

GRĖSMĖS KALBAI: IŠŠŪKIS KALBOS TECHNOLOGIJOMS

Gyvename skaitmeninės revoliucijos, turinčios didžiulį poveikį bendravimui ir visuomenės raidai, metu. Naujaisi skaitmeninės informacijos pateikimo ir bendravimo technologijų išradimai kartais prilyginami Johaneso Gutenbergio spausdinimo mašinos išradimui. Ką ši analogija gali byloti apie Europos informacinės visuomenės ir ypač apie mūsų kalbų ateitį?

Esame skaitmeninės revoliucijos liudininkai.
Ši revoliucija prilygsta Gutenbergio
spausdinimo mašinos išradimui.

Gutenbergio išradimas lėmė svarbius proveržius komunikacijos ir žinių mainų srityje, tokius kaip Martino Lutherio atliktą Biblijos vertimą į gimtąją kalbą ir kt. Vėlesniais amžiais buvo sukurta kultūrinių metodologijų, lengvinančių kalbos apdorojimą ir žinių mainus:

- Ortografinis ir gramatinis labiausiai paplitusių kalbų standartizavimas suteikė galimybių sparčiau plisti naujoms mokslinėms ir intelektinėms idėjoms.
- Bendrinių kalbų susiformavimas suteikė piliečiams galimybių bendrauti tam tikruose (dažniausiai politinių valstybių sienų apibrėžtuose) plotuose.
- Kalbų mokymas ir vertimai iš vienos kalbos į kitą suteikė galimybių keisti informaciją skirtingomis kalbomis.
- Redagavimo ir bibliografinės gairės užtikrino spausdintos informacijos kokybę bei prieinamumą.

- Skirtingų žiniasklaidos priemonių – laikraščių, radijo, televizijos, knygų ir kitokių – atsiradimas tenkino skirtingus komunikavimo poreikius.

Per pastarąjį dvidešimtmetį informacinės technologijos padėjo automatizuoti ir palengvinti daugybę procesų:

- Kompiuterinė programinė įranga pakeitė spausdinimą mašinėle ir rankinį tekstų rinkimą.
- Programa *Microsoft PowerPoint* pakeitė projektoriumi rodomas skaidres.
- Elektroniniu paštu siųsti ir gauti dokumentus galime sparčiau nei faksu.
- Programa *Skype* teikia galimybę pigiai skambinti internetu ir rengti virtualius susitikimus.
- Garso ir vaizdo kodavimo formatai palengvina daugialypės terpės turinio keitimą.
- Paieškos sistemos teikia galimybių naudojant esminius žodžius pasiekti reikiamus tinklalapius.
- Internetinės paslaugos, tokios kaip *Google Translate*, teikia galimybių sparčiai, nors ir ne visada tiksliai, versti tekstus.
- Socialinės terpės, tokios kaip *Facebook*, *Twitter* ir *Google+*, lengvina bendravimą, bendradarbiavimą ir keitimąsi informacija.

Nors tokios priemonės ir programos yra naudingos, kol kas jos nepajėgia išlaikyti visapusiškai tvarios daugialybės Europos visuomenės, kurioje galėtų vykti netrikdoma informacijos ir prekių sklaida.

2.1 KALBŲ BARJERAI – KLIUVINYS EUROPOS INFORMACINEI VISUOMENEI

Negalime tiksliai numatyti, kokia bus ateities informacinė visuomenė. Tačiau tikėtina, kad komunikacinių technologijų revoliucija skirtingomis kalbomis kalbantiems žmonėms teikia naujų būdų suartėti. Dėl to žmonės patiria spaudimą mokytis naujų kalbų, o programinės įrangos kūrėjai – kurti naujus technologinius sprendinius, kurie užtikrintų susikalbėjimą ir prieigą prie bendrų žinių. Dėl naujų medijų rūšių pasaulinėje ekonominėje ir informacinėje erdvėje sąveikauja vis daugiau kalbų ir jomis kalbančių žmonių informacijos.

Pasaulinėje ekonominėje ir informacinėje erdvėje sąveikauja vis daugiau kalbų ir jomis kalbančių žmonių informacijos.

Šiuo metu išpopuliarėjusios bendrauti skirtos priemonės (*Wikipedia, Facebook, Twitter, YouTube* ir visai neseniai – *Google+*) yra vos ledkalnio viršūnė.

Šiandien galime akimirksniu iš kito pasaulio krašto parsisiųsti keletą gigabaitų teksto ir tik po to pamatyti, kad jis parašytas kalba, kurios nesuprantame. Anot vienos iš Europos Komisijos pastarojo laikotarpio ataskaitų, 57 proc. interneto naudotojų Europoje perka prekes ir paslaugas ne savo gimtąja kalba (anglų kalba yra labiausiai paplitusi užsienio kalba, po jos – prancūzų, vokiečių ir ispanų kalbos). 55 proc. naudotojų skaito internete pateikiamą informaciją užsienio kalbomis ir vos 35 proc. užsienio kalbomis rašo elektroninius laiškus arba komentarus interneto tinklalapiuose [2]. Prieš kelerius metus anglų kalba gal ir buvo interneto *lingua franca* – didžioji dalis internete pateikiamos informacijos buvo ja parašyta – tačiau dabar padėtis radikaliai pasikeitė. Internetą užplūdo informacija kitomis Europos, taip pat Azijos ir Vidurio Rytų kalbomis.

Neįtikėtina, bet atrodo, kad ši skaitmeninė takoskyra dėl kalbų barjerų nesulaukė labai daug visuomenės dėmesio. Tačiau ji kelia itin aktualų klausimą: kurios Europos kalbos klestės į tinklą sujungtoje informacinėje ir žinių visuomenėje, o kurioms lemta išnykti?

2.2 GRĖSMĖ KALBOMS

Nors spausdinimo mašina ir palengvino keitimąsi informacija, tai lėmė kai kurių Europos kalbų išnykimą. Spaudinių regioninėmis ir mažumos kalbomis buvo mažai, o tokios kalbos kaip kornų ir dalmatų išvis neturėjo raštijos ir tai labai apribojo jų vartojimo sritis. Ar ir interneto poveikis mūsų kalboms bus toks pats?

Europoje kalbama apie 80 kalbų. Jos yra vienas iš gausiausių ir svarbiausių regiono kultūros turtų bei esminė Europos unikalaus socialinio modelio dalis [3]. Nors tokios kalbos kaip anglų ir ispanų greičiausiai išliks besivystančioje skaitmeninėje rinkoje, daugelis Europos kalbų gali tapti nereikalingos į tinklą susietoje visuomenėje. Tai susilpnintų Europos padėtį pasaulyje ir paženktų strateginiam tikslui užtikrinti kiekvieno Europos piliečio neatsižvelgiant į vartojamą kalbą galimybes būti visateisiu Sąjungos nariu.

Europos kalbų įvairovė – svarbi jos kultūros turto dalis.

Anot UNESCO ataskaitos apie daugiakalbystę, kalbos sudaro esminę terpę džiaugtis pamatinėmis teisėmis, tokiomis kaip teisė į politinę raišką, švietimą ir dalyvavimą visuomenės gyvenime [4].

2.3 KALBOS TECHNOLOGIJOS – NAUJŲ GALIMYBIŲ KŪRĖJOS

Anksčiau investicijos ir pastangos išsaugoti kalbų įvairovę buvo skiriamos kalbų mokymui ir vertimams. Apy-

Europai reikia stabilių, lengvai prieinamų kalbos technologijų, pritaikytų visoms Europos kalboms.

- rasti informacijos naudojantis internetinės paieškos sistema;
- patikrinti, ar tekste nėra rašybos, skyrybos ir gramatinių klaidų;
- peržiūrėti produktų rekomendacijas internetinėje parduotuvėje;
- suprasti automobilio navigacinės sistemos žodines instrukcijas;
- versti tinklalapius naudojant internetines programas.

Jei norės neprarasti pozicijos tarp pasaulio inovacijų lyderių, Europai reikės stabilų, lengvai prieinamų ir į svarbiausias programines įrangos terpes integruotų KT, pritaikytų visoms Europos kalboms. Neturėdami reikiamo lygio KT, artimiausioje ateityje nesugebėsime užtikrinti kalbos vartotojams interaktyvaus daugiakalbio bendravimo daugialypėje terpėje.

Spausdinto žodžio pasaulyje didysis technologinis proveržis buvo galimybė greitai kopijuoti vaizdą ar tekstą tinkamo galingumo spausdinimo mašina. Žmonėms teko sunkus darbas – ieškoti žinių, jas skaityti, versti ir apibendrinti. Mes turėjome laukti Thomo Edisono išradimų, padėjusių įrašyti šnekamąją kalbą. Ši technologija suteikė galimybių gaminti analogines kopijas.

Kaip ir daugumos technologijų atveju, pirmosios kalbinės programos, pavyzdžiui, balsu valdomos naudotojų sąsajos ir dialogų sistemos, buvo skirtos itin specializuotoms sritims, jų veikimas gana ribotas. Tačiau švietimo ir pramogų rinkoje esama didžiulių galimybių diegti KT žaidimams, kultūrinio paveldo sąvokoms, mokomosioms ir pramoginėms priemonėms, bibliote-

koms, imitacinėms aplinkoms ir mokomosioms programoms. Mobiliosios informavimo paslaugos, kompiuterinė programinė kalbų mokymo įranga, e.mokymo aplinka, vertinimo priemonės ir programos plagiatui aptikti, tai vos kelios sritys, kuriose KT gali būti ypač vertingos. Tokių daugialypės terpės bendravimo platformų kaip *Twitter* ir *Facebook* populiarumas leidžia manyti, kad ateityje prireiks ir pažangių KT, skirtų stebėti dalyvių žinutes, apibendrinti diskusijas, rodyti nuomonių tendencijas, aptikti emociingus atsakymus, nustatyti autorių teisių pažeidimus arba susekti netinkamo naudojimo atvejus.

Kalbos technologijos padeda įveikti kalbų įvairovės „negalią“.

KT teikia daug galimybių Europos Sąjungai. Jos gali padėti išspręsti sudėtingą Europos daugiakalbystės problemą – skirtingos kalbos gali kartu gyvuoti Europos įmonėse, organizacijose ir mokyklose. Gyventojai turi bendrauti be Europos bendrąją rinką skaidančių kalbų barjerų, o KT gali padėti jiems įveikti šią kliūtį bei teikti daugiau galimybių laisvai ir viešai kalbėti savomis kalbomis. Žvelgiant dar toliau į priekį, Europos novatoriškos daugiakalbės technologijos pažymės gaires mūsų tarptautiniams partneriams, pradėsiantiems kurti technologijas savoms daugiakalbėms bendruomenėms. KT gali būti vertinamos kaip „pagalbinės“, padedančios visiems įveikti kalbų įvairovės „negalią“ ir suteikiančios skirtingoms kalbos vartotojų bendruomenėms daugiau galimybių bendrauti. Galiausiai viena iš aktualių mokslinių tyrimų sričių yra KT pritaikymas vykdant gelbėjimo operacijas nelaimių zonose, kur kiekvienas veiksmas gali lemti gyvybę ir mirtį: ateityje sumanieji robotai, galintys bendrauti įvairiomis kalbomis, turės daug daugiau galimybių gelbėti gyvybes.

2.5 IŠŠŪKIAI, KURIUOS TURI ĮVEIKTI KALBOS TECHNOLOGIJOS

Nors KT pastaraisiais metais padarė nemažą pažangą, dabartiniai technologinės pažangos ir produkcijos naujovių taikymo tempai yra pernelyg lėti. Plačiau taikomos tekstų rašymo sistemų rašybos ir gramatikos tikrintuvės paprastai yra vienakalbės, o ir įdiegtos toli gražu ne visoms kalboms.

Šiuo metu technologinė pažanga yra pernelyg lėta.

Internetinės automatinio vertimo paslaugos yra naudingos, kai norima greitai gauti apytikslį dokumento turinio vertimą, tačiau susiduriama su sunkumais, kai reikia itin tikslaus ir išsamaus vertimo. Dėl natūraliosios kalbos sudėtingumo kalbų modeliavimas programinėje įrangoje ir jų testavimas realioje aplinkoje yra ilgas, brangus procesas, kurį reikia nuolat finansuoti. Vis dėlto Europa privalo išsaugoti lyderės pozicijas sprendama daugiakalbės bendruomenės technologines problemas ir rasti naujų būdų spartinti plėtrą visose srityse. Metodai galėtų būti susiję tiek su pažanga skaitmeninėje (kompiuterių) srityje, tiek su tokiomis metodikomis kaip individualių užduočių perdavimas žmonių grupėms ar bendruomenei (angl. *crowdsourcing*).

2.6 KALBOS ĮVALDYMAS: ŽMONĖS IR MAŠINOS

Norėdami suprasti, kaip kompiuteriai „išmoksta“ kalbą ir kodėl juos taip sunku užprogramuoti ją vartoti, trumpai pažvelkime į tai, kaip pirmąją ir antrąją kalbą išmoksta žmonės, o po to – kaip veikia KT sistemos. Žmonės įgyja kalbos įgūdžių dviem skirtingais būdais. Kūdikiams išmoksta kalbą girdėdami, kaip bendrauja jų

tėvai, broliai, seserys ir kiti šeimos nariai. Dvejų metų amžiaus vaikai pradeda patys kalbėti – iš pradžių tai būna atskiri žodžiai ir trumpos frazės. Tai įmanoma todėl, kad žmonės genetiškai gali mėgdžioti ir praktiškai taikyti, ką išgirdo.

Vyresnio amžiaus vaikams išmokti antrąją kalbą yra kiek sunkiau, ypač jeigu vaikas auga bendruomenėje, kuriai ta antroji kalba nėra gimtoji. Mokyklose paprastai mokoma užsienio kalbos gramatinės struktūros, žodyno ir rašybos taisyklių, t. y. atliekami pratimai, kuriais kalbos žinios įtvirtinamos pagal abstrakčias taisykles, lenteles ir pavyzdžius.

Dviejų pagrindinių tipų KT sistemos kalbos išmoka panašiais būdais. Statistiniai (arba pagrįstais duomenimis) metodais duomenys surenkami iš gausybės konkrečių tekstų pavyzdžių. Vienakalbiai tekstai gali būti naudojami tos pačios kalbos mokymo tikslais, pavyzdžiui, tikrinti, ar nėra rašybos klaidų, tačiau automatinio vertimo sistemai „išmokyti“ būtini lygiagretūs tekstai dviem ar daugiau kalbų. Iš tokių tekstų automatinio vertimo algoritmas „išmoka“ žodžių, trumpų frazių ir visų sakinių vertimo modelius.

Žmonės įgyja kalbos įgūdžių dvejopai:
mokydamiesi iš pavyzdžių ir mokydamiesi
pagrindinių kalbos taisyklių.

Tokiam statistiniais metodais pagrįstam vertimui gali prireikti milijonų sakinių, o vertimo kokybė gerėja didėjant išanalizuotų tekstų kiekiui. Tai viena iš priežasčių, kodėl paieškos sistemų teikėjai pageidauja sukaupti kaip galima daugiau rašytinės informacijos. Tekstų rašymo ir tokių paslaugų kaip *Google Search* ir *Google Translate* rašybos klaidų taisymo funkcijos pagrįstos statistiniais metodais. Didžiausias statistikos pranašumas yra

tas, kad mašina „mokosi“ greitai, įveikdama nesibaigiančius mokomuosius ciklus, nors kartais kokybė gali būti labai įvairi.

Antrasis KT ir ypač automatinio vertimo metodas yra kurti taisyklėmis pagrįstas sistemas. Lingvistikos, kompiuterinės lingvistikos ir kompiuterių mokslo ekspertai visų pirma turi užkoduoti gramatinę analizę (vertimo taisykles) ir sudaryti žodyno sąrašus (leksikonus). Tam reikia daug pastangų, intensyvaus darbo ir laiko. Kai kurios pagrindinės taisyklėmis pagrįstos automatinio vertimo sistemos buvo tobulinamos daugiau nei du dešimtmečius. Pagrindinis taisyklėmis pagrįstų sistemų pranašumas yra tas, kad ekspertai gali geriau kontroliuoti kalbos apdorojimą. Tai leidžia nuosekliai taisyti programinės įrangos klaidas ir palaikyti grįžtamąjį ryšį su naudotoju, ypač tuo atveju, kai taisyklėmis pagrįstos sistemos taikomos kalbų mokymui. Tačiau dėl didelių finansinių sąnaudų taisyklėmis pagrįstų KT turi susikūrusios tik didžiosios kalbos.

Kadangi paprastai statistinių ir taisyklėmis pagrįstų sistemų pranašumai ir trūkumai kompensuoja vieni kitus, šiuo metu atliekami tyrimai siekiant sukurti hibridinius būdus, sujungiančius šias dvi metodologijas. Tačiau kol kas didesnės sėkmės pasiekta ne taikant šiuos metodus verslo poreikiams, bet tyrimų laboratorijose.

Taigi galima konstatuoti, kad didžioji dalis programų, kurias šiuo metu naudoja informacinė visuomenė, labai priklauso nuo KT. Tai ypač pasakytina apie Europos ekonominę ir informacinę erdvę su jos daugiakalbe bendruomene. Nors pastaraisiais metais KT srityje padaryta reikšminga pažanga, vis dar egzistuoja daugybė galimybių gerinti KT sistemų kokybę. Kituose skyriuose aptarsime lietuvių kalbos vietą Europos informacinėje visuomenėje ir įvertinsime dabartinę lietuvių KT būklę.

The Lithuanian Language in the Digital Age

Rehm, G.; Uszkoreit, H. (Eds.)

2012, VI, 85 p. 24 illus. in color., Softcover

ISBN: 978-3-642-30757-7