

LEKŠINĖS MAŠININIO VERTIMO KLAIDOS: BEEKVIVALENČIŲ ŽODŽIŲ VERTIMAS

JURGITA CVILIKAITĖ

Vytauto Didžiojo universitetas
j.cvilikaite@hmf.vdu.lt

Pagrindiniai žodžiai: mašininis vertimas, daugiareikšmiškumas, homonimija, realija, lakūna, beekvivalenčiai žodžiai.

2008 m. pradžioje pradėjusi veikti VDU sukurta internetinė mašininio vertimo sistema, verčianti tekstą iš anglų kalbos į lietuvių, buvo ne kartą pristatyta Lietuvos vartotojams. Ne vienas jau spėjo ją išbandyti. Kaip visada, esama ir giriančių, ir peikiančių. Straipsnyje aprašomas šios programos kokybės tyrimas, aptariamos esminės problemos, su kuriomis programoje susiduriama. Dėmesys sutelkiamas į sudėtingą beekvivalenčių žodžių vertimą, apibūdinami terminai *realija*, *lakūna* ir *beekvivalenčiai žodžiai*, pateikiami argumentai, kodėl terminas *beekvivalenčiai žodžiai* yra tinkamiausias. Tiriama medžiaga atskleidžia įvairialypę beekvivalenčių žodžių vertimo mašininio vertimo programa problematiką.

Anglų – lietuvių kalbų mašininio vertimo sistemos testavimas

Anglų – lietuvių kalbų mašininio vertimo sistema yra Vytauto Didžiojo universiteto vykdyto projekto „Internetinė informacijos vertimo priemonė“ pagrindinis tikslas ir rezultatas. Šio projekto tikslas buvo sukurti viešą paslaugą internete, leidžiančią automatiškai versti bendro informacinio pobūdžio tekstus iš anglų kalbos į lietuvių 70 proc. tikslumu (Daudaravičius 2006, 15).

Programai testuoti pasinaudota kitų mokslininkų vertinimo pavyzdžiu ir metodika. A. Calude (2003) analizavo ir aprašė, kaip „Systran“ mašininio vertimo programa verčia neilgus (apie 200 žodžių) įvairių stilių tekstus iš vokiečių kalbos į anglų kalbą. Minėtos programos vertimo kokybę vertino abi kalbas mokantys kalbininkai. Vertinti buvo pasirinktos techninės instrukcijos, žurnalų ir laikraščių straipsnių ištraukos ir literatūrinio teksto atkarpos. Tokį pasirinkimo lėmė pats tyrimo tikslas: buvo siekiama išsiaiškinti, ar mašininio vertimo klaidos gali priklausyti nuo verčiamo teksto stiliaus, koks yra skirtingo stiliaus vertimo patikimumas. Buvo keliama hipotezė, kad ribotos semantinės srities tekstai (pavyzdžiui, techniniai) yra verčiami geriau



(t. y. tiksliau ir patikimiau) nei grožinės literatūros kūriniai.

Tyrimo rezultatai parodė, kad mašininio vertimo patikimumą riboja tekstų pobūdis. Tai nėra netikėta išvada. Kiekvienas vertėjas žino, kad tekstų, kuriuose vyrauja sudėtingos struktūros sakiniai, gausu pragmatinės informacijos, didelė leksikos įvairovė, o ir patys žodžiai nėra glaudžiai susiję su konkrečia sritimi, vertimas yra sudėtingas. Tad minėto autoriaus tyrimas tik patvirtino tai: automatiškai išversto literatūrinio teksto nebuvo įmanoma suprasti, o mažiausiai klaidų padaryta verčiant technines instrukcijas. Tuo remiantis daroma išvada, kad visų stilių tekstus įmanoma versti automatiškai. Jeigu vartotojo tikslas yra suprasti pagrindinę teksto mintį, įsitikinti, ar tekstas įdomus, ar verta į jį gilintis, tada tokios vertimo kokybės pakanka. Tačiau geresnių mašininio vertimo rezultatų galima tikėtis tik verčiant technines instrukcijas (Calude 2003, 16–17).

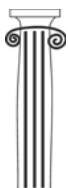
Panašiai buvo vertinama ir anglų – lietuvių kalbų mašininio vertimo sistema. Remiantis lietuvių kalbininkų (Pikčilingis 1971; Župerka 1997) siūlomomis stilių klasifikacijomis tyrimui pasirinkti pagrindinių rašytinės kalbos stilių – mokslinio, administracinio, publicistinio, meninio – tekstai. Buitinio (arba šnekamosios kalbos) stiliaus tekstai neanalizuoti: neįmanoma rasti natūralių anglų šnekamosios kalbos tekstų, išverstų į lietuvių kalbą. Tyrimo pradžioje buvo mėginama versti interneto dienoraščių (vadinamųjų *blogų*) bei pokalbių svetainių dialogų tekstų atkarpas. Tačiau interneto pokalbiuose vartojama kalba yra specifinė, gausu nepilnų sakinių, sutrumpinimų, sutartinių simbolių (pvz., *lol* – „laughing out loud“ ir kt.), todėl tokių tekstų mašininį vertimą apskritai buvo neįmanoma suprasti ar kaip nors analizuoti.

Kiekvieno stiliaus tekstų rinkinių apimtis – apie 27 tūkst. žodžių, o visų anglų kalbos tekstų apimtis buvo apie 107 tūkst. žodžių.

Pagrindinius mašininio vertimo kokybės vertinimo modelius apibendrina J. White (2003): tinkamumo (angl. *feasibility*), vidinio vertinimo, konstatuojamojo vertinimo, patogumo naudoti (angl. *usability*), operacinio ir palyginamojo vertinimo testai. Tinkamumo testais siekiama išsiaiškinti, ar tam tikra teorija ar metodas gali pagerinti mašininį vertimą. Vidinio vertinimo būdu testuojami tam tikri sistemos komponentai. Vidinio vertinimo testai esti dvejopi: „juodosios dėžės“ (kai analizuojami originalo ir vertimo kalbų tekstai, nesigilinant, kas vyksta pačios sistemos viduje) ir „stiklinės dėžės“ (kai siekiama išsiaiškinti, kaip veikia atskiri mašininio vertimo sistemos komponentai). Konstatuojamųjų vertinimų rezultatai paprastai būna tam tikra vertimo kokybės statistika (skaičiuojamos klaidos) arba tiesiog vertėjų ir mašininio vertimo sistemos atliktų vertimų palyginimas. Patogumo naudoti testais siekiama išsiaiškinti, ar vartotojai, kuriems buvo sukurta mašininio vertimo sistema, gali ja tinkamai naudotis. Palyginamasis metodas taikomas lyginant įvairias mašininio vertimo sistemas tarpusavyje (White 2003, 222).

Šis tyrimas atliktas remiantis dviem metodais: „juodosios dėžės“ (t. y. nesigilinama, kaip veikia pati programa, tik analizuojami originalo ir vertimo kalbos testai) ir konstatuojamuoju vertinimu (gretinami programos ir vertėjų išversti tie patys anglų kalbos tekstai).

Labai problemiškas ir diskutuotinas yra vertimo kokybės apibrėžimas. Koks



tekstas esti kokybiškai išverstas? Tikriausiai čia negalima kalbėti apie vieną paramet-
rą. Paprastai vertimo kokybė nustatoma tam tikru kriterijų kompleksu, iš kurių svar-
biausi: taisyklingumas (gramatinis) ir tikslumas (prasmės, turinio perteikimo tikslu-
mas). Kai kurių mokslininkų nuomone (Daudaravičius 2006, 17), svarbiausiu laikyti-
nas teksto turinio, prasmės perteikimas vertime. Analizuojant automatiškai išverstus
tekstus buvo susitelkta į prasmės perteikimą. Savaimė suprantama, taip analizuojant
verstinius tekstus svarbiausias yra žodžių (leksikos vienetų) parinkimas. Šiuo aspek-
tu aprašomoji programa testuota iki pristatant vartotojams. Tą darė šeši testuotojai
(tarp jų – šio straipsnio autorė). Siekiant įvertinti sakinio, leksinio ir gramatinio lyg-
mens ekvivalentiškumą, buvo analizuojama 100 įvairaus ilgio sakinių. Tyrimo metu
nustatyta, kad leksikos vertimo klaidų rasta daugiausiai – 43,2 procento (Rimkutė ir
Kovalevskaitė 2008, 258–259).

Analizuota, ar sakinyje suprantamas. Jeigu jis sunkiai suprantamas, buvo ieško-
ma netikslių vertimo vietų ir jų atitikmenų originalo kalbos tekste. Vadinas, nuo
verstinio teksto buvo einama prie originalo kalbos vienetų, kuriuos sunku išvesti
automatiškai. Toks principas akcentuojamas ir kituose panašiuose tyrimuose. Nusta-
tyta, kad vertimo kokybė pirmiausia priklauso nuo to, kaip tiksliai bus išanalizuota
kalbos, iš kurios verčiama, teksto atkarpa. Mašininio vertimo procese kalbos, iš ku-
rios verčiama, analizė sudaro apie 90 proc. veiksmų, lemiančių vertimo kokybę, ir
tik apie 10 proc. – kalbos, į kurią verčiama, sintezė (Rimkutė ir Kovalevskaitė 2007).
Todėl ir anglų – lietuvių mašininio vertimo programos klaidos analizuotos iš origina-
lo kalbos pozicijų.

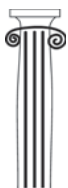
Pagrindinės leksinės anglų – lietuvių kalbų mašininio vertimo klaidos

Mokslinėje literatūroje (Arnold et al 1994; Gross 1992; Calude 2003; Rimkutė
ir Kovalevskaitė 2007) minimos kalbinės ir nekalbinės mašininio vertimo klaidų
priežastys. Iš kalbinių priežasčių, darančių įtaką mašininio vertimo kokybei, dažniau-
siai išskiriamas konteksto neatitinkančių daugiareikšmių žodžių reikšmių parinki-
mas, sintaksinis kalbos daugiareikšmiškumas, struktūriniai ir leksiniai originalo ir
vertimo kalbų skirtumai. Kai kada dar atkreipiamas dėmesys ir į idiomų, kolokacijų
bei leksinių ertmių vertimą (Arnold et al 1994, 105).

Grossas (1992, 109–110) teigia, kad kompiuteriai kol kas negeba kaip žmonės
suvokti kontekstinės ir pragmatinės informacijos. Tą patį galima pasakyti ir apie
kultūrinius skirtumus, kurie verčiant ypač aktualūs. Šiai nuomonei pritaria Arnold
ir bendra autoriai (1994, 108). Mašininio vertimo sistemoms sudėtinga perteikti ir
kalbos vienetų ekspresines funkcijas. Be dalykinio turinio, tekste esti ir emocijų (juo-
kaujama, daroma įtaka ir kt.). Calude (2003) taip pat pastebi, kad mašininio vertimo
modelis negali atskleisti kūrybiškumo ir visų kalbos funkcijų. Apskritai nekalbinėje
plotmėje teigiama, kad vertimas yra subjektyvus veiksmas. Net vertėjai tą patį tekstą
išverčia skirtingai, tad negali būti vienintelio teisingo vertimo (Calude 2003, 5–6).

Prie nekalbinių mašininio vertimo klaidų priežasčių priskirtinos ir techninės ypa-
tybės. Čia svarbūs du aspektai – leksinis ir kalbos generavimo. Pastebėta, kad mašini-





nio vertimo sistema žodžių atitikmenų ieško paprastai tik viename bendro pobūdžio žodyne, kuriame stengiamasi palikti po vieną variantą kiekvienai jo reikšmei (Rimkutė ir Kovalevskaitė 2007). Tačiau žodžiai gali turėti daug reikšmių ir galbūt konkrečiam tekstui taiklesnė būtų ne ta reikšmė, kuri palikta žodyne. Tad kokybiški bendrieji žodynai, specializuoti žodžių sąrašai, atskirų reikšmių kontekstų, vartojimo sričių žymos yra labai svarbūs, siekiant pagerinti mašininio vertimo kokybę.

Kita svarbi mašininio vertimo savybė yra vertimo principas. Vertimo atitikmenys gali būti parenkami ir tekstas generuojamas dviem principais – statistiniu arba formaliuoju, pagrįstu taisyklėmis. Anglų – lietuvių kalbų mašininio vertimo sistema yra mišraus tipo, nes naudojami ir statistiniai, ir formalieji metodai (Daudaravičius 2006, 15). Naudojantis aprašomąja vertimo programa matyti, kad dabartinis mašininis vertimas iš anglų kalbos į lietuvių yra pernelyg pažodinis. Tai leidžia manyti, kad minėtų dviejų vertimo sistemos veikimo principų nepakanka. Pavyzdžiui, analizuojant originalo tekstą yra atsižvelgiama į kai kurias gramatines anglų kalbos taisykles (pavyzdžiui, gramatinių laikų sudarymą), vis dėlto dar yra gramatikos sričių, kurias būtina aprašyti (pavyzdžiui, anglų kalbos posesyvinės konstrukcijos su prielinksniu „of“). Pažodinis vertimo išpūdis menkina ir teksto turinio perteikimą. Tai ypač akivaizdu, kai lyginamos ilgesnės vertimo atkarpos, kuriose sistema negeba identifikuoti vieno reikšmės vieneto ir skaido jį į atskirus žodžius, verčia pažodžiui. Tačiau tokį vertimo principą verta išnaudoti aprašant kalbą. Pastebėta, kad žodyniniai pagrindinių reikšmių atitikmenys ir pažodinis žodžių junginių vertimo būdas yra priemonė, leidžianti parodyti kelių kalbų conceptualiųjų metaforų skirtynes, išryškėjančias vertime (Marcinkevičienė 2006, 111). Galima teigti, kad šis būdas taip pat tinka ir kitiems originalo kalbos žodžių junginiams atskleisti. Taip identifikavus ir aprašius tinkamo ilgio originalo kalbos vertimo vienetus ir parinkus jų atitikmenis, neabejotinai pagerėtų mašininio vertimo kokybė.

Susisteminius analizuojamas mašininio vertimo klaidas matyti, kad jos atsirado ne dėl stiliaus specifikos, o apskritai dėl kalbos specifikos, tokių chrestomatinių leksikos reiškinių kaip daugiareikšmiškumas ir homonimija, kalbinio ir nekalbinio konteksto, tokių techninių mašininio vertimo savybių kaip žodyno kokybė bei pažodinis vertimo principas. Tipiškiausios leksinio pobūdžio mašininio vertimo klaidos yra šios:

- 1) netinkamai parinkta daugiareikšmio žodžio reikšmė ir anglų kalbos homonimo atitikmuo;
- 2) lietuvių kalbos sakinio prasmė prarasta dėl pažodinio vertimo, nes nenustatytas vertimo vienetas;
- 3) lietuvių kalba stokoja beekvivalenčių anglų kalbos žodžių atitikmenų.

Daugiausiai vertimo klaidų programa padaro parinkdama netinkamą daugiareikšmio žodžio reikšmę. Čia svarbūs ir pačios programos leksiniai ištekliai, jos naudojamas žodynas, reikšmių filiacija bei atskirų reikšmių pažymos. Kadangi projekto tikslas buvo sukurti tokią vertimo programą, kuri verstų bendro pobūdžio tekstus, programa nenaudoja specializuotų žodynų. Su daugiareikšmiškumu ir homonimi-

ja susijusių vertimo klaidų gausa rodo, kad šios programos žodyną privalu taisyti, plėsti, papildyti tam tikromis žymomis, leisiančiomis tiksliau parinkti vertimo atitikmenis. Tą akivaizdžiai iliustruoja ir toliau pateikiami pavyzdžiai, kuriuose netinkamai išversti daugiareikšmiai anglų kalbos žodžiai.

Anglų kalbos sakinys

<lang value=„en“><s>Such an arrangement at EU level will promote on the one hand effectiveness, clarity and legal certainty, and, on the other a more open and **transparent** public procurement regime.</s></lang>

Mašininis vertimas

<lang value=„lt“><s>Toks susitarimas ES lygmenyje skatins viena vertus efektyvumą, aiškumą ir teisėtą įsitikinimą, ir, ant kito labiau atidarytas ir **vaiskus** visuomeninis įsigijimo režimas.</s></lang>

Žmogaus vertimas

<lang value=„lt“><s>Toks susitarimas ES lygiu galėtų, kaip tikimasi, padidinti efektyvumą, suteikti daugiau aiškumo ir teisinio tikrumo, viešųjų pirkimų tvarka taptų atviresnė ir **skaidresnė**.</s></p></lang>

Būdvardžio *transparent* reikšmė vertimo žodyne išversta žodžiais *permatomas*, *vaiskus* ir *skaidrus*. Vertimo programa parinko *vaiskus*, kuris, deja, iškreipė anglų kalbos sakinio prasmę. Vertėjas parinko kitą atitikmenį, labiau derantį prie administracinio stiliaus.

Kitame pavyzdyje matyti, kad atitikmenys klaidingai parenkami dėl kelių priežasčių iš karto.

Anglų kalbos sakinys

<lang value=„en“><s>Lord Henry went out to the garden and found Dorian Gray **burying his face in** the great cool lilac-blossoms, feverishly drinking in their perfume as if it had been wine.</s></lang>

Mašininis vertimas

<lang value=„lt“><s>Valdovas Henry išėjo sodo ir surado Dorėnų Pilką **laidojimą jo veido** dideliame vėsiame alyviniame sužydėjime, karštligiškai geriančiame jų kvapaluose, tarytum jis buvo vynas.</s></lang>

Žmogaus vertimas

<lang=„lt“><p><s>Išėjęs į sodą, lordas Henris rado Dorianą Grėjų veidu **įsikniaubusį** į didelius vėsius alyvų žiedus, karštligiškai geriantį jų kvapą, tarsi tai būtų vynas.</s></lang>

Teksto atkarpa nevykusiai išversta ne tik dėl vaizdingos kalbos (grožinis stilius), homonimijos (*burying* yra veiksmazodžio esamojo laiko dalyvis arba gerundijus bei daiktavardis), bet ir dėl konteksto, nenustačius vertimo vieneto ribų (*įsikniaubti* – tai ilgesnio vieneto – *to burry one's face in smth* – atitikmuo).

Kaip rodo tiriamoji medžiaga, homonimija – sunkiai įveikiama kliūtis dabartinei anglų – lietuvių vertimo programai. Šioje vertimo klaidų grupėje išsiskyrė anglų kalbos tikriniai daiktavardžiai: *Basil* (bazilikas – Bezilis), *Dorian* (dorėnų – Doria-



nas), *Gray* (pilkuma, pilkas – Grėjus), *Eve* (išvakarės – Eva, Ieva), *Sibyl Vane* (Sibilė Mentė – Sibilė Vein). Tokioms klaidoms atsirasti įtakos turėjo tyrimo objekto pasirinkimas: analizuojama knygos „Picture of Dorian Gray“ („Doriano Grėjaus paveikslas“) ištrauka, kurioje yra nemažai vardų, galinčių žymėti ir bendrinius žodžius. Tad kai kurios vertimo atkarpos tiesiog komiškos, pvz., *Dorėnų Pilkuma diskutuoja su Baziliku*. Tokių tikrinių vardų ir bendrinių žodžių painiojimo klaidų aptikta ir anksčiau tiriant šios programos kokybę (pvz., *Turkey* – kalakutas ir Turkija, *China* – Kinija ir porcelianas, *Spears* – asmens pavardė ir ietys) (Rimkutė ir Kovalevskaitė 2008, 260). Nemažai anglų kalbos homonimijai priskiriamų klaidų atsiranda dėl anglų kalbos priesagų *-ed* ir *-ing* daugiareikšmiškumo. Pavyzdžiui:

Anglų kalbos sakiny

<lang value=„en“><s>Good-bye, Lord Henry, you are quite delightful and dreadfully **demoralizing**.</s></lang>

Mašininis vertimas

<lang value=„lt“><s>Viso labo, Valdovas Henry, jūs esate ganėtinai žavus ir baisiai **demoralizacija**.</s></lang>

Žmogaus vertimas

Likite sveiki, lorde Henri, jūs labai mielas ir siaubingai visus **demoralizuoja**.</s></lang>

Šis pavyzdys įdomus tuo, kad anglų kalbos sakiny ir du jo vertimo variantai atskleidžia netgi tris *demoralizing* gramatinės reikšmės: būdvardžio (anglų kalbos sakinyje), daiktavardžio (mašiniame vertime) ir veiksmažodžio (žmogaus vertime). Vertėjas sąmoningai pakeitė kalbos dalį, siekdamas išvengti pažodinio ir nenatūralaus vertimo (*siaubingai demoralizuojantis*). Kita vertus, mašininio vertimo programa nesugeba atlikti tokios rūšies transformacijos ir net neatpažino homoniminės anglų kalbos formos.

Kaip jau minėta, programa nesugeba nustatyti vertimo vieneto ribų. Kitaip sakant, jei jos žodyne nėra nurodytas tam tikras daugiažodis junginys (pavyzdžiui, dviem žodžiais užrašytas sudurtinis anglų kalbos žodis, frazeologizmas, frazinis veiksmažodis, kolokacija), kiekvienas žodis verčiamas atskirai, neretai dar parenkama ir klaidinga reikšmė.

Sudurtiniai žodžiai:

full-fledged – pažodžiui išverstas *pilnai plunksnuotas* (turėtų būti *visavertis*);

printing house – spausdinimo namai (*spaustuvė*);

good-looking – geras žiūrėdamas (*gražus*);

cross-cutting – skersiniai trumpėjantys (*daugiasritis*);

Frazeologizmai:

stepping stones – mindė akmenys (*tarpiniai tikslai*);

homeward leg – namų link koja (*skrydis atgal*);

outward leg – matoma koja (*skrydis į priekį*);

to imagine oneself in someone's shoes – įsivaizduoti savo jo batuose (*įsivaizduoti save kieno kailyje*);



Kolokacijos

full-length portrait – pilno ilgio portretas (portretas visu ūgiu);

heavy temptation – sunki pagunda (didelė pagunda);

Women are a decorative sex – Moterys yra **dekoratyvinis seksas** (Moterys yra **dailioji lytis**);

A laugh ran around the table. – **Juokas bėgo aplink stalą.** (**Juokas apskriejo stalą.**);

And now I must bid good-bye to your excellent aunt. – Ir dabar turiu **pasiūlyti kainą viso labo** jūsų puikiai tetai. (O dabar turiu **atsisveikinti** su jūsų nuostabiąja teta.).

Automatinis beekvivalenčių žodžių vertimas

Tikras iššūkis mašininio vertimo programai yra angliški žodžiai, kurie lietuvių kalboje neturi atitikmenų. Mokslinėje literatūroje tokie originalo kalbos žodžiai vadinami *lakūnomis* (arba *lakunomis*) ir *realijomis*. Šių terminų vartoseną priklauso nuo jais įvardijamų sąvokų pobūdžio. Lakunos apibrėžiamos kaip „žodžiai, kurie dėl tam tikrų priežasčių neturi atitikmenų (netgi pastovių žodžių junginių) kitoje kalboje (Armalytė 1990, 123). A. Gudavičius (2007, 93) pateikia aiškesnę apibrėžimą: „...lakuna žymi žodžio nebuvimą vienoje kalboje kitos kalbos atžvilgiu, kai abiejų kalbų atspindimoje tikrovėje atitinkamas daiktas, reiškiny ir pan. egzistuoja.“ Jo nuomone, lakunų buvimo priežastis yra pasaulio kategorizacijos ypatumai kalboje, o lakunomis laikomas siauresnės arba platesnės reikšmės žodžio nebuvimas vienoje iš gretinamų kalbų. Pavyzdžiui, anglų kalboje nėra atitikmens lietuviškam žodžiui *para*.

Terminas *realija* pavadina tautos buities, istorijos, kultūros ar kitų visuomenės gyvenimo sričių daiktus, reiškinius, faktus ar įvykius (Ambrasas-Sasnava 1978, 135). Vis dėlto reikia pabrėžti, kad terminas *realija* yra suprantamas ir dažnai mokslinėje literatūroje vartojamas dvejopai – kaip realija – objektas ir kaip realija – žodis (Vlachov 1980, 6–7). Gudavičiaus (2000, 89) požiūriu, toks terminas nėra tinkamas, nes „realijos bendrinėje kalboje yra patys daiktai ir reiškiniai, o čia reikia įvardyti žodžius, pavadinančius realijas.“ Jis siūlo atsisakyti šio termino, o vietoje jo vartoti tiesiog *beekvivalentę leksiką* arba *beekvivalenčią žodžiai*. Pastarasis terminas yra logiškas, nes žymi visumos, kurią galėtume vadinti beekvivalente leksika, dalį (beekvivalentis žodis), tad teiktinas vietoje dviprasmiškai mokslinėje literatūroje vartojamo *realija*.

Vertimo programos klaidingai išversti beekvivalenčiai žodžiai priklauso kelioms grupėms: kitų tautų buities dalykus pavadinantys žodžiai (*Persian saddle-bags* – *persiškos gūnios*), specifinės buities realijas žymintys žodžiai (*hansom* – karietos tipas), aprangos pavadinimai (*liveried* – *apsirengęs livrėja*), mūsų kultūroje nežinomų dalykų rūšys (*bourdon* – varpo rūšis, žodis neverstas, kompensuotas veiksmazodžiu *gaudė*), religines nuostatas įvardijantys žodžiai ar jų junginiai (*to commit a suttee* – susideginti kartu su vyru po jo mirties) ir kt. 1 lentelėje žr. anglų kalbos beekvivalenčių žodžių vertimą mašininio būdu ir vertėjo pateiktus atitikmenis.



Įvairias realijas žymintys anglų kalbos žodžiai ir jų vertimas

Žodis ar žodžių junginys	Mašininis vertimas	Žmogaus vertimas
<i>Persian saddle-bags</i>	<i>persų persisveriami krepšiai</i>	<i>persiškos gūnios</i>
<i>garter</i>	<i>guma</i>	<i>ordinas</i>
<i>a fluted Georgian urn</i>	<i>rievėta gruzinų laidojimo urna</i>	<i>didelis gruziniškas arbatinukas</i>
<i>hansom</i>	<i>hansom</i>	<i>karieta</i>
<i>Prince Charming</i>	<i>princas, žavintis</i>	<i>pasakų princas</i>
<i>guinea</i>	<i>Gvinėja</i>	<i>ginėja</i>
<i>to commit suttee</i>	<i>pavesti suttee</i>	<i>susideginti kartu su juo (po vyro mirties)</i>
<i>bourdon</i>	<i>bourdon</i>	<i>gaudė (veiksm.)</i>
<i>liveried</i>	<i>liveried</i>	<i>apsivilkusi (veiksm.)</i>
<i>Spa</i>	<i>vandens procedūrų salonas</i>	<i>Spa miestelis</i>
<i>bushrangers</i>	-	<i>bušreindžeriai</i>

Vertėjas beekvivalenčius žodžius iš vienos kalbos į kitą gali perteikti keliais būdais: a) apibūdindamas realiją, pateikdamas papildomą komentarą; b) pavartodamas panašią realiją vertimo kalboje žymintį žodį; c) transkribuodamas originalo kalbos žodį. Mašininio vertimo programa beekvivalenčius žodžius verčia tik tada, kai jie yra jos žodyne. Kitu atveju programa gali elgtis įvairiai:

- palikti anglų kalbos žodį, pvz.: *bourdon*, *hansom*;
- apskritai praleisti žodį, kaip buvo pasielgta su *bushrangers*;
- versti pažodžiui (*Prince Charming*, *Persian saddle-bags*);
- pavartoti žodyne pateiktą reikšmės aiškinimą (*Spa*).

Atskirai reikėtų aptarti kelis atvejus. *Bushrangers* programa visai nevertė, o vertėjo pateiktas abejotinas skolinys *bušreindžeriai*. Žodyne pateiktas paaiškinimas, kad žodis reiškia pabėgusius kalinius, su nuoroda į australų istoriją. Iš tiesų šiuo žodžiu žymima sąvoka neturi nieko bendro su Teksaso reindžeriais, nes bušreindžeriai yra nusikaltėliai, kurie Australijoje slapstydami krūmuose tykodavo savo būsimų aukų ir taip išgyvendavo.

Sparčiai lietuvių kalboje įvairiuose pavadinimuose atsirandantis *Spa* automatiškai išverstas taip, kaip yra paaiškintas žodyne – „vandens procedūrų salonas“. Valstybinės lietuvių kalbos komisijos svetainėje rašoma: *Santrumpa SPA gali reikšti tam tikrą medicinos įstaigą, kurorto, grožio salono tipą, gydymo ar sveikatinimo būdą. Su minėtają santrumpa sutampa Belgijos miesto pavadinimas Spa. Miestas garsus kaip kurortas, kuriame gydoma mineraliniu vandeniu. Manoma, kad SPA galima sieti su lotyniškais posakiais: Senare per aquam. Sanum per aquam. Salute per aquam. Solus per aquam ir t. t.*

Beekvivalenčiai žodžiai problemiški ne tik dėl to, kad neturi atitikmenų vertimo kalboje. Jie išryškina ir kitas spręstinas mašininio vertimo problemas. Pavyzdžiui,



beekvivalentis žodis gali būti tik viena daugiareikšmio žodžio reikšmė. *Garter* yra daugiareikšmis daiktavardis, kurio pirmoji reikšmė iš tiesų yra „guma, keliaraištis“. Tačiau žodis taip pat žymi ir ordino tipą, kurį nešioja Keliaraiščio ordino nariai. Šis ordinas kilęs iš riterių ordino viduramžių Anglijoje, o dabar simbolizuoja pačius aukščiausius Jungtinės Karalystės kilmingųjų sluoksnius, kuriems priklauso tik Velso Princas ir dvidešimt keturi nariai, vadinamieji kompanionai.

Kuriozišku galima pavadinti *Georgian urn* vertimą. Tiek vertėjas, tiek mašininio vertimo programa *Georgian* interpretuoja kaip „gruzinių“. Pastaruoju atveju *urn* verčiama „laidojimo urna“. Taigi turime du variantus – „gruzinių laidojimo urną“ ir „gruzinišką arbatinuką“. Tačiau žinant sakinio kontekstą (Londono aukštuomenės namuose į kambarį įnešama arbata), mažai tikėtina, kad kalbama apie gruzinų arbatinuką. Galbūt vertėjui derėjo labiau pasidomėti ir sužinoti, kad *Georgian* taip pat nusako Didžiosios Britanijos karalių Jurgių I–IV epochą, trukusią nuo 1714 iki 1830 m., kurios metu mene kaip tik vyravo neoklasikinis stilius. Galbūt spraga, atsiradusi dėl negebėjimo perteikti „Jurgio epochos“, galėjo būti kompensuota kitaip, neutraliu žodžiu *arbatinuką* ar žodžių junginiais *senovinį arbatinuką*, *Jurgio laikų arbatinuką*. Taigi matome, kad realijų vertimas kelia sunkumų ir vertėjams.

Daiktavardis, *guinea* taip pat yra tipiškas kelių probleminių vertimo atvejų pavyzdys. Pirmiausia, tai realiją – šalies valiutą – įvardijantis daiktavardis, o antra, tai ir šalies pavadinimas, tad turime tikrosios homonimijos atvejį.

Beekvivalentis žodis gali būti vartojamas stabiliaame žodžių junginyje, kolokacijoje, ir nulemti pažodinę vertimo kalbos struktūrą, kurią generuoja mašininio vertimo programa. Pavyzdžiui, iš indų kultūros atkeliavęs daiktavardis *sutee* ir reiškiantis „suti, našlės susideginimą su vyro lavonu“, vartojamas kolokacijoje *to commit a sutee*, kuri išverta „pavesti sutee“, t. y. pavartota viena iš veiksmazodžio *to commit* reikšmių ir paliktas neišverstas nežinomas žodis.

Išvados

Atlikus kokybinį anglų – lietuvių kalbų mašininio vertimo sistemos darbo tyrimą, nepasitvirtino hipotezė, kad mašininio vertimo kokybė priklauso nuo verčiamo teksto stiliaus specifikos. Tam įtakos galėjo turėti programoje vartojamas bendrojo pobūdžio žodynas, neorientuotas į jokią specializuotą sritį. Kita vertus, tyrimo rezultatus reikėtų vertinti labai atsargiai. Neatmestina ir tai, kad galbūt anglų – lietuvių kalbos vertimo programa nėra pakankamai gera tam, kad būtų galima rasti stiliaus vertimo skirtumų.

Leksinių mašininio vertimo klaidų analizė parodė, kad jas lemia įprasti sąvokų įvardijimo reiškiniai: daugiareikšmiškumas, homonimija, sąvokų įvardijimas žodžių junginiais ar jų neįvardijimas. Klaidoms atsirasti įtakos turi ir programos vartojamo žodyno dydis ir struktūra, programos vertimo modelis (pažodinis vertimas).

Klaidų analizė tik parodo kryptis, kaip turi būti tobulinama programa. Akivaizdu, kad beekvivalentių žodžių vertimas nėra didžiausia bėda. Anglų – lietuvių kalbų mašininio vertimo programą reikia tobulinti: pateikti papildomų reikšmių ir homonimų žymų, kur įmanoma, pateikti tam tikrų taisyklių ir nuolat pildyti žodyną.



Literatūra

- Anglų – lietuvių kalbų mašininio vertimo sistema: <http://vertimas.vdu.lt/twsas/>.
- Ambrasas-Sasnava 1978 – Ambrasas-Sasnava K. *Vertimo mokslas*. Vilnius: Mokslas.
- Armalytė 1990 – Armalytė O., Pažūsis L. *Vertimo teorijos pradmenys*. Vilnius: VU.
- Arnold 1994 – Arnold D. J. et al. *Machine Translation: and Introductory Guide*. London: Blackwells-NCC.
- Calude 2003 – Calude A. S. Machine Translation of Various Text Genres <http://www.calude.net/andreea/MT.pdf>.
- Daudaravičius 2006 – Daudaravičius V. Pradžia į begalybę. Mašininis vertimas ir lietuvių kalba. *Darbai ir dienos* (45), 7–18.
- Gross 1992 – Gross, A. Limitations of computers as translation tools. *Computers in translation – a practical appraisal*, ed. J. Newton, London: Routledge, 96–130.
- Gudavičius 2000 – Gudavičius A. *Etnoligvistika*. Šiauliai: Šiaulių universitetas.
- Gudavičius 2007 – Gudavičius A. *Gretinamoji semantika*. Šiauliai: Šiaulių universiteto leidykla.
- Marcinkevičienė 2006 – Marcinkevičienė R. Konceptualioji metafora vertime. *Darbai ir dienos* (45), 109–118.
- Pikčilingis 1971 – Pikčilingis J. *Lietuvių kalbos stilistika I*. Vilnius: Mintis.
- Rimkutė 2007 – Rimkutė E., Kovalevskaitė J. Mašininis vertimas – greitoji pagalba globalėjančiam pasauliui. *Gimtoji kalba*. Nr. 9.
- Rimkutė 2008 – Rimkutė E., Kovalevskaitė J. Linguistic Evaluation of the First English-Lithuanian Machine Translation System. *Proceedings of the Third Baltic Conference on Human Language Technologies*. Vytautas Magnus University, Institute of the Lithuanian Language, 257–264.
- Valstybinė lietuvių kalbos komisija: <http://www.vlkk.lt/>
- Vlachov 1980 – Влахов С., Флорин С. *Непереводимое в переводе*. Москва: Международные отношения.
- White 2003 – White J. S. How to evaluate machine translation. *Computers and translation – a translator's guide*. Ed. H. Somers, Amsterdam, Philadelphia: John Benjamins, 211–244.
- Župerka 1997 – Župerka K. *Stilistika*. Šiauliai: ŠU.



Jurgita Cvilikaitė

LEKŠINĖS MAŠININIO VERTIMO KLAIDOS: BEEKVIVALENČIŲ ŽODŽIŲ VERTIMAS

Santrauka

Pagrindiniai žodžiai: *mašininis vertimas, daugiareikšmiškumas, homonimija, realija, lakūna, beekvivalenčiai žodžiai.*

2008 m. pradžioje pristatytą internetinę anglų – lietuvių kalbų mašininio vertimo programą jau išbandė daug vartotojų, buvo ir esama visokių nuomonių. Straipsnyje pateikiami kai kurie kokybiniai šios programos darbo tyrimo rezultatai, aptariamos svarbesnės problemos. Daugiau dėmesio skirta beekvivalenčių žodžių vertimui. Aptarti terminai *realija, lakūna* ir *beekvivalenčiai žodžiai*. Pateikiami argumentai, kodėl pastarasis terminas yra tinkamiausias. Tyrimo rezultatai atskleidė daugialypę automatizuoto beekvivalenčių žodžių vertimo problematiką.

Tyrimo pradžioje kelta hipotezė, kad mašininio vertimo kokybė priklauso nuo verčiamo teksto stiliaus, nepasitvirtino. Tam įtakos galėjo turėti programos naudojamas bendro pobūdžio žodynas. Kita vertus, tokius rezultatus reikėtų interpretuoti atsargiai. Gali būti, kad anglų – lietuvių kalbų mašininio vertimo sistema nėra gera ir negali identifikuoti skirtingo stiliaus vertimo skirtumų. Mašininio vertimo leksikos klaidų analizė rodo, kad jas lemia įvardijimo reiškiniai: daugiareikšmiškumas, homonimija, įvardijimas žodžių junginiais ir tam tikrų sąvokų neįvardijimas. Vertimo klaidų atsiranda ir dėl žodyno dydžio, jo struktūros, vertimo principo (dažniausiai pažodinis vertimas).

Mašininio vertimo leksikos klaidų analizė parodo tik kryptis, kuriomis turėtų būti gerinama programa. Savaimė suprantama, beekvivalenčių žodžių vertimas nėra didžiausia programos problema. Anglų – lietuvių kalbų mašininio vertimo programą reikia tobulinti: išplėsti žodyną, įtraukti daugiau frazių ir pažymų apie žodžio reikšmių vartosenos sritis; sukurti taisykles, leidžiančias atskirti anglų kalbos homonimus ir parinkti jiems tinkamus lietuvių kalbos atitikmenis.

Jurgita Cvilikaitė

LEXICAL MACHINE TRANSLATION MISTAKES: TRANSLATION OF NON-EQUIVALENT WORDS

Summary

Keywords: *machine translation, polysemy, homonymy, realia, lacuna, non-equivalent words.*

The on-line machine translation programme, which was created by Vytautas Magnus University and introduced at the beginning of 2008, translates texts from English into the

Philologia 2008 (13)



Lithuanian language. Many users have already tried this programme. It is natural that it is both praised and criticized. The article presents a quality research of this programme. The main problems of the programme are discussed. More attention is paid to the complicated translation of non-equivalent words. Terms “*realia*”, “*lacuna*” and “non-equivalent words” are discussed; arguments in favour of the latter term are provided. The research data shows many-sided problems of machine translation of non-equivalent words.

The raised hypothesis for the research that machine translation quality depends on the text style was not confirmed. This might have been due to the non-specialized dictionary of the programme. On the other hand, the results of the research should be interpreted with care. It is possible that the English – Lithuanian machine translation programme is not good enough so that style translation differences could be identified. The analysis of lexical machine translation mistakes showed that the mistakes are conditioned by ordinary phenomena of nomination: polysemy, homonymy, nomination by multiword units and the lack of nomination. The size and structure of the dictionary, as well as the model of translation (literal translation), also influence the appearance of translation mistakes.

The analysis of the mistakes only shows the directions in which the machine translation programme should be improved. The translation of non-equivalent words is obviously not the major problem of the programme. The English – Lithuanian machine translation programme has to be improved: additional sense and homonym notations should be added; where possible, rules should be provided; the dictionary should be continuously expanded with new words.

