institute of Mathematics and Computer Science of the University of Latvia. *viesturs.lasmanis@lumii.lv*

Levāne-Petrova, Kristīne – *Mg. philol.*, linguist. Institute of Mathematics and Computer Science of the University of Latvia. *kristine.levane-petrova@lumii.lv*

Ļaksa-Timinska, Ilze – *Mg. philol.*, literary scholar. Institute of Literature, Folklore and Art of the University of Latvia. *ilze.laksa-timinska@ulfmi.lv*

Martena, Sanita – *Dr. philol.*, linguist. Rezekne Academy of Technologies. *sanita.martena@rta.lv*

Matulis, Haralds – *Mg. phil.*, philosopher. Institute of Literature, Folklore and Art of the University of Latvia. *haralds.matulis@ulfmi.lv*

Nau, Nicole – *Dr. habil. phil.*, linguist. Adam Mickiewicz University in Poznań, Faculty of Modern Languages and Literature, and Rezekne Academy of Technologies. *nicole.nau@rta.lv*

Nešpore-Bērzkalne, Gunta – *Mg. philol.*, linguist. Institute of Mathematics and Computer Science of the University of Latvia. *gunta.nespore@lumii.lv*

Oga, Jānis – *Dr. philol.*, literary scholar. Institute of Literature, Folklore and Art of the University of Latvia. *janis.oga@ulfmi.lv*

Pokratniece, Kristīne – *Mg. philol.*, folklorist. Institute of Mathematics and Computer Science of the University of Latvia. *kristine.pokratniece@lumii.lv*

Rapa, Sanda – *Dr. philol.*, linguist. Latvian Language Institute of the University of Latvia. *sanda.rapa@lu.lv*

Rašmane, Anita – *Mg. math., Mg. sc. soc.*, mathematician, information scientist. National Library of Latvia. *anita.rasmane@lnb.lv*

Reinsone, Sanita – *Dr. philol.*, folklorist. Institute of Literature, Folklore and Art of the University of Latvia. *sanita.reinsone@ulfmi.lv*

Rituma, Laura – *Mg. philol.*, linguist. Institute of Mathematics and Computer Science of the University of Latvia. *laura.rituma@lumii.lv*

Saulespurēns, Valdis – *Mg. sc. comp.*, research technology analyst. National Library of Latvia. *valdis.saulespurens@lnb.lv*

Saulīte, Baiba – *Dr. philol.*, linguist. Institute of Mathematics and Computer Science of the University of Latvia. *baiba.valkovska@lumii.lv*

Siliņa-Pinķe, Renāte – *Dr. philol.*, linguist. Latvian Language Institute of the University of Latvia. *renate.silina-pinke@lu.lv*

Skadiņa, Inguna – *Dr. sc. comp.*, computer scientist. Institute of Mathematics and Computer Science of the University of Latvia. *inguna.skadina@lumii.lv*

Autori / Authors

Skrūzmane, Elga – *Mg. philol.*, linguist. Latvian Language Institute of the University of Latvia.
elga.skruzmane@lu.lv

Spektors, Andrejs – *Dr. phys.*, mathematician. Institute of Mathematics and Computer Science of the University of Latvia.
andrejs.spektors@lumii.lv

Stepanovs, Kaspars – *Dr.*, radiologist. Riga East University Hospital.
kstepanovs@gmail.com

Trumpa, Anta – *Dr. philol.*, linguist. Latvian Language Institute of the University of Latvia.
anta.trumpa@lu.lv

Ungure, Elza – *Mg. hum.*, anthropologist. National Library of Latvia.
elza.ungure@lnb.lv

Vanags, Pēteris – *Dr. habil. philol.*, linguist. Latvian Language Institute of the University of Latvia.
peteris.vanags@lu.lv

Voitkāns, Arnis – *Mg. phys.*, physicist, computer scientist. Institute of Mathematics and Computer Science of the University of Latvia.
arnis.voitkans@lu.lv

LETONICA

Redakcijas adrese / Address of the editorial office
Mūkusalas ielā 3, Rīga, LV-1423, Latvija
Tālrunis / Phone: +371 67229017

Galvenais redaktors / Editor-in-chief
ARTIS OSTUPS
artis.ostups@lulfmi.lv

Redakcijas asistents / Assistant
IVARS ŠTEINBERGS

Numura redaktori / Issue editors
SANDIS LAIME,
ANDA BAKLĀNE

Literārā redakcija / Proofreaders
EVELĪNA ZILGALVE,
LAINE KRISTBERGA

Numurs finansēts Valsts pētījumu programmas “Humanitāro zinātņu digitālie resursi” projektā “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (VPP-IZM-DH-2020/1-0001) un izdots ar Liepājas Universitātes un Latvijas Universitātes atbalstu.



**LATVIJAS
UNIVERSITĀTE**



**LIEPĀJAS
UNIVERSITĀTE**



VPP
Valsts pētījumu
programma

ISSN 1407-3110

ISSN 2592-9453

(digitālā versija)

© LU LFMI, 2022

Latvijas digitālo
humanitāro zinātņu
rīki un resursi

Latviešu folkloras
krātuves digitālais
arhīvs *garamantas.lv*

Metadatos balstīta
dienasgrāmatu
teksta korpusa analīze

Liepājas Universitātes
folkloras krājums

Literārās telpas
kartēšana

Ceļojumu kartēšana un
sociālpolitiskās tendences

Latento Dirihlē
sadalījumu modelis
tematu analīzē

LNB atvērtie dati
pētniecībai

Georga Manceļa tekstu
atveide mūsdienu
rakstībā

Latgaliešu valodas
korpus

Mūsdienu latgaliešu
valodas runas korpusa
izveide

Specializēta runas
korpusa un izrunas
vārdnīcas izveide

Latviešu valodas
uzdevumu automātiskā
ģenerēšana

Latviešu valodas
FrameNet korpus

Vārdnīcas digitālā
struktūra

CLARIN valodas
resursi un rīki