

SKAITMENINIŲ TECHNOLOGIJŲ KELIAMĖ IŠŠŪKIAI IR SIŪLOMI BŪDAI JUOS ĮVEIKTI

24-oji Jono Jablonskio konferencija

Rudenimis nuo 1993 m. Lietuvių kalbos instituto ir Vilniaus universiteto Lietuvių kalbos katedros rengiamų Jono Jablonskio konferencijų visuma atskleidžia, kad bene dažniausiai jose rinktasi nagrinėti gana konkrečias norminimo ir vartosenos problemas arba normos ir vartosenos santykį (pvz.: *Kalbos normos ir vartosenos*, 1994; *Nauji kalbos reiškiniai ir jų vertinimas*, 1998; *Normų kaita ir vartosenos*, 1999; *Kalbos normų ir vartosenos lygmenys*, 2000; *Kalbos vartosenos ir tvarkyba*, 2001)¹. Tačiau, ko gero, nebuvo nė vienos konferencijos, kurioje vienaip ar kitaip neiškiltų ir fundamentalesnių klausimų – bendrinės kalbos sampratos, ribų, paskirties, santykio su visuomene, mainymosi besikeičiant aplinkybėmis, išlikimo. O maždaug nuo amžių sandūros, įvairioms kalbą veikiančioms aplinkybėms vis esmingiau keičiantis, tokie klausimai ir tikslingai iškeliami kaip pagrindinis svarstymų objektas (plg.: *Bendrinės kalbos ribos*, 2002; *Bendrinė kalba ir visuomenė*, 2005; *Bendrinė kalba ir kitos kalbos atmainos*, 2006; *Kalbos variantiškumas ir kalbinės nuostatos*, 2009; *Lietuvių kalbos gyvavimas ir atsinaujinimas šiuolaikinėje visuomenėje*, 2013). Konferencijų raida atskleidžia ir kitą polinkį: žvilgsniui kryptant prie pamatinių dalykų, taigi šiuo požiūriu gilėjant, juo kartu aprėpiama vis platesnė perspektyva. Jau keturios konferencijos surengtos kaip tarptautinės², ir tai reiškia, kad jose svarstyta ne tik lietuvių bendrinės kalbos būklė ar samprata, bet ir kitų šalių plėtojamos bendrinės kalbos koncepcijos arba, jei tokios koncepcijos netaikomos, kalbos planavimo, politikos, tvarkybos kryptys.

Šiūmetėje 24-ojoje konferencijoje atsivėrė ypač plataus masto perspektyva – europinė, aprėpianti daugiakalbę Europos Sąjungą ir joje kylančią kalbų įvairovės problemą. Pagal sritį, į kurią susitelkta, Lietuvių kalbos institute vykusią konferenciją **SKAITMENINIAI KALBOS IŠTEKLIAI, JŲ PLĖTROS KRYPTYS IR PANAUDOS GALIMYBĖS** (2017 m. rugsėjo 29 d.) būtų galima sieti ir su dar platesnės aprėpties plotme – globaliąja. Ir nors būdvardžiui *globalus*, *-i* pakeisti siūloma lietuviškų žodžių, tokių kaip *pasaulinis*, *-ė*; *visuotinis*, *-ė*; *didžiulis*, *-ė*

¹ Kai kurių J. Jablonskio konferencijų pranešimų tezes galima rasti Lietuvių kalbos instituto interneto tinklalapyje adresu <http://lki.lt/tarptautin%C4%97-jono-jablonskio-konferencija/#1501063755178-1d439107-84e9>. *Bendrinėje kalboje* (anksčiau – *Kalbos kultūra*) spausdinamos konferencijų apžvalgos.

² Kaip tarptautinės konferencijos buvo organizuojamos 2005, 2009, 2010 ir 2016 metais.

ar *milžiniškas*, $-a^3$, šiemetinius svarstymus tiksliausia įvardyti kelis semantinius atspalvius apimančiu tarptautiniu. Pirma, skaitmeninė erdvė yra ne tik pasaulinė, bet ir visuotinė, milžiniška ar net beribė, juolab kad pranešėjų kaip skaitmeninių technologijų sritis minėti ne tik kompiuteriai, bet ir išmanieji laikrodžiai, telefonai, balsu valdomi dulkių siurbliai, robotai, apskritai dirbtinis intelektas. Antra, žodis *globalus*, $-i$ iškart parodo sąsajas su globalizacija, globalėjimu, o tokių akivaizdžių sąsajų turi ne tik pasirinkta nagrinėti kalbos vartojimo sritis, bet ir konferencijoje keltas fundamentalusis klausimas.

Tą klausimą apibendrintų matematikos mokslų habilituoto daktaro Laimučio Telksnio pateikta formuluotė: „Kaip elgtis mums, keliems milijonams lietuvių, kad neišnyktume daugiau kaip 7 milijardų rašančiųjų ir šnekančiųjų žmonių bei ateinančių išmaniųjų mašinų okeane?“. Tačiau tiek plenariniame posėdyje akademiko pateiktas konkretus ir dalykiškas atsakymas, tiek kituose keturiuose konferencijos posėdžiuose filologų, matematikų, programuotojų ir inžinierių vienas po kito pristatomi jungtiniai projektai rodė, kad bendroji strateginė kryptis puikiai žinoma. Tai kalbos ir jos sandaros diegimas į skaitmeninę erdvę ir aktyvus rūpinimasis, kad ji ten funkcionuotų – tiek rašytiniu, tiek ir žodiniu pavidalu. Strateginė kryptis konferencijoje veikiau ne svarstyta, o vertinta kaip vienintelė mažos šalies kalbos išlikimo skaitmeniniame amžiuje galimybė, vienintelis galimas pasirinkimas. Europarlamentaras Algirdas Saudargas jį palygino su knygnešiams tekusiu pasirinkimu kirilica rašančioje aplinkoje rinktis lietuvišką raštą lotyniškėmis raidėmis, o Vilniaus universiteto Filologijos fakulteto projektų vadovas dr. Audrius Valotka minėjo nuoseklią pasirinkimų grandinę: ką nors užrašyti molyje ar pergamente pasirinkusios tautos išliko, paskui istorijos kelyje išliko tos tautos, kurios Gutenbergo išrastu spaudos presu spausdino tekstus; dabar atėjo neišvengiamo rinkimosi keltis į skaitmeninę erdvę eilė.

Kiek dažniau nei strateginė kryptis konferencijoje minėti sunkumai, kylantys aiškiai suvokiama kryptimi einant, tačiau ir jie nebuvo konferencijos svarstymų pagrindas. Ne vieno kaip problema įvardyta tai, kad technologijų pasaulyje viešpatauja Jungtinių Amerikos Valstijų (toliau – JAV) bendrovės, tokios kaip *Microsoft*, *Amazon* ar *Google*, o jas mažų šalių rinka, juolab jų kalba, mąža tedomina. Todėl keliskart eita prie išvados, kad kalbos vartojimu skaitmeninėje erdvėje ir jos išlikimu reikia rūpintis patiems. Tačiau ne vienas tuo jau besirūpinantis minėjo išteklių, įskaitant žmogiškuosius, stygių. Suprantama, mažai šaliai kalbos technologijų srityje varžytis su rinkos rykliais ar milžinėmis nekart įvardytomis vadintomis JAV privačiojo kapitalo bendrovėmis nėra didelių galimybių, tačiau iš Europos Sąjungos (toliau – ES) situaciją aptarusio

³ Pavyzdžiui, tokie pakaitai nurodomi Valstybinės lietuvių kalbos komisijos teikiamame Konsultacijų banke: <http://www.vlkk.lt/konsultacijos/5531-globalus>.

ispanų mokslininko Rafaelio Riveros pranešimo aiškėjo, kad tų bendrovių viešpatavimas informacinių technologijų pasaulyje kelia problemų ir ES lygiu. Kita vertus, pačioje ES irgi kryptama į anglų kalbą, o tai turėtų būti lietuvių kalbos diegimą ir gyvavimą skaitmeninėje erdvėje ypač apsunkinantis veiksnys.

Tačiau konferencija atvėrė visiškai kitokią perspektyvą. Plenarinį posėdį pranešimu *Gyvoji kalba dirbtiniame prote* pradėjęs Europos Parlamente dirbantis Nepriklausomybės Akto signataras Algirdas Saudargas, paskatintas 60 mokslinių tyrimų centrų 34 šalyse apimančio meistriškumo tinklo META-NET atlikto tyrimo, parodžiusio, kad pagal būklę kalbos technologijų požiūriu lietuvių kalbai tenka žemiausias balas, t. y. ji vertinama „kaip turinti menką technologinę paspirtį arba jos visai neturinti“, o tai reiškia, kad kalbai gresia išnykimas, jau prieš kelerius metus Europos Parlamente inicijavo tyrimą *Kalbų lygybė skaitmeniniame amžiuje*. Tyrimas kaip sąlyga, kad ES būtų pasiekta tikroji integracija, atskleidė ne vienakalbystę, o kaip tik kalbų įvairovę. Vėliau jis papildytas *Gimtosios kalbos projektu*, kuriame siūlomos įvairius lygmenis apimančios strateginės kryptys, kaip ES pasiekti technologinį proveržį, kuris leistų kiekvienam Sąjungos piliečiui paslaugas bendrojoje skaitmeninėje erdvėje gauti ir bendrauti nepriklausomai nuo vartojamos kalbos.

Pristatyti atliktojo tyrimo ir jo pagrindu parengtos studijos atvyko vienas iš tyrimo vykdytojų – konsultacijų bendrovės *iClaves* direktorius Rafaelis Rivera. Jis įvardijo kalbų įvairovės ES keliamą prieštarą: viena vertus, ji, kaip Europos kultūrų įvairovės apraiška, laikoma vienu didžiausių 24 oficialiųjų ir daugiau nei 60 nacionalinių bei regioninių mažumų kalbų turinčios sąjungos turtu, kita vertus, susidarantys kalbos barjerai yra viena pagrindinių kliūčių kurti išties integruotą sąjungą. Dėl jų ypač nukenčia mažesnio išsilavinimo ir vyresni žmonės, mažesniųjų ir regioninių mažumų kalbų atstovai. Taip susidaro kalbinė atskirtis. Didelės yra ekonominės bei socialinės kalbinių barjerų pasekmės: neišnaudojamos visos bendrosios rinkos, laisvo prekių ir darbo jėgos judėjimo, gyventojų dalyvavimo politiniuose procesuose, elektroninės prekybos teikiamos galimybės, o teisę jomis naudotis turintys piliečiai negali ja reikiamai pasinaudoti. Ir ypač aktuali ši problema tampa skaitmeniniame amžiuje. Randantis naujų technologijų, tokių kaip savimokos neuroniniai tinklai, atsiranda būdų kalbos barjerus įveikti, tačiau tam trukdo Europos kalbos technologijų rinkos susiskaidymas, tyrimų nekoordinuotumas, nepakankamas finansavimas, o ir ES politikų darbotvarkėje šiems klausimams neskiriama pakankamai dėmesio. Visos tos priežastys bei problemos ir paaiškina

europinio gimtosios kalbos projekto, kuriame numatomas išsamus institucinį, mokslinį, pramonės, rinkos ir viešojo sektoriaus lygmenis apimančių priemonių kompleksas, radimąsi⁴.

Kalbos technologijų srityje dirbantis Vilniaus universiteto Matematikos ir informatikos instituto profesorius, interneto plėtros Lietuvoje pradininkas, ne tik Lietuvoje išradimais garsėjantis mokslininkas habil. dr. Laimutis Telksnys pranešime *Rašytinė ir sakytinė lietuvių kalba įprastoje ir elektroninėje terpėse*⁵ atsakydamas į paties iškeltą klausimą apie galimybę išlikti milijardiniame okeane, kaip būtinas išlikimo sąlygas formulavo gana konkrečius darbus: kad popieriniuose dokumentuose ir elektroninėje terpėje išliktų taisyklinga lietuvių kalba, reikia rašytinės bei sakytinės kalbos taisyklių ir techninės bei programinės įrangos, transkribatorių, kurie leistų nelietuvių rašytinės kalbos ženklus pavaizduoti lietuviškais rašytinės kalbos ženklais, o sakytinės kalbos garsus pateikti lietuvių sakytinės kalbos ženklais, tinkančiais automatinei lietuvių šnekos žodžių garsų sintezei. Taip būtų užtikrinta, kad tiek su kompiuteriais, tiek su telefonais, kitais kalbančiais išmaniaisiais daiktais ir robotais būtų bendraujama taisyklinga lietuvių rašytine ir sakytine kalba, o Lietuvos elektroninėse duomenų saugyklose nelietuviški asmenvardžiai būtų pateikiami lietuviškų rašto ženklų kodais bei originalo rašmenimis; jų tarimas balsu būtų pateikiamas garso įrašais, padarytais juos tariant lietuvių kalbos fonemų garsais ir garso įrašais originalo kalba.

Paties profesoriaus ir kitų Matematikos ir informatikos instituto specialistų taisyklingos šnekamosios lietuvių kalbos vartojimu skaitmeninėje erdvėje jau gerokai pasirūpinta – tai aiškėjo iš pirmajame posėdyje **ŠNEKOS ATPAŽINIMAS IR SINTEZĖ** skaitomų pranešimų. (O ir visuose kituose konferencijos posėdžiuose maža teminėti sunkumai, veikiau dalykiškai ir konstruktyviai pristatyti dirbami darbai ir vardytos tolesnės užduotys.) Bendradarbiaudama su filologais – Lietuvių kalbos institutu, Vilniaus universiteto Filologijos fakultetu, Vilniaus edukologijos universitetu ir Šiaulių universitetu – habil. dr. L. Telksnio vadovaujama Vilniaus universiteto Matematikos ir informatikos instituto bei Matematikos ir informatikos fakulteto specialistų grupė 2013–2015 m. įgyvendino projektą *Lietuvių šneka valdomos paslaugos (LIEPA)*. Jo rezultatas – septynios visiems prieinamos paslaugos, teikiamos taisyklinga sakytine lietuvių kalba: ląstelės sandarą mokyklose padedantis aiškintis *Pažintuvas*, su UNESCO paveldo objektais supažindinantis *Ieškotuvą*, balso komandomis kompiuterį ir paiešką leidžiantys valdyti

⁴ Studija *Language equality in the digital age. Towards a Human Language Project* ir jos santrauka anglų kalba prieinamos viešai. Jos pateikiamos Europos Parlamento interneto svetainėje ([http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU\(2017\)598621](http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2017)598621)), tačiau tekstą atgaminti ar versti nekomerciniais tikslais galima tik gavus Europos Parlamento sutikimą.

⁵ Pagal pranešimą parengtą straipsnį žr. *Bendrinės kalbos* šio tomo „Apžvalgų“ skyriuje.

Valdytuvas ir *Naršytuvas*, taisyklingos lietuvių kalbos mokantys *Pagalbininkas* ir *Tartuvas* bei akliesiems skirtas kompiuterinis sintezatorius, skaitantis tekstą kompiuterio ekrane. Kartu sukurtos infrastruktūros paslaugos: elektroninio teksto skaitytuvas, interneto puslapių atidarymo balsu programa, lietuvių šnekos atpažinimo variklis, garsynas, balso įrašų laboratorija⁶. Atrodytų, nuveikta daug, tačiau L. Telksnys, skaitydamas kartu su to paties instituto Projektų skyriaus vyresnioju specialistu Gediminu Navicku parengtą pranešimą *Lietuvių šneka valdomos paslaugos. Padėtis. Perspektyvos*, šiuos darbus įvardijo tik kaip durų pravėrimą lietuvių šnekamajai kalbai patekti į elektroninę erdvę. Kaip projekto rezultatus pranešėjas buvo linkęs minėti ne tiek sukurtąsias paslaugas, kiek išryškėjusią informatikų bei filologų turimų žinių „sulydymo“ svarbą, taip pat pavyzdžio reikšmę – gamybininkams ir paslaugų teikėjams parodomos lietuvių šneka valdomų paslaugų panaudos galimybės. Šiuo metu kylančiu gana skubiu uždaviniu, tenkančiu jau ne tiek kalbos technologijų kūrėjams bei plėtojams, pasaulines technologijų srities kryptis išvelgiantis mokslininkas laikė robotų įtraukimą į mokymo programas, bet prieš tai – dar ir mokytojų parengimą.

Tik kaip vartų pravėrimą projekto *LIEPA* darbus įvardijo ir Vilniaus universiteto projektų vadovas dr. Audrius Valotka. Pranešime *Tolyn nuo Gutenbergo spaudos preso: lietuvių šneka naujausiose technologijose* (bendraautoris G. Navickas) jis teigė, kad „šiuo metu lietuvių šneka informacinėse technologijose yra panašioje situacijoje, kaip prieš 30 metų, kai buvo sprendžiami lietuviškų diakritinių ženklų kompiuterijoje naudojimo klausimai“. Tokia būklė skatina imtis tolesnių darbų – projekto *LIEPA 2*. Šio jau pradėto vykdyti projekto rezultatas turėtų būti infrastruktūriniai sprendimai: 1 000 valandų lietuvių šnekos garsynas, 3 000 žodžių lietuvių šnekos atpažintuvas, tobulesnis lietuvių šnekos sintezatorius. Bus sukurta ir naujų paslaugų: vaikų sprendimų priėmimo gebėjimus ugdantis humanoidinis robotas, asmens kontaktų mobiliajame telefone valdytuvas balsu, taksi iškvietuvas, interneto naujienų skaitytuvas, tarpkalbinis komunikatorius, apimantis lietuvių ir kinų kalbas. Tačiau kaip svarbiausias naujojo projekto rezultatas numatyta galimybė natūralia šneka bendrauti su daiktai (mobiliaisiais telefonais, planšetėmis, išmaniaisiais laikrodžiais, robotais), duoti komandas žmogaus balsu ir suprasti jų atsakymus.

Dar vienas Matematikos ir informatikos instituto atstovas – docentas Pijus Kasparaitis pranešime *Lietuviško balso kompiuterinės sintezės dabartis ir perspektyvos* (bendra autoriai Vilniaus universiteto Matematikos ir informatikos fakulteto asistentas dr. Gintaras Skersys ir lektorė

⁶ Jos išsamiau aprašomos Integruotų lietuvių kalbos ir raštijos išteklių svetainėje *Raštija.lt*: <https://liepa.rastija.lt>.

Margarita Beniušė⁷, pasitelkęs filmuotą medžiagą ir interneto svetaines, iliustravo projekto *LIEPA* metu sukurto sintezatoriaus panaudos galimybes: sintezatorius panaudotas per 2014 m. teatrų festivalį *Sirenos* rodytą spektaklį *Remote Vilnius* (žiūrovai, atliekantys kartu ir aktorių vaidmenį, vaikščiojo Vilniaus gatvėmis per ausines klausydami sintetinių balsų); dalies internetinių laikraščių (*Izinius.lt*, *ukininkopatarejas.lt*, *vilnius.lt*, *delfi.lt*) straipsnius galima ne tik perskaityti, bet ir išklausyti; Vilkaviškio autobusų stotyje sintetiniu balsu skaitomi pranešimai keleiviams; teksto įgarsinimo paslauga *RoboBraille* leidžia bet kokius tekstinius dokumentus automatiškai paversti į garso failus; tikrove yra tapęs ir lietuviškai šnekantis robotas – humanoidas *NAO* kvietė aplankyti Lietuvos standą Briuselyje surengtame ES šalių prisistatymo renginyje, deklamavo Kristijono Donelaičio *Metus* Seime vykusioje konferencijoje *Ar kalbės robotai lietuviškai?* Vardydamas būsimas balso sintezės panaudos galimybes, kurias įgyvendinant jau žengta tam tikrų žingsnių, P. Kasparaitis minėjo straipsnių įgarsinimą *Linux* pagrindu veikiančiose svetainėse, virtualiuosius asistentus, užkandžių ar gėrimų parke galėsiantį pasiūlyti mobilųjų pardavimų automatą *MIO mobile kiosk*.

Kituose trijuose posėdžiuose pristatyti darbai, daugiausia susiję su rašytinės kalbos išsaugojimo pastangomis, kėlė įspūdį, kad nuo žodynų, tekstynų, duomenų bazių, kartotekų perkėlimo į skaitmeninę erdvę vis labiau einama prie tokio kalbos skaitmeninimo, kuris leistų kalbos ištekliams naudotis ne tik kalbos vartotojams bei tyrėjams, bet ir kalbos technologijų programoms, dirbtiniam intelektui. Tokią strateginę kryptį, aktualėjančią dirbtiniam intelektui darant vis didesnę pažangą, tiek straipsniuose spaudoje⁸, tiek ir šioje konferencijoje pabrėžė kalbos strategas A. Saudargas. Konferencijai pateiktose tezėse ji nusakyta taip: „dirbtinio intelekto raida konverguoja į hibridinį pavidalą, kuriame neuroniniai tinklai atitinka nesąmoningus smegenų mechanizmus, o sąmoningą proto veiklą modeliuoja tradicinis, vadinamasis simbolinis dirbtinis intelektas. Tas pats tinka ir kalbos technologijoms, jei jas suprasime kaip kalbos mechanizmų žmogaus smegenyse (gyvosios kalbos) modelius. Todėl kiekviena kalbinė bendruomenė privalo rūpintis, kad kalbos technologijos, atitinkančios simbolinį dirbtinį intelektą (vadinamosios „taisyklėmis grįstos“ programos) išsamiai ir tiksliai atspindėtų visą gimtosios kalbos sandarą, o neuroninių tinklų savimokai būtų sukurta turininga gimtosios kalbos (ir gimtosios kultūros) aplinka“.

⁷ Pagal pranešimą parengtą straipsnį žr. *Bendrinės kalbos* šio tomo „Apžvalgų“ skyriuje.

⁸ Pavyzdžiui, straipsnyje *Fulgura plango*, paskelbtame dienraštyje *Bernardinai.lt* (<http://www.bernardinai.lt/straipsnis/2012-12-05-algirdas-saudargas-fulgura-plango/91787>), straipsnyje *Gimtoji kalba: ko Google neišspręs*, paskelbtame *Lietuvos žiniuose* (<http://Izinius.lt/Izinius/Mokslas-ir-svietimas/gimtoji-kalba-ko-google-neisspres/246316>).

Antrajame konferencijos posėdyje **SKAITMENINIAI LIETUVIŲ KALBOS IŠTEKLIAI**, pristatant Lietuvių kalbos instituto vykdomą naują projektą, irgi teigta, kad kalbos ištekliai, kylant naujiems poreikiams, iššūkiams ir lūkesčiams, turi kisti: turi rasti naujų skaitmeninių kalbos išteklių ideologijų ir koncepcijų, kurtis prasmių koridoriai, sąvokų inžinerijos, informacijos paieškos galimybių įvairovė ir kt. Bendrųjų reikalų infrastruktūros skyriaus vadovės Deimantės Budriūnaitės perskaitytame pranešime *E. kalba – skaitmeninių kalbos išteklių naudojimo(si) inovacija* pristatytas naujasis instituto projektas, jos vykdomas kartu su direktore Jolanta Zabarskaite ir *Netcode* projektų vadovu Skirmantu Šermukšniu, iš tikrųjų yra ankstesniojo kalbos skaitmeninimo projekto tęsė. Tuokart buvo suskaitmeninta 11 vienakalbių ir dvikalbių žodynų bei 5 kartotekos ir taip sukurta *Lietuvių kalbos išteklių informacinė sistema*⁹, o nauju projektu jau siekiama ne tik integruoti naujus išteklius, bet ir kurti technologijas, sprendimus, įrankius išteklių semantinei paieškai atlikti, jos rezultatams vizualizuoti, technologijų pagrindu kurti paslaugas vartotojams *Paieška žodžių tinkle*, *E. rinkodara*, *E. sąvokos* ir *E. patarimai*. Numatoma galimybė išteklius pritaikyti rinkodaros tikslams, teikiant viešąsias ir privačias elektronines paslaugas, taip pat užtikrinti didesnę žodžių darybos įrankių ir kalbos patarimų prieinamumą internete, kurti galimybes lietuvių kalbos žodžių tinklą susieti su daugiau kaip 200 kitų kalbų žodžių tinklais.

Rūpinimąsi į internetą jau įdiegtų lietuvių kalbos išteklių naudojimu rodė ir dvi Lietuvių kalbos instituto atstovų pristatytos duomenų bazės, kurių viena integruota į Lietuvių kalbos išteklių informacinę sistemą, o kitą rengiamasi integruoti. Lietuvių kalbos naujažodžių duomenyno iniciatorė ir kūrėja Lietuvių kalbos instituto Bendrinės kalbos tyrimų centro vyriausioji mokslo darbuotoja dr. Rita Miliūnaitė kuriamą bazę apibūdino kaip fiksuoti ir leksikografiškai aprašyti XXI a. pradžios lietuvių kalbos naujažodžiams skirtą internetinį žinyną¹⁰, tačiau jos pranešimas atskleidė, kad dabartiniu jo kūrimo etapu bene labiausiai rūpinamas dalykas yra šio duomenyno pravartumas vartotojams. To siekiant, pritaikyta paieška pagal antraštinį žodį, naujažodžio kilmę, originalo formą, rašybos variantus, vartojimo sritį, kai kuriuos ypatingesnius požymius, pagal naujažodį pavartojusio žymesnio žmogaus ar jo autoriaus pavardę, o turint galvoje atskiras tikslines vartotojų grupes, kuriama išplėstinė paieškos sistema. Susidomėjimas tiek duomenynu, tiek naujažodžiais didelis, taigi kaip problema kol kas minėta tik tai, kad vartotojų atsiunčiamų duomenų nespėjama apdoroti. Kita vertus, siuntinių gausa paneigia du visuomenėje gajus mitus – kad lietuvių kalba miršta ir kad žodžius kuria kalbininkai. Vis dėlto

⁹ *Lietuvių kalbos išteklių informacinė sistema* pateikiama adresu <http://lkiis.lki.lt/home>.

¹⁰ Į *Lietuvių kalbos naujažodžių duomenyną* patenkama adresu <http://naujazodziai.lki.lt>.

dr. Daiva Murmulaitytė, taip pat dirbanti Bendrinės kalbos tyrimų centre, pranešime *Naujažodžių darybos tyrimų perspektyvos* (Lietuvių kalbos naujažodžių duomenyno *atvejis*) siūlė pasirūpinti ir didesniu šio duomenyno pravartumu kalbos tyrėjams. Tam reiktų jį atitinkamai pritaikyti. Siūlyta naujažodžius žymėti indeksais – jais fiksuoti darinio pamatinio žodžio kalbos dalį, darybos pamato pakitimus, analogijos, vertimo vaidmenį darant naujažodį, netipinės darybos apraiškas ir kt. Pranešėjos pastebėjimu, mėginimas tai padaryti atskleidė tokių naujažodžių darybos reiškinių, kurie yra maža tetirti ar net netirti iš viso.

Vardydamas kito duomenyno – Lietuvos vietovardžių geoinformacinės duomenų bazės¹¹ sudarymo tikslus, Lietuvių kalbos instituto Baltų kalbų ir vardyno tyrimų centro vyresnysis mokslo darbuotojas dr. Laimutis Bilkis irgi minėjo tiek praktinius tikslus, t. y. taisyklingos vartosenos mokymą, tiek ir kalbos tyrimus (taip pat ir kalbos išsaugojimą bei sklaidą). Kad šia baze orientuojamasi ir į kalbos vartotojus, ir į tyrėjus, rodo paieškos ir informacijos pateikimo galimybės: informacijos apie vietovardžius galima ieškoti pagal vardą, giminę, skaičių, kirčiuotę, darybos būdą, kilmę pagal pamatinio žodžio priklausymą kalbai ir kilmę pagal pamatinio žodžio priklausymą leksinei grupei, o vietovardžių analizės srityje pateikiamas vietovardžių visų linksnių kirčiavimas, kilmės bei darybos aiškinimai. Į bazę įkelta apie 25 000 vienetų vietovardžių. Jau pasirūpinta ir kai kuriomis sąsajomis: integruojant bazę į *Lietuvių kalbos išteklių informacinę sistemą*, sukurtos nuorodos į kitą vietovardį, iš kurio ieškomasis kilęs, į *Lietuvių pavardžių duomenų bazėje* esantį asmenvardį, iš kurio vietovardis kilęs, į istorinius to vietovardžio užrašymus *Istorinių vietovardžių duomenų bazėje*. Dar numatomos sąsajos su bendrine ar tarminė leksika, ypač semantinės, darybinės ir lokalinės, tačiau paminėta keletas problemų – ne tik vietovardžių gausa, bet ir finansavimo fragmentiškumas.

Akivaizdų poslinkį nuo kalbos duomenų kaupimo prie jų panaudojimo paieškų rodė ir tekstynų lingvistikos atstovų dirbami darbai. Trečiajame konferencijos posėdyje **TEKSTYNŲ LINGVISTIKA** šiuos darbus Lietuvoje daugiausia dirbantys Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centro mokslininkai kartu su Baltijos pažangių technologijų instituto atstovais pristatė du grupinius projektus, atskleidžiančius siekį tekstynus anotuoti ir taip juos panaudoti kalbos tyrimams ir kitiems kalbos ištekliams kurti. Pirmosios tekstynų lingvistikos atstovų grupės – Erikos Rimkutės, Agnės Bielinskienės, Loïc Boizou ir Andriaus Utkos – parengtame pranešime¹² pristatytas morfologiškai anotuotas tekstynas *MATAS* ir sintaksiškai anotuotas tekstynas *ALKSNIS*. Tekstynas *MATAS* parengtas 1 mln. žodžių tekstyno pagrindu,

¹¹ Ši bazė pasiekama ir adresu <http://lkiis.lki.lt/lietuvos-vietovardziu-geoinformacine-duomenu-baze>.

¹² Pagal pranešimą parengtą straipsnį žr. *Bendrinės kalbos* šio tomo „Kalbos technologijų“ skyriuje.

pritaikius statistinius modelius, jam anotuoti naudotas Kompiuterinės lingvistikos centro parengtas morfologinis anotatorius. *ALKSNIO* anotavimas grįstas automatinio morfologinio ir sintaksinio anotavimo principais, taikytas sintaksinių priklausomybių modelis. Antrosios mokslininkų grupės – Erikos Rimkutės, Agnės Bielinškienės, Loïc Boizou, Ievos Bumbulienės, Jolantos Kovalevskaitės, Tomo Krilavičiaus, Justinos Mandravickaitės ir Lauros Vilkaitės – vykdomas projektas atskleidė, kad, anotavus *Delfi.lt* tekstyną, pereinama prie dabartinės rašytinės lietuvių kalbos pastoviųjų žodžių junginių tyrimo metodikos kūrimo, lietuvių kalbos kolokacijų žodyno rengimo. Tekstyno pagrindu sudaromoje pastoviųjų žodžių junginių duomenų bazėje apie junginius ketinama pateikti gramatinę, leksinę informaciją, vartosenos dažnį, teksto rubriką, konkordanso pavyzdžius ir pan. Šiuo metu bazėje yra apie 2 700 lemų ir daugiau nei 35 000 kaitybinių formų.

Praktines tekstynų panaudojimo galimybes atskleidė ir Vilniaus universiteto Lietuvių kalbos katedros dėstytojų Gintarės Judžentytės bei dr. Vilmos Zubaitienės parengtas pranešimas *Tekstynais paremti akademinų frazių tyrimai: formalioji struktūra ir semantika*. Apžvelgus ligtolinius bandymus tekstynų pagrindu tirti mokslo kalbą, pristatytas ir pačių autorių vykdomas projektas, kuriuo, remiantis Vilniaus universitete tam tikslui kuriamu Studentų rašto darbų tekstynu, siekiama iširti akademinės kalbos frazių struktūrą bei semantiką. Naudojantis tekstyno medžiaga ir atlikto tyrimo rezultatais, bus sudaromas akademinės kalbos frazių sąvadas, pravartus įvairių studijų pakopų ir specialybių studentams.

Ketvirtasis konferencijos posėdis **AUTOMATIZUOTA KALBOS SANDAROS IR TEKSTŲ ANALIZĖ** atskleidė dar sudėtingesnes naujų technologijų panaudojimo kalbos analizei galimybes. Vilniaus universiteto Taikomųjų mokslų instituto inžinierius Virginijus Dadurkevičius, pranešime *Lietuvių kalbos gramatika skaitmeniniame atvirojo kodo pasaulyje*¹³ apžvelgęs Lietuvių kalbos instituto programuotojo Vytauto Zinkevičiaus nuveiktus morfologinės analizės srities darbus, beveik du dešimtmečius buvusius nepamainoma ir vienintele moderna lietuvių kalbos tyrimo ir kompiuterinių taikymų priemone, pristatė ir tam tikrą slinktį kompiuterinės morfologijos srityje, kurią leido pasiekti *Hunspell* platformos panaudojimas. Kalbos morfologiją platformoje aprašo du tekstiniai failai, abiejų sudarymo pagrindas – *Dabartinės lietuvių kalbos gramatika* ir Vilniaus universiteto, Vytauto Didžiojo universiteto bei Lietuvos Respublikos Seimo tekstynai. Pranešėjo teigimu, iš viso užrašyta maždaug 5 000 adresuojamų taisyklių grupių, atrinkta per 170 000 lemų. Vykdamas atranką, apsiribota tik taisyklinga šiuolaikine lietuvių kalba. Pasak pranešėjo, šis morfologinis analizatorius teisingai

¹³ Pagal pranešimą parengtą straipsnį žr. *Bendrinės kalbos* šio tomo „Kalbos technologijų“ skyriuje.

interpretuoja vidutiniškai 99 iš 100 žodžių. Naujasis morfologinės analizės būdas sėkmingai pritaikytas Vytauto Didžiojo universiteto sukurtoje *Lietuvių kalbos sintaksinės ir semantinės analizės informacinėje sistemoje*¹⁴, *Teisės aktų registre*¹⁵, o ypač sėkmingai – *Solr/Lucene* dokumentų indeksavimo ir paieškos sistemoje (anot pranešėjo, pavyko ne tik tiksliai indeksuoti bei surasti visas norimas žodžių formas, bet ir šimtus kartų pagreitinti indeksavimo procesą). Tolesniu konkrečiu darbu laikomas dviejų žodyno versijų – norminės, tinkančios rašybai tikrinti, ir norminės bei nenorminės kalbos versijos, tinkančios taikomiesiems gyvosios kalbos analizės uždaviniams spręsti, – parengimas.

Senųjų raštų tyrimams skaitmeninės eros atveriamas galimybes atskleidė Lietuvių kalbos instituto Raštijos paveldo tyrimų centro darbuotojas dr. Mindaugas Šinkūnas pranešime *Kas nuo ko nusirašė? Biblijos vertimų istorijos tyrimo automatizavimas ir vizualizacija*. Jis pristatė jau išbandytą metodiką, kuri leidžia palengvinti daugiausia sunkumų senųjų raštų tyrėjams keliantį citatų panašumo Biblijos vertimuose vertinimą. Pranešėjo teigimu, jos perspektyvumą rodo tai, kad taikant programinę analizę atlikto Bretkūno *Postilės* (1591) bibliinių eilučių palyginimo rezultatai atitinka ankstesnių tyrėjų kitais metodais gautas išvadas. Svarbiausiu darbo rezultatu pranešėjo laikyta priemonė, parodanti orientacinį šaltinių ar jų dalių panašumo santykį. Ji atveria galimybes tolesniems, išsamesniems tyrimams.

Žinoma, konferencijoje neliko nepaliesta ir nemaža nesėkmių patyrusi kalbos technologijų sritis – automatizuotas vertimas, juolab kad jai pirmiausia taikomas terminas *kalbos technologijos*. Pranešime *Mašininis vertimas lietuvių kalbai*¹⁶ paetapiui apžvelgęs visus mašininio pasiekimus ir nesėkmes (šaltojo karo metu besiradusį pažodinį vertimą, taisyklėmis grindžiamą vertimą, statistinius vertimo metodus), Vilniaus universiteto Taikomųjų mokslų instituto inžinierius Danielius Algirdas Ralys darė išvadą, kad šios srities pasiekimai neaplenkia ir Lietuvos. Atsiradus neuroniniams tinklams, konstruojantiems transformavimo taisykles, naujo proveržio viltis automatizuoto vertimo srityje siejama su jais. Atkreipus dėmesį į tai, kad naujausios neuroninio mašininio vertimo sistemos verčia geriau už vidutinišką vertėją, pristatyti Vilniaus universiteto planai 2018 m. imtis naujos kartos neuroninio vertimo anglų, lietuvių, lenkų, prancūzų, rusų ir vokiečių kalbomis projekto. Vis dėlto pastebėta, kad reikia stengtis ne tik kopijuoti tai, kas išrasta, bet ir kurti sava, nauja.

Lietuvių kalbos institute vykdomas projektas atskleidė, kad tokių pastangų išties esama. Kad Lietuvai svarbu kurti alternatyvius automatinio vertimo variantus, Lietuvių kalbos instituto

¹⁴ *Lietuvių kalbos sintaksinės ir semantinės analizės informacinės sistemos* adresas – <https://semantika.lt>.

¹⁵ Teisės aktų registro adresas – <https://www.e-tar.lt/portal/index.html>.

¹⁶ Pagal pranešimą parengtą straipsnį žr. *Bendrinės kalbos* šio tomo „Kalbos technologijų“ skyriuje.

Bendrinės kalbos tyrimų centro mokslo darbuotoja dr. Daiva Šveikauskienė ir programuotojas Vytautas Šveikauskas nusprendė remdamiesi 2017 m. *Tildės* duomenimis, rodančiais, kad statistinius metodus naudojančios *Google* vertimų sistemos pateikiamų vertimų į lietuvių kalbą (ir iš jos) kokybė yra prastesnė nei latvių ar estų kalbų. Alternatyvių vertimo variantų perspektyva autorių siejama su jų kuriama lietuvių kalbos skaitmenine gramatika. Pasak pranešėjų, šiuo metu parengtas tik bandomasis gramatikos pavyzdys, teapimantis keletą žodžių, bet ir jis rodo, kad vertimo kokybė kur kas geresnė. Iš dviejų lygmenų – abstrakčiojo (visoms kalboms bendros formos) ir konkrečiojo (konkrečios kalbos specifinius bruožus atspindinčio lygmens) – sudaryta skaitmeninė gramatika leidžia įveikti bene esminę automatinio vertimo problemą – reikšmės perteikimo. Sakinį vertimo metu ji leidžia transformuoti į abstraktų semantinį pavidalą, iš kurio generuojamas kitos kalbos sakinytis remiantis tik tos kalbos, į kurią jis yra verčiamas, savybėmis, visiškai neatsižvelgiant į kalbos, iš kurios jis buvo verstas, morfologiją, o tai ir sudaro sąlygas gauti daug tikslesnius reikšmės perteikimo rezultatus. Beje, dabartinis bendrovės *Google* pateikiamas vertimas į lietuvių kalbą minėtas kone kas antro konferencijos dalyvio – tiek kaip kalbos technologijos panaudojimo nesėkmė, tiek kaip grėsmė kalbai, tiek kaip pavyzdys, akivaizdžiai rodantis, kad kalbos technologijomis reikia rūpintis patiems.

Šiometinę Jono Jablonskio konferenciją apibendrinant, savaiame kyla keletas tarpusavyje gana susijusių išvadų. Pirmiausia, naujųjų technologijų amžius atveria neįtikėtinų galimybių kalbos tyrimams – tai revoliuciją kalbotyroje padarę tekstynai ir kitokios kompiuterinės duomenų saugos, technologijų atveriamos vis tobulesnės paieškos, apibendrinimo, statistikos pateikimo galimybės, dar labiau tobulėjančios plėtojantis dirbtiniam intelektui. Tačiau kalbos gyvybei naujosios technologijos kelia ypač didelių iššūkių, ir grėsmė kyla ne tik bendrinei, bet ir visai lietuvių kalbai. Vis dėlto būtent jos siūlo būdų, kaip iššūkius įveikti, ir ypač palankias galimybes išlikti mažos šalies kalbai sudaro naujoji europinė iniciatyva. Tačiau tiek ji, tiek skaitmeninių technologijų plėtra išryškina kaip niekad aktualų pastangų telkimo ir sintezės poreikį. Jau ir šioje konferencijoje pristatyti projektai rodo, kad tokiomis sąlygomis tiek kalbos gyvavimo užtikrinimas, tiek ir jos tyrimas tampa vienu kalbos specialistų žiniomis ir pajėgomis nebeįmanomas įgyvendinti uždavinys – projektai filologų vykdyti su informatikais, matematikais, programuotojais. Dirbtinio intelekto plėtra, kaip ne kartą yra nurodęs A. Saudargas, kelia jau ne tik skirtingų sričių specialistų pastangų telkimo, bet ir tarpkrypčio mokslo, besiformuojančio kalbos, kompiuterijos, kognityvinių mokslų sandūroje, kūrimo poreikį¹⁷. O ES

¹⁷ Pavyzdžiui, straipsnyje *Lietuvių kalba privalo tapti raštinga ir skaitmeninėje epochoje*, išspausdintame interneto dienrašyje *Bernardinai.lt* (<http://www.bernardinai.lt/straipsnis/2013-01-25-algirdas-saudargas-lietuviu-kalba-privalo-tapti-rastinga-ir-skaitmenineje-epochoje/94233>).

gimtosios kalbos projektas atskleidžia ir visos valstybės pastangų – tiek viešojo, tiek privačiojo sektorių, politikų, mokslininkų, technologų – sutelkimo būtinybę.

Tik iš pirmo žvilgsnio šiemetinės konferencijos tema galėjo pasirodyti su bendrinės kalbos kūrėjo vardu teturinti maža ką bendra. Tarsi užbėgdama už akių galinčiam kilti sąsajos nebuvimo įspūdžiui, viena iš konferencijos organizatorių dr. Rita Miliūnaitė vieną sąsają įvardijo dar tardama renginio atidarymo žodį: nūdien, kalbą keliant į skaitmeninę erdvę, taip pat kaip ir J. Jablonskio laikais, siekiama kurti universalų, patogų, visuomenės poreikius tenkinantį kalbos variantą. Viena kita sąsaja įžodinta ir kitų konferencijos dalyvių: P. Kasparaitis, reiškdamas viltį, kad projektas *LIEPA* taps rūpinimosi lietuvių kalba sinonimu, pridūrė, kad J. Jablonskis tam pritartų, o D. A. Ralys bendrinės kalbos kūrėją prisiminė kaip gramatikos taisyklių, nuo kurių žengta prie automatizuotos morfologinės analizės, kūrimo pradininką. Tačiau pati konferencija sąsajų bei siekių bendrumo atskleidė kur kas daugiau, ir pirmiausia – rūpinimąsi lietuvių kalbos (iš)likimu, taisyklingos kalbos vartojimu, pasauliui ir kalbą veikiančioms aplinkybėms neįtikėtinai kintant.

Gauta 2017 10 22

RAMUNĖ VASKELAITĖ

ramvasram@gmail.com