

# EUROPOS KALBŲ MAŠININIS VERTIMAS: PROBLEMAS, IŠŠŪKIAI, ATEITIES PERSPEKTYVOS

Andrejs Vasiljevs  
„Tilde“, LATVIJA  
andrejs@tilde.com

## Kalbos barjerai

Europos Sąjunga stengiasi pašalinti visas vidaus kliūtis laisvai prekybai ir judėjimui. Tačiau vis dar egzistuoja viena nematoma kliūtis, kuri trukdo užtikrinti, kad visų ES valstybių piliečiai gautų informaciją ir galėtų bendrauti tarpusavyje. Ta kliūtis – tai kalbos barjerai.

Kalba – esminis socialinio gyvenimo ir bendravimo verslo tikslais elementas, svarbiausia tapatybės dalis, kiekvienos tautos socialinis ir kultūrinis turtas. Daugiakalbystė – Europos turtas ir įvairovės šaltinis. Tačiau kalbinė įvairovė taip pat sudaro kliūtis verslui, žmonių bendravimui ir bendradarbiavimui Europoje ir pasaulyje. Kalbų barjerai tampa kliūtimi, apsunkinančia Europos vienybės ir konkurencingumo užtikrinimo užduotį.

Veikiausiai ne vienas esame neįtikėję pasijutę nemokėdami kokios nors kalbos. 2011–2012 m. Eurobarometro tyrimų duomenimis, tik 39 procentai Europos interneto vartotojų bendrauja internetu (t.y. rašo elektroninius laiškus, naudojami programomis „Twitter“, „Facebook“ ir t.t.) ne savo gimtąja kalba. 46 procentai europiečių nemoka kalbėti kita nei gimtoji kalba ir maždaug tiek pat žmonių nieko neskaito užsienio kalba. 43 procentai Europos interneto vartotojų neperka prekių ir paslaugų, jei jos parduodamos ne jų gimtąja kalba. 6 procentams Europos gyventojų kasdien reikia vertimo.

Kalbų įvairovė turi savo kainą. Verslui tai kainuoja brangiai. Jei maža Lietuvos įmonė norėtų prekiauti savo prekėmis visoje Europoje, jai būtų labai brangu išversti visus tekstus ir informacinę medžiagą į 30 kalbų. Tai brangus malonumas ir valstybėms. ES institucijos oficialių dokumentų vertimui į 24 oficialiąsias ES kalbas išleidžia 1,1 mlrd. eurų. Vertėjams tenka dirbti su 552 kalbų kombinacijomis.

## Taikomosios mašininio vertimo programos

Kaip sumažinti kalbų barjerą, kartu išsaugant ir skatinant kalbų vystymąsi ir įvairovę? Eksponentinis daugiakalbės informacijos kiekio augimas gerokai viršija vertėjų galimybes viską išversti.

Kaip ir daugelyje kitų sričių, čia galima pasitelkti į pagalbą technologijas. Kalbų technologijos – tai veiksmingiausia priemonė kainos ir laiko atžvilgiu kalbos barjerui įveikti. Mašininis vertimas – vienintelis lengvai įgyvendinamas sprendimas, leidžiantis greitai ir pigiai gauti informaciją, parašytą užsienio kalbomis. Štai kodėl mašininis vertimas itin svarbi technologija daugiakalbei Europai.

Dar prieš keletą dešimtmečių atrodė neįmanoma, kad kompiuteriai ir robotai galės mus suprasti ir versti į daugelį kalbų. Dabar tokios kalbų technologijos jau tampa mūsų gyvenimo dalimi. Tokių pasaulio masto paslaugų kaip „Google Translate“ ir „Microsoft Translator“ dėka mašininis vertimas per pastaruosius kelerius metus imtas naudoti gerokai dažniau.

Automatinis vertimas negali pakeisti vertėjo, tačiau tokį vertimą galima pritaikyti daugelyje sričių. Mašininis vertimas suteikia galimybę iš karto gauti informaciją, užrašytą užsienio kalba, kuri kitaip būtų nenaudojama arba pareikalautų daug laiko ir išlaidų vertimui. Mašininis vertimas gali būti naudojamas daugiakalbėms interneto svetainėms kurti, jis gali palengvinti informacijos paiešką ir analizę nepriklausomai nuo kalbos, kuria ta informacija yra pateikta, jis gali būti taikomas ne tik rašytiniam tekstui, bet ir sakybinei kalbai, pvz. vaizdinės medžiagos dubliavimui. Mašininis vertimas didina vertėjų produktyvumą, nes jie gauna juodraštinį

vertimo variantą, kurį tereikia patobulinti ir suredaguoti. Kai kuriose specializuotose srityse naudojant mašininį vertimą galima susilaukti beveik tinkamos leidybai išversto teksto kokybės.

Technologijų naudojimo potencialą akivaizdžiai rodo ir Europos Komisija – viena stambiausių mašininio vertimo vartotojų pasaulyje. Automatinis vertimas yra kasdieninė darbo priemonė, su kuria dirba daugiau kaip 2000 Europos Komisijos vertėjų; ji padeda didinti darbo našumą ir mažinti išlaidas.

Inovatyvios kalbų technologijos gali gerokai padidinti didelių ir mažų įmonių konkurencingumą. Tas pats pasakytina apie valstybinę valdžią. Mašininio vertimo dėka vyriausybės gali sutaupyti pinigų ir teikti geresnes paslaugas įvairių kultūrų atstovams, kurie gyvena Europos teritorijoje.

Kalbų technologijos apskritai ir mašininis vertimas konkrečiai yra priemonės, naudingos mažinant skaitmeninę atskirtį ir įveikiant kalbų barjerą bei skatinant bendravimą ir prekybą. Jos atlieka šias funkcijas:

- suteikia piliečiams galimybių gauti informaciją iš visos Europos;
- mažindamos kalbų barjerus ir leisdamos įeiti į naujas rinkas gerina Europos konkurencingumą;
- padeda naudotis turtingu ir įvairiu Europos kultūros paveldu;
- didina profesionalių vertėjų darbo našumą ir mažina vertimo kainas;
- mažina globalizacijos pavojų tautinei tapatybei;
- skatina elektroninės demokratijos sklaidą ir piliečių dalyvavimą priimant visai Europai svarbius politinius sprendimus.

## Iššūkis

XX a. viduryje, kai mokslininkai ėmė domėtis šia sritimi, buvo pernelyg optimistiškai manoma, kad automatinis vertimas bus atliekamas žodynų pagrindu arba paprasčiausiai pažodžiui verčiant vienos kalbos žodžius į kitą kalbą. Tačiau netrukus suvokta, kad kalbų struktūros skiriasi, žodžiai turi daug prasmių, o kalbos išraiškos priemonės labai turtingos, todėl ši užduotis yra labai sudėtinga. Geras vertėjas supranta teksto prasmę ir perteikia ją kita kalba, stengdamasis išlikti kuo arčiau originalo. Kadangi kompiuteris negali suprasti prasmės taip pat, kaip žmogus, teko taikyti įvairias sintaksinės ir semantinės analizės priemones bei statistinius metodus.

Statistiniai metodai pasiteisino ir itin dažnai taikomi mašininio vertimo priemonėse, ypač „Google Translate“ priemonėje ir daugelyje kitų sistemų. Tačiau vertimo kokybė dar toli gražu netobula. Europoje reikėtų sukurti naujos kartos technologijas, kurios padėtų gerinti vertimo kokybę ir labiau priartėti prie gyvos kalbos struktūros ir prasmės.

Itin problemiškos yra gausios mažosios Europos kalbos, kurioms būdinga sudėtinga morfologinė ir lingvistinė struktūra. Dėl ekonominių priežasčių verslo dėmesys labiausiai sutelktas ties didžiosiomis kalbomis. Pavyzdžiui, nors „Google Translate“ prieinamas daugiau kaip 70 kalbų, tačiau vertimo į Baltų kalbų grupės kalbas kokybė yra kur kas prastesnė, nei į anglų, prancūzų, ispanų ir kitas didesnes kalbas.

Vertindami kalbų technologijų plėtros lygį, ekspertai padarė išvadą, kad 21 ES kalbos skaitmeninės priemonės (mašininis vertimas ir kitos kalbų technologijos) yra prastai išvystytos, o šių kalbų ateičiai skaitmeniniame amžiuje gresia pavojus išnykti.

Šiai technologijų atskirčiai įveikti būtini tikslingi veiksmai. Kaip yra pasakęs Latvijos Ministras Pirmininkas Valdis Dombrovskis, „Vienintelis būdas užtikrinti mūsų kalbos gyvavimą ateityje yra suteikti jos vartotojams tokias pat galimybes, kokias turi kitų kalbų vartotojai.“

## Vizija

Europos meistriškumo tinklas META-NET sukūrė įspūdingą vertimo technologijų vystymosi viziją ir strategiją. META-NET strateginėje tyrimų darbotvarkėje<sup>1</sup> teigiama, kad jeigu Europos bendrija padarys kryptingą šuolį pirmyn, iki 2020 m. jai gali pavykti sugriauti daugumą kalbos barjerų ir sukurti visavertę skaitmeninę rinką. Europiečiai galės bendrauti savo gimtąja kalba ne tik tarpusavyje ir su savo valdžia, bet taip pat galės naudotis interneto paslaugomis gimtąja kalba.

Siekiamo, kad klestėtų daugiakalbė Europos visuomenė, kurioje visi piliečiai galėtų naudotis visomis paslaugomis, gauti visą informaciją, naudotis visomis komunikacijos priemonėmis ir valdyti visas technologijas savo gimtąja kalba. Tai bus pasaulis, kuriame kalbos barjeras nebekels kliūčių raštiškam ir žodiniam bendravimui ir kuriame net specializuoti aukštos kokybės vertimai kainuos palyginti nebrangiai.

## Pasiekimai

Ateities vizija remiasi naujausiais pasiekimais. Europos Bendrijų Septintosios mokslinių tyrimų, technologijų plėtros ir demonstravimo veiklos pagrindų programos ir Konkurencingumo ir inovacijų bendrosios programos (CIP) dėka sukurta keletas pirmaujančių pasaulyje mašininio vertimo technologijų. Bene geriausias jų pavyzdys – tai priemonė „Moses“ – plačiausiai pasaulyje naudojama mašininio vertimo priemonė.

„Tilde“ didžiuojasi lyderio pozicijomis vystant internetinę platformą LetsMT!, skirtą specializuotoms daugiakalbėms mašininio vertimo sistemoms kurti. LetsMT! grindžiama naujausiais pasiekimais statistinio mašininio vertimo srityje. Tai priemonė, kuri mokosi versti analizuodama milijonus išverstų sakinių. LetsMT! platforma suteikia galimybes kurti įvairias specializuoto mašininio vertimo sistemas, kurios ypač tinka mažesnėms kalboms ir specializuotam vertimui. Jos padėjo įmonei „Tilde“ sukurti daug mašininio vertimo sistemų, tinkančių konkrečioms sritims ir užduotims, kurių vertimo kokybė kur kas geresnė už „Google Translate“. Pažangos pasiekta ir kuriant TaaS (terminologijos paslauga) platformą, skirtą neautomatinio ir mašininio vertimo taisyklių terminų problemai spręsti.

Latvija ir Lietuva lietuvių ir latvių kalbų sklaidai siekia finansavimo iš ES struktūrinių fondų. Neseniai Lietuva pradėjo plataus masto programą, skirtą vystyti lietuvių kalbą informacinės visuomenės poreikiams tenkinti. Mašininis vertimas – vienas pagrindinių šios programos projektų. „Tilde“ atsakinga už viešos mašininio vertimo paslaugos Lietuvos piliečiams sukūrimą. Latvijoje mašininio vertimo plėtra užtikrinama remiantis Kompetencijų centro programa „Kalbos krantas“. Latvijos vyriausybė padarė svarbų žingsnį pradėdama kurti specializuotą e-valdžios mašininio vertimo priemonę, kad internetu teikiamos paslaugos būtų prieinamos Latvijos ir užsienio daugiakalbėms bendruomenėms. Estija ir kelios kitos valstybės taip pat turi specialias programas kalbų technologijų plėtrai, įskaitant mašininio vertimo išteklius.

## Tikslingų veiksmų poreikis

Siekiant įgyvendinti viziją, reikalingi bendri veiksmai, į kuriuos turi būti įtraukiamos pagrindinės suinteresuotosios šalys – vyriausybės, mokslininkai ir pramonės atstovai. Reikalingi šie darbai:

- pažangūs tyrimai siekiant didelio proveržio mašininio vertimo kokybės srityje;
- palankių sąlygų sudarymas atvirai Europos vertimo ir kalbinių paslaugų infrastruktūrai kurti;
- mašininio vertimo naudojimo naujose taikomosiose programose ir integravimo įvairiose informacijos paslaugose ir sistemose skatinimas;
- automatinio vertimo taikymas sprendžiant svarbius visuomenės gyvenimo klausimus tokiose srityse kaip švietimas, darbo rinka, demokratinis visuomenės dalyvavimas ir viešosios paslaugos;
- technologijų atskirties problemos sprendimas mašininio vertimo ir kitose kalbos technologijų srityse;

<sup>1</sup> META-NET, [www.meta-net.eu/sra](http://www.meta-net.eu/sra)

- kalbinių išteklių mašiniam vertimui ir kitoms kalbos technologijoms kūrimo ir sklaidos skatinimas;
- mašininio vertimo priemonių ir metodų vystymas;
- profesionalių vertėjų ir įmonių įtraukimas į visą mokslinių tyrimų ir inovacijų procesą.

Norint, kad visa tai taptų tikrove, reikia mobilizuoti ir koordinuoti tokios veiklos dalyvių veiksmus nacionaliniu ir tarptautiniu lygiu, inicijuoti ir sinchronizuoti mokslinių tyrimų ir plėtros veiklą valstybiniu lygiu, siekti finansavimo iš ES mokslinių tyrimų ir plėtros programų proveržiui inovacijų srityje pasiekti ir Europos lyderystei pasaulyje įtvirtinti.