

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 ilgstošī un mērķtiecīgi izstrādāti digitāli latviešu valodas resursi, rīki un izmantotas digitālas 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 komunikāciju un praktisku izmantošanu. Matemātisko un statistisko metožu izmantojums joprojām ir nozīmīgs digitālās pētniecības stūrakmens, 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 plašākai sabiedrībai paredzētas digitālas mantojuma kolekcijas un izstādes, savukārt sabiedrības iesaistes kampaņas 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 veltīti 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-Kristsones 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-Kristsones raksts sniedz izsmelošu 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 korpusi, kas veidoti 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 daudzveidīgāku 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 korpusi 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 iespējamus 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 tādus 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 Spektors 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 aplūkotie 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 korpusss 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* korpusi” 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 iespiestām un elektroniskām vārdnīcām, raksturo dažādus elektronisko vārdnīcu strukturēšanas veidus, to priekšrocības un ierobežojumus, kā arī sniedz ieskatu latviešu valodas elektronisko vārdnīcu veidošanas vēsturē un tagadnē.

Daudzveidīgo valodas resursu pārskatu noslēdz Ingūnas Skadiņas, Ilzes Auziņas, Roberta Darģa un Arņa Voitkāna publikācija “CLARIN valodas resursu un rīku pētniecības infrastruktūra humanitārajām un sociālajām zinātnēm”. Autori uzsver starptautiskās valodas digitālo resursu infrastruktūras CLARIN nozīmi un iepazīstina lasītāju ar CLARIN ERIC un CLARIN Latvijas mezglpunkta CLARIN–LV piedāvātajām iespējām. Demonstrējot infrastruktūras funkcionalitātes, kā piemēri aplūkoti CLARIN jau iekļautie Latvijā veidotie digitālie valodas resursi.