

MAIK platformos „Coursera“ turinio postredagavimas

Dalia Lapinskaitė, Dalia Mankauskienė

Vilniaus universitetas
Filologijos fakultetas
Literatūros, kultūros ir vertimo tyrimų institutas
Institute for Literary, Cultural and Translation Studies
Faculty of philology
Vilnius University
dalia.lapinskaite@yahoo.com
<https://orcid.org/0000-0002-0895-5413>
dalia.mankauskiene@ff.vu.lt
<https://orcid.org/0000-0003-2806-0892>

Santrauka. Straipsnyje pristatomi MAIK (masinių atvirųjų internetinių kursų) platformos „Coursera“ mokomojo turinio, išversto mašininio vertimo (MV) sistema „Smartling“, postredagavimo ypatumai. Didžioji dalis platformos „Coursera“ turinio pateikiama anglų kalba ir tai yra viena iš priežasčių, kodėl įsitraukimas į šiuos kursus Lietuvoje yra mažas. Augant internetinių išteklių paklausai, poreikis išversti kursus į lietuvių kalbą tapo akivaizdus ir tam vis dažniau pasitelkiamos MV sistemos. Straipsnyje aprašyti eksperimento, atlikto su MV sistema „Smartling“, rezultatai. Eksperimente dalyvavo 10 dalyvių – 6 profesionalūs ir 4 neprofesionalūs vertėjai, kurie postredagavo platformos „Coursera“ kurso „Gerovės mokslas“ ištrauką. Postredagavimo procesas stebėtas naudojant įrankį „Translog-II“, fiksuojantį dalyvių klavišų paspaudimus. Straipsnyje pateikta MV klaidų klasifikacija ir jų dažnumas. Vienas iš svarbiausių MV sistemos „Smartling“ postredagavimo ypatumų yra teksto skaidymas į subtitrų eilutes, dėl kurio ir kyla dauguma gramatinių klaidų. Tarp klaidų, kurias lėmė ne skaidymas, dažniausios – MV klaidos, atsirandančios dėl žodžių daugiareikšmiškumo, pažodinio vertimo ir įvardžių vartosenos. Po postredagavimo užduoties dalyviai atsakė į klausimyno klausimus apie sistemos „Smartling“ funkcionalumą. 7 iš 10 tyrimo dalyvių šią sistemą įvertino patenkinamai. Tyrimo rezultatai parodė, kad sistema „Smartling“ nėra pakankamai pritaikyta lietuvių kalbai, todėl postredaguodami vertėjai turi įdėti daug kognityvinių pastangų.

Pagrindiniai žodžiai: postredagavimas, mašininis vertimas, „Coursera“, „Smartling“, mašininio vertimo klaidos

MOOC Coursera Content Post-editing

Abstract. This paper presents the post-editing features of the machine translation (MT) system *Smartling* used to translate the learning content of MOOC (Massive Open Online Course) Coursera. Most of the Coursera content is delivered in English, which is one of the reasons for the low uptake of these courses in Lithuania. With the growing demand for online resources, the need to translate courses into Lithuanian has become evident and MT systems are increasingly used for that purpose. This paper describes the results of an experiment carried out with the *Smartling* MT system. The experiment involved 10 participants, 6 professional and 4 non-professional translators, who post-edited a passage from the Coursera course *The Science of Wellbeing*. The post-editing process was monitored using the Translog-II tool, which captures the participants' keystrokes. The paper presents the classification and frequency of MT errors. One of the most important post-editing features of the *Smartling* MT system is the splitting of the text into subtitle lines, which is the cause of most grammatical errors. Among the errors not attributable to this text division are those caused by the polysemy of the words, literal translation and the use of pronouns. After the post-editing task, participants filled in a short questionnaire about the functionality of the *Smartling* system. 7 out of 10 participants rated the performance of this system as satisfactory. The results of the study showed that *Smartling* is not sufficiently tailored to the Lithuanian language, and that translators have to use a lot of cognitive effort in post-editing.

Keywords: post-editing, machine translation, Coursera, Smartling, machine translation errors

Įvadas

Per pasaulinę pandemiją žymiai išaugo poreikis mokytis nuotoliniu būdu. Nuo 2019 m. masinių atvirų internetinių kursų (MAIK) paslaugą teikiančios platformos „Coursera“ dalyvių skaičius padidėjo 32 proc. ir 2022 m. viršijo 100 milijonų¹. Siekdama pašalinti neigiamus COVID-19 padarinius, „Coursera“, bendradarbiaudama su daugiau nei šimto šalių vyriausybėmis, pristatė Darbo jėgos gaivinimo iniciatyvą (angl. *Workforce Recovery Initiative*), pagal kurią bedarbiams teikiama nemokama prieiga prie mokymų ir galimybė persikvalifikuoti (Vandenbosch 2020). Šioje iniciatyvoje dalyvauja ir Lietuvos užimtumo tarnyba². Labai svarbu, kad aukščiausio lygio pasauliniai kursai Lietuvoje būtų prieinami gimtąja kalba, todėl yra poreikis kuo greičiau išversti kiek galima daugiau siūlomų kursų, kad lietuviai galėtų sužinoti apie naujausius tyrimus, technologijas, gilinti savo žinias, kelti kvalifikaciją, plėsti akiratį ir tapti lygiaverčiais globalios rinkos dalyviais.

Vis dėlto Lietuvos gyventojų įsitraukimas yra mažas³, todėl kuo geresnės kokybės kursų vertimas bus pateiktas vartotojui, tuo didesnė tikimybė, kad jų lankomumas išaugs. „Coursera“ siūlo kursų vertimą į lietuvių kalbą atlikti vertimo sistemoje

¹ <https://www.classcentral.com/report/coursera-q1-2022/>

² <https://uzt.lt/news/norintiems-tobuleti-nemokami-coursera-org-kursai/>

³ <https://globalhighered.files.wordpress.com/2013/11/courseramapoct2013.jpg>

„Smartling“. Straipsnyje aprašomi platformos „Coursera“ mokymo medžiagos, verstos sistema „Smartling“ iš anglų kalbos į lietuvių kalbą, vertimo ypatumai ir eksperimentas, atliktas siekiant palyginti profesionalių ir neprofesionalių vertėjų postredagavimo procesą ir suklasifikuoti jų aptiktas MV klaidas. Vertėjų postredagavimo proceso eiga ir atlikti mašininio vertimo pakeitimai analizuoti naudojant klavišų paspaudimų registravimo įrankį „Translog-II“ (Jakobsen 2011) empiriniu tyrimo metodu. Be to, vertėjai atsakė į klausimus apie sistemos „Smartling“ funkcionalumą ir kokybę. Šio tyrimo tikslas – nustatyti pagrindines ir dažniausiai pasitaikančias MV klaidas verčiant platformos „Coursera“ turinį iš anglų kalbos į lietuvių kalbą vertimo sistema „Smartling“, įvertinti šių vertimų postredagavimo ypatumus.

Sutelktinis „Coursera“ turinio vertimas

Sutelktinės veiklos (angl. *crowdsourcing*) terminą 2006 m. pasiūlė žurnalo „Wired“ autorius Jeff Howe, norėdamas apibūdinti augančią tendenciją, kai laisvu nuo darbo metu žmonės kuria turinį, sprendžia problemas ar net atlieka tyrimus įvairioms įmonėms, pavyzdžiui, farmacijos ar televizijos srityse (McDonough Dolmaya 2012). Technologijos ir mėgėjiško vertimo simbiozės rezultatas yra sutelktinis vertimas – situacija, kai vertėjai savanoriai bendradarbiauja versdami tam tikrą internetinį turinį. Vertėjai savanoriai dažniausiai yra neprofesionalai, norintys prisidėti prie įvairiausio turinio tekstų vertimo dėl savo pažiūrų ar humanistinių iniciatyvų (Jiménez-Crespo 2017). Milijonai tokių savanorių vertėjų padeda susidoroti su vertimo iššūkiais tiek verčiant MV sistemomis, tiek MV redaguojant, nes 2022 m. net 60,4 proc. interneto turinio buvo anglų kalba (Arfaoui 2022). Nuo 2011 m. šių vertėjų paslaugomis naudojasi šimtai įmonių, kad patenkintų augančius vertimo poreikius, tarp jų – ir tokios milžinės, kaip TED, „Twitter“, „Facebook“ (Kelly, Ray and DePalma 2021).

„Coursera“ turi gausią vertėjų savanorių bendruomenę visame pasaulyje. Jie verčia angliškų paskaitų vaizdo įrašų subtitrus į 40 pasaulio kalbų. Anot platformos „Coursera“ kūrėjų, sutelktinis vertimas yra vienintelė vertimo galimybė, kai reikia versti tokios didelės apimties turinį (Hu, O’Brien and Kenny 2020). Vertėjai savanoriai gali versti jiems patogiu metu ir tik tiek, kiek jie nori. „Coursera“ motyvuoja vertėjus apdovanojama juos ženkleliais. Jie skiriami už išverstus žodžius: pirmas ženklelis skiriamas išverius pirmus 15 000 žodžių, antras – 50 000, trečias – 100 000 žodžių. Tai tarptautinio pripažinimo įvertinimas – vertėjas gali apie savo pasiekimus paskelbti „LinkedIn“ paskyroje ar savo CV. Vertėjai, verčiantys į lietuvių kalbą, yra pavieniai vertėjai, nesusibūrę į bendruomenę. Nors „Coursera“ siekia, kad išverstus tekstus redaguotų redaktoriai, lietuvių kalbos redaktorių nėra.

MV sistema „Smartling“

Didelės MAIK turinio apimties vertimams naudojamos specialios platformos, kurios padeda organizuoti vertimo eigą, stebi vertimo procesą ir suteikia vertėjams galimybes naudotis naujausiais vertimo įrankiais. Verčiant platformos „Coursera“ mokymo turinį naudojama MV sistema „Smartling“. Tai debesijos vertimų valdymo sistema (angl. *a cloud-based Translation Management System, TMS*), siūlanti visą įvairovę funkcijų: patikimą turinio valdymą, naujų įrankių integraciją su jau esamais, mašininį mokymąsi (angl. *machine learning*) ir kt. Verčiant darbas vyksta „Smartling“ paskyroje, naršyklėje.

Pasirinkęs norimą versti kursą, vertėjas atsidaro sistemoje „Smartling“ vertimui atlikti paruoštą langą, kuriame mato vaizdo įrašą, originalo tekstą, dešiniajame viršutiniame polangyje siūlomą mašininį vertimo variantą. Paskaitos tekstas suskirstytas eilutėmis, paruoštas subtitruoti. Vaizdo įrašas skamba tiek laiko, kol pasakomas vienos eilutės, kurią vertėjas verčia, tekstas. Kairėje pusėje pateikiamas tekstas originalo kalba, o dešinėje vertėjas spausdina verčiamą tekstą ir jį patvirtina. Vertėjas gali pasinaudoti siūlomu MV variantu (gali būti net keletas siūlymų), jį nukopijuoti į išversto teksto eilutę ir patvirtinti. Sistema automatiškai išverčia anksčiau tekste pasitaikiusius identiškus sakinius.

Programa „Smartling“ vertimą ir tikrina. Pavyzdžiui, jei angliškas tekstas į kitą kalbą išverčiamas labai trumpais sakiniais, vertėjas tokio vertimo negali išsaugoti. Vertimą galima išsaugoti tik patvirtinus, kad trumpesnė išversto teksto eilutė tinkamai perteikia ilgesnės originalo teksto eilutės prasmę. Svarbu pažymėti, kad verčiant tam tikrą kursą platformoje „Coursera“ galima naudotis ir kitų vertėjų vertimais ieškant atitikmenų vertimo atmintyje. Dar viena svarbi ypatybė verčiant sistemoje „Smartling“ – subtitravimas. Simbolių skaičius kiekviename segmente yra ribotas. Segmente gali tilpti dvi subtitrų eilutės. Vertėjas turi pats pažymėti, kur įterpti tarpelio simbolį, ir perskirti tekstą į eilutes.

MV postredagavimas

Sparčiai augant interneto vartotojų skaičiui ir informacijos prieinamumui, postredagavimo poreikis nuolat didėja. Tai patvirtino ir 2017 m. išleistas tarptautinis postredagavimo standartas ISO 18587, kuriame nurodoma, kad postredagavimas – tai mašininio vertimo redagavimas ir taisymas. O'Brien (2011:1) siūlo kiek platesnę postredagavimo apibrėžtį: „mašininio būdu išversto neapdoroto teksto taisymas, kurį vertėjas atlieka vadovaudamasis konkrečiomis gairėmis ir kokybės kriterijais“. Taigi, postredagavimo veikla priešpriešinama su žmogaus atliekamu teksto vertimu ir MV sistemos apdoroto teksto postredagavimas kartais įvardijamas kaip atskira vertimo forma, nes, nors teksto

šaltiniai, skaitymo ir rašymo procesai skiriasi, vertėjui reikia atlikti tuos pačius veiksmus kaip ir verčiant. Vis dėlto postredaguojant nereikia versti originalo teksto nuo pradžių, nes MV sistemos perima darbą vertimo juodraščio parengimo stadijoje (do Carmo 2017; Moorkens and O'Brien 2017). Tačiau taip pat galima teigti, kad MV postredagavimas ir vertėjo vertimas yra du skirtingi procesai. Koby (2001) išskyrė tris vertimo ir postredagavimo skirtumus:

- klaidų dažnumas;
- klaidų pasikartojimas – versdamas tekstą vertėjas gali suklysti ir žodį išversti neteisingai vieną kartą, o MV apdorotame tekste tas pats žodis gali būti išverstas neteisingai daug kartų, nes taip jis įvestas sistemoje;
- klaidų tipas – pvz., Nitzke (2020: 33) teigimu, MV retai pasitaikys spausdinimo ar rašybos klaidų, tačiau bus leksikos, sintaksės ir lokalizavimo klaidų.

Versdamas vertėjas daugiau dėmesio skiria originalo teksto skaitymui ir suvokimui, bando įsigilinti (Carl et al. 2011: 138). Tyrėjai pastebėjo, kad postredaguodami vertėjai laiko originalo teksto skaitymui skiria mažiau – jų dėmesio išlaikymo trukmė ir žvilgsnio fiksacija skaitant originalo teksto segmentus yra trumpesnė. Tyrimas atskleidė, kad postredagavimas reikalauja mažiau laiko sąnaudų nei vertimas, todėl įvardijus dažniausias MV klaidas vertėjai gali iš anksto pasiruošti ir numatyti pataisymus (Carl et al. 2011: 141).

Postredagavimas sumažina originalo teksto vertimui atlikti reikalingą laiką ir pastangas. Visgi, ieškodami atsakymo į klausimą, ar geriau versti originalo tekstą nuo pradžių pačiam vertėjui, ar geriau postredaguoti MV apdorotą tekstą, vertimo studijų tyrėjai siekia išmatuoti vertimui atlikti reikalingas pastangas. Krings (2001: 178) išskyrė šias pastangas:

- laiko pastangos (angl. *temporal effort*) – kiek reikia laiko postredagavimo procesui, t. y. kiek užtrunka sutvarkyti MV sistemos apdorotą tekstą, kad jis taptų aukštos kokybės tekstu vertimo kalba;
- techninės pastangos (angl. *technical effort*) – kiek reikia laiko pataisymams atlikti, t. y. klavišų paspaudimai keičiant žodžių tvarką, įterpiant ar išbraukiant žodžius;
- kognityvinės pastangos (angl. *cognitive effort*) – koks kognityvinis krūvis tenka identifikuojant klaidas ir atliekant pataisymus.

Šios pastangos yra konceptualiai skirtingos, tačiau kartu ir susijusios: laiko pastangos priklauso nuo kognityvinių ir techninių pastangų, o techninis procesas – nuo kognityvinių pastangų (Krings 2001, 178). Pavyzdžiui, vienas tyrimas parodė, kad vertėjai sugaišta daugiau laiko taisydami MV sakinius, reikalaujančius daugiau kognityvinių pastangų (Koponen et al. 2012).

MV kokybės vertinimas

Vertėjo pastangos postredaguojant MV priklauso ir nuo vertimo kokybės. Vienam vertimo atvejui gali būti siūdomi keli teisingi ir neteisingi variantai, todėl vertimo sistemoms sudėtinga kiekvieną kartą pasirinkti tinkamą. Vertėjas taip pat turi gebėti įvertinti, kuriais siūlomais vertimo variantais galima pasikliauti, o į kuriuos reikėtų žiūrėti įtariai ir patikrinti. Vertėjas, žinodamas dažniausiai pasitaikančias MV klaidas ar iššūkius, postredagavimą gali atlikti greičiau ir našiau.

MV kokybės vertinimas – taip pat sudėtingas procesas. Išskiriami du pagrindiniai būdai, kuriais galima vertinti MV kokybę: automatinės priemonės ir žmogaus vertinimas (Farrús et al. 2010). Pastaruoju atveju dažniausiai vertimas pateikiamas dvikalbiams vertintojams, mokantiems ir originalo, ir vertimo kalbą, kad jie suteiktų konkrečiam segmentui kokybės balą, pvz., nuo 1 = prastas iki 5 = puikus (žr. Guzmán et al. 2015). Paprastai naudojami šie kriterijai: adekvatumas, t. y. prasmės išsaugojimas; sklandumas, t. y. gramatiškumas; bendra kokybė (pagrįsta abiejų kriterijų deriniu) ir numatomos postredagavimo kognityvinės pastangos (Popović 2018). Skirtingus to paties teksto šaltinio MV variantus taip pat galima lyginti reitinguojant (Callison-Burch et al. 2007), t. y. vertintojas turėtų nurodyti, ar geresnė kiekvieno šaltinio sakinio A, ar B versija, nesuteikdamas jokio absoliutaus balo.

Žmogaus vertinimas dažnai trunka labai ilgai, gali būti labai brangus ir nenuoseklus procesas, todėl MV vertinimui ypač aktualūs automatinio vertinimo rodiklių metodai. Šių metodų tikslas – įvertinti MV sistemos ir etaloninio vertimo (paprastai žmogaus vertimo) panašumus ir skirtumus pagal iš anksto nustatytus lingvistinius kriterijus (Koehn 2009). Vieni pirmųjų 2002 m. automatinio vertinimo metodą pristatė mokslininkai Kishore Papineni, Salim Roukos, Todd Ward ir Wei-Jing Zhu. Tai iki šiol populiarī metrika – BLEU (angl. *Bilingual evaluation understudy*) (Papineni et al. 2002). Vėliau sukurta ir daug kitų, pavyzdžiui, NIST metrika (Doddington 2002), GTM (angl. *General Text Matcher*) (Turian, Shen and Melamed 2003), taip pat TER (angl. *Translation Error Rate*) ir HTER (angl. *Human-targeted TER*) (Snover et al. 2006), METEOR (Lavie and Agarwal 2007), „MaxSim“ (Chan and Ng 2008), ULC (Giménez and Márquez 2008), RTE (Padó et al. 2009), „wpBleu“ (Popović and Ney 2009), „chrF“ (Popović 2015).

Vis dėlto automatinės priemonės nesuteikia pakankamai informacijos apie tiriamą vertimo sistemą, jos stiprybes ir silpnybes. Dėl šių priežasčių atsirado klaidų klasifikacijos, kuriomis vadovaujantis nustatomos ir klasifikuojamos klaidos vertimo tekste, siekiant apibrėžti klaidų klases ir suteikti daugiau informacijos, reikalingos sistemoms tobulinti. Klaidų analize dažniausiai siekiama nustatyti klaidų pasiskirstymą pagal nustatytas klaidų klases, klaidų pasiskirstymą kiekvienoje klaidų klasėje, skirtingų klaidų tipų poveikį įvairiems postredagavimo aspektams ir pan. (Popović 2018).

Kaip ir automatinio vertinimo, taip ir klaidų klasifikacijos sistemų sukurta ne viena (Font Llitjós, Carbonell and Lavie 2005; Vilar et al. 2006; Temnikova 2010; Kirchoff, Capurro and Turner 2012; Federico et al. 2014; Lommel et al. 2014; Daems et al. 2017). Klaidos gali būti klasifikuojamos ir pasitelkus jau minėtus automatinio vertinimo rodiklių metodus. Šiame tyrime svarbu paminėti, kad postredagavimą galima laikyti netiesiogine klaidų anotacija, nes kiekvienas pakeitimas iš tikrųjų yra klaidos taisymas (Popović 2018: 17– 18). Tačiau, nors postredagaujant atliekami pakeitimai, kurie, redaktoriaus nuomone, yra būtini, jie nebūtinai visada susiję su tikromis klaidomis. Tyrimuose pastebėta, kad postredagaujant kartais redaguojama per daug, atliekant laisvuosius pakeitimus (angl. *preferential changes*) tais atvejais, kai MV jau buvo teisingas prasmės ir taisyklingas gramatikos požiūriu (de Almeida 2013; Koponen, Salmi and Nikulin 2019; Nitzke and Gros 2020). Pavyzdžiui, de Almeida tyrimo (2013) nustatyta, kad laisvieji pakeitimai sudaro nuo 16 iki 25 proc. pakeitimų, o Koponen et al. (2019) rezultatai parodė, kad 26 proc. redagavimų buvo nereikalingi.

Šiame tyrime remiamasi Petkevičiūtės ir Taulyno (2011) pasiūlyta MV iš anglų kalbos į lietuvių kalbą klaidų klasifikacija, pagal kurią klaidos skirstomos į dvi grupes: lingvistines ir sistemes klaidas.

- Lingvistinės klaidos:
 - morfologijos (gramatikos) klaidos – tai netaisyklingas skaičiaus, giminės, linksnio, veiksmažodžių asmens ir formos derinimas, klaidingas kalbos dalies parinkimas, prielinksnių ir prieveiksnių vertimas;
 - leksikos klaidos – tai neišversti žodžiai, parinkta netinkama daugiareikšmio žodžio reikšmė, neteisingas sutrumpintų žodžių, įvardžių, santrumpų, žodžių, sujungtų brūkšneliu, vertimas, pažodinis posakių ir frazeologizmų vertimas.
- Sisteminės klaidos – tai žodynų, programos kodų klaidos, kurių lingvistinio, o kartais ir argumentuoto loginio paaikškinimo nėra.

Tyrimo dalyviai ir tyrimo eiga

Siekiant nustatyti MV sistema „Smartling“ išversto platformos „Coursera“ turinio postredagavimo ypatumus buvo parinkta „Smartling“ išversta ištrauka iš vieno populiariausių platformos „Coursera“ kursų – „Gerovės mokslas“ (angl. *The Science of Well-Being*⁴). Šį kursą dėsto Jeilio universiteto psichologijos profesorė Laurie R. Santos. Jo reitingas 5 žvaigždučių sistemoje vertinamas 4,9 balo pagal 370 tūkst. įvertinimų. Į šiuos kursus užsirašė 3,8 mln. žmonių. Ištraukos ilgis – 445 žodžiai (43 subtitrų eilu-

⁴ <https://www.coursera.org/learn/the-science-of-well-being>

tės). Šios ištraukos ilgis leido dalyviams susidaryti įspūdį apie paskaitos turinį, kontekstą ir dėstytojos kalbėjimo stilių.

Tyrime dalyvavo 10 dalyvių – 6 profesionalūs vertėjai ir 4 anglų kalbą gerai mokantys asmenys. Profesionalūs vertėjai turėjo nuo 10 iki 40 metų vertimo patirties įvairiose srityse. Neprofesionalių vertėjų grupėje dalyvių anglų kalbos žinios atitiko B2–C1 kalbos mokėjimo lygį, tačiau šie asmenys ankstesnės vertimo patirties neturėjo. Visų dalyvių gimtoji kalba – lietuvių.

Duomenys tyrimui buvo surinkti per 10 individualių sesijų, trukusių po 30 min. Dalyvių užduotis buvo postredaguoti MV sistemos apdorotą tekstą ir užpildyti trumpą klausimyną (žr. 2 priedą). Pradžioje jie peržiūrėjo paskaitos vaizdo įrašo ištrauką, kad geriau suprastų kontekstą, tada programoje „Translog-II“ postredagavo vertimo sistemos „Smartling“ atlikto MV tekstą. Programa „Translog-II“ stebi ir registruoja vertėjo elgesį, vertimo trukmę, verstinio teksto kūrimo ir rašymo dinamiką. Klavišų registratorius veikia fone ir netrukdo rašymo procesams. „Translog-II“ fiksuoja tikslų vertimo proceso pradžios, pabaigos ir visų judesių laiką. Dalyviai, užbaigę postredagavimą, atsakė į klausimus apie sistemos funkcionalumą ir išsakė pastabas apie postredagavimo ypatumus. Dėl techninių kliūčių 2 profesionalių vertėjų postredagavimo rezultatų išsaugoti nepavyko, tačiau į jų nuomonę atsižvelgiama nagrinėjant atsakymus į klausimyno klausimus.

Nagrinėjant postredagavimo procesą, atlikta MV klaidų analizė lyginant MV ir postredaguotus teksto variantus. Analizei naudoti dalyvių atlikti klavišų paspaudimai, MV taisymai ir postredagavimo trukmė. Atsižvelgta ir į tai, ar taisymas rodo MV klaidą, ar yra laisvasis pakeitimas (laisvieji pakeitimai nebuvo skaičiuojami kaip MV klaida). Klaidos skirstytos remiantis literatūros apžvalgoje pateikiamu klaidingų kompiuterinio vertimo atvejų skirstymu į lingvistinių (gramatinių ir leksinių) ir sisteminių klaidų tipus (Petkevičiūtė and Tamulynas 2011). Taip pat analizuota, kiek profesionalūs ir neprofesionalūs vertėjai taisė MV tekstą, tačiau reikėtų paminėti, kad ištaisyto teksto kokybė šiame darbe nenagrinėjama: svarbiau, kad vertėjai norėjo konkretų segmentą taisyti, o ne tai, kaip jie jį ištaisė.

Tyrimo rezultatai

Išnagrinėjus „Translog-II“ pateiktus statistinius duomenis apie postredagavimo procesą, klavišų paspaudimus ir keitimus, galima išvelgti profesionalių ir neprofesionalių vertėjų aktyvumo skirtumų (1 lentelė). Duomenys rodo, kad neprofesionalūs vertėjai atliko daugiau veiksmų ir parašė daugiau teksto. Jie taip pat daugiau taisė. Tai reiškia, kad neprofesionalūs vertėjai skyrė daugiau techninių pastangų nei kognityvinių. Tai gali būti ir patirties trūkumo rodiklis, kai vertėjas abejoja savo sprendimais ir ieško tinkamo varianto iš karto rašydamas ir taisydamas tekstą, o ne gerai apsvarstęs.

1 lentelė. Profesionalių ir neprofesionalių vertėjų aktyvumo (klavišų paspaudimų) rodikliai

Rodiklis	Profesionalūs vertėjai	Neprofesionalūs vertėjai
Dalyvio veiksmas	nuo 775 iki 1552	nuo 1540 iki 1686
Teksto rašymas	nuo 437 iki 578	nuo 657 iki 839
Teksto trynimasis	nuo 74 iki 241	nuo 233 iki 344
Dalyvio veiksmas per minutę	nuo 32 iki 48	nuo 51 iki 61
Teksto rašymas per minutę	nuo 16 iki 18	nuo 22 iki 32

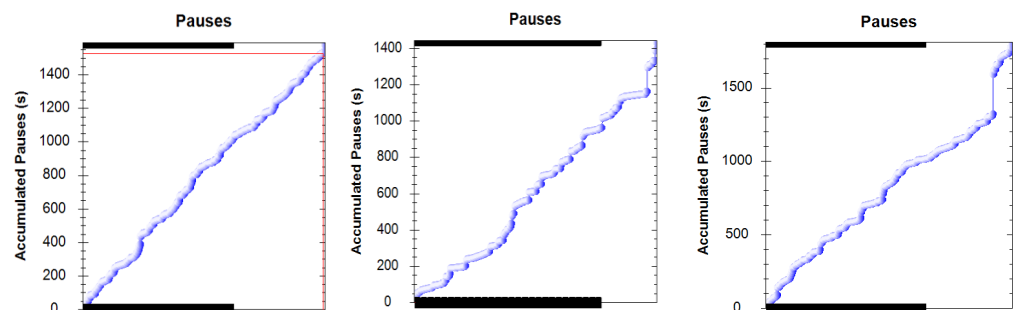
Kita vertus, profesionalūs vertėjai postredagavo mažiau eilučių nei neprofesionalūs vertėjai (2 lentelė). Nė vienas profesionalus vertėjas nespėjo suredaguoti viso teksto, iš neprofesionalių vertėjų tai pavyko padaryti tik 2, tačiau šie vertėjai mažiausiai taisė tekstą.

Remiantis šiais duomenimis ir Koponen et al. (2012) išvada, kad vertėjai sugaišta daugiau laiko taisydami MV sakinius, reikalaujančius daugiau kognityvinių pastangų, darytina prielaida, kad profesionalūs vertėjai daugiau laiko skyrė būtent kognityvinėms pastangoms. Tikėtina, kad profesionalūs vertėjai įtariau žiūri į MV ir stengiasi išlaikyti aukštą vertimo kokybę, daugiau dėmesio skirdami originalo ir vertimo tekstų analizei. Neprofesionalūs vertėjai daugiau pasitiki MV ir atlieka tik pačius būtiniausius taisyimus. Tai galima matyti ir iš postredaguotų eilučių skaičiaus rodiklio – neprofesionalūs vertėjai patikrino daugiau eilučių nei profesionalūs.

2 lentelė. Postredaguotų eilučių skaičius

Profesionalūs vertėjai		Neprofesionalūs vertėjai	
Prof1	32	Neprof1	43
Prof2	29	Neprof2	43
Prof3	13	Neprof3	33
Prof4	32	Neprof4	35

Analizuojant „Translog-II“ pateiktą informaciją apie pauzes, galima matyti, kuriose postredagavimo vietose vertėjai sugaišo daugiausia laiko (1 pav.).

**1 pav.** Pauzių sanauptų vaizdas

Pateiktose trijų skirtingų vertėjų pausių sanauų nuotraukose galima pastebėti, kad dinamika labai skirtinga. Dėl šios priežasties sudėtinga apibendrinti, kur vertėjai bendrai susidūrė su kognityviniais sunkumais, nes tai priklauso nuo kiekvieno dalyvio vertimo stiliaus.

MV klaidos

Postredaguotoje ištraukoje nustatytos 67 klaidos, tačiau nustatytų klaidų absoliutus skaičius nėra vienareikšmiškai teisingas, nes galima diskutuoti dėl kai kurių atvejų, pavyzdžiui, kaip turėtų būti vertinama viena iš naujo versta eilutė – ar skaičiuojami visi pakeisti žodžiai, ar viena eilutė, kaip pataisytas pažodinis vertimas ir pan. Visgi šiuo tyrimu nesiekama kiekybinio tikslumo, o norima sužinoti apytikslų klaidų santykį skirtingose kategorijose. Tuo labiau kad tokiam neilgame tekste negalima tikėtis daug visų tipų klaidų.

3 lentelė. MV klaidų klasifikacija

Klaidų kategorijos	Klaidų skaičius	Vidutiniškai klaidą taisyusių profesionalių vertėjų, %	Vidutiniškai klaidą taisyusių neprofesionalių vertėjų, %
Gramatinės klaidos	16	97	80
Giminė, skaičius, asmuo	8	96	81
Linksniai	6	100	88
Priešdėliai	1	75	50
Neigiami veiksmazodžiai	1	100	50
Leksikos klaidos	50	83	74
Pažodinis vertimas	24	82	73
Daugiareikšmiškumas	16	86	77
Įvardžiai	8	75	72
Idiomos	2	88	75
Sisteminės klaidos	1	100	100
Iš viso:	67	87	76

Kaip matyti 3 lentelėje, didžiausi kliuviniai atliekant postredagavimo užduotį buvo pažodinis vertimas, žodžių daugiareikšmiškumas ir įvardžių vartojimas. Giminės, skaičiaus ir asmens kategorijos iš pradžių vertintos atskirai, tačiau kadangi jos labai susijusios ir dažniausiai nulemtos tų pačių priežasčių, vėliau sujungtos į vieną bendrą MV klaidų kategoriją.

Kadangi ne visi vertėjai postredagavo visas eilutes, 3 lentelėje pateikiami vidutiniškai klaidą taisyusių vertėjų rodikliai. Kaip matyti, visi profesionalūs vertėjai ištaisė tik

linksnių klaidas ir vieną neigiamo veiksmožodžio atvejį, kitas klaidas taisė ne visi vertėjai. Duomenys rodo, kad neprofesionalūs vertėjai taisė mažiau visų kategorijų klaidų. Didžiausias skirtumas – giminės, skaičiaus, asmens kategorijoje: profesionalūs vertėjai vidutiniškai taisė 96 proc. atvejų, o neprofesionalūs vertėjai – 81 proc.; mažiausias – taisant įvardžių klaidas: profesionalūs vertėjai vidutiniškai taisė 75 proc. atvejų, o neprofesionalūs vertėjai – 72 proc.

Prieš pateikiant konkrečius įvairių klaidų pavyzdžius, reikėtų atkreipti dėmesį į tai, kad nemaža dalis lingvistinių (ypač gramatinių) klaidų kyla iš to, kad tekstas skaidomas į subtitrų eilutes ir kiekvieną eilutę sistema „Smartling“ verčia atskirai, neatsižvelgdama į tai, kad eilutėje pateikiama tik dalis sakinio. Pirmame pavyzdyje matyti, kad *geri pažymiai* verčiami vardininku todėl, kad MV požiūriu tai yra segmento pradžia, nesusijusi su ankstesniame segmente esančiu veiksmiu *seek*. Dėl tos pačios priežasties neišverstas ir neiginys, nurodantis, kad *nereikėtų siekti gerų pažymių*.

They both fall under the guideline of we should be seeking out not	Jie abu patenka į gairę, kad neturėtume ieškoti
good grades and not big salary but we should be seeking out healthier practices.	geri pažymiai ir nedidelis atlyginimas, bet turėtume ieškoti sveikesnės praktikos.

Kitas pavyzdys taip pat atskleidžia šią problemą, nes antroji eilutė neteisingai verčiama todėl, kad neatsižvelgiama į pirmoje eilutėje likusį prielinksnį *to*, nurodantį, kad *pratimų intervencija* siekiama tikslo – *padidinti gyventojų laimę*.

They tried to see whether or not they could use an exercise intervention to	Jie bandė išsiaiškinti, ar galėtų panaudoti pratimų intervenciją
really increase happiness in a population	tikrai padidina gyventojų laimę

Nepaisant šio iššūkio, visi tyrimo dalyviai atsižvelgė į teksto segmentavimą ir postredagavo priskirtų eilučių MV teksto suskirstymą. Tekstas lietuvių kalba atitiko tekstą priskirtame segmente anglų kalba. Vertėjai trumpino tekstą, trynė pasikartojančius žodžius, tačiau vertimo prasmė atitiko originalo tekstą visuose segmentuose. Vis dėlto tikėtina, kad tai turėjo įtakos jų nuomonei apie sistemą „Smartling“. Toliau pateikiami lingvistinių klaidų pavyzdžiai, kuriems skaidymas į subtitrų eilutes neturėjo įtakos.

Lingvistinės klaidos

MV sistemos apdorotame tekste pasitaikė nesuderintų linksnių, skaičiaus, asmens ir giminės atvejų, tačiau visi atvejai, išskyrus vieną, tikėtina, atsirado dėl teksto segmenta-

vimo. Šis atvejis, kuriam įtakos segmentavimas neturi, nors ir neteisingas, neturėtų būti laikomas MV klaida, nes tik iš to, kad ekrane kalba moteris, galima spręsti, jog versti reikėjo kitaip. MV sistema tokio konteksto negali žinoti.

Originalo tekstas	MV	Vertėjų taisymai
And this is the point where I get to be a terrible salesman in the course,	Ir tai yra taškas, kai aš galiu būti baisus pardavėjas kursuose,	siaubinga prekeivė; grėsminga siūlytoja; siaubinga pardavėja; įtaigi pardavėja; baisi pardavėja

Visus nesuderintų linksnių, skaičiaus, asmens ir giminės atvejus taisė tiek profesionalūs, tiek neprofesionalūs vertėjai, išskyrus vieną eilutę, kurioje netinkamą giminę taisė tik profesionalūs vertėjai. Tikėtina, kad neprofesionalūs vertėjai, taip pat kaip ir MV sistema, neatsižvelgė į platesnį kontekstą, kad sakydama *it's free* kalbėtoja kalba vis dar apie anksčiau paminėtą *happy pill*.

Originalo tekstas	MV	Vertėjų taisymai
It's not like some sketchy drug and it's even free .	Tai ne kaip koks nors eskizinis narkotikas ir netgi nemokamas	nemokama

Kita klaida, kurią taisė ne visi – netaisyklingas priešdėlių vartojimas. Lietuvių kalboje kai kuriais atvejais patartina vartoti veiksmažodį be priešdėlių, o MV pasiūlė vertimą su priešdėliu *pa-*, kuris šiame kontekste neatlieka savo funkcijos žymėti atliktinį veiksmą. Profesorė pataria *daryti* tam tikrus veiksmus, išsiugdyti įprotį, o ne vieną kartą *padaryti*. Šią klaidą taisė ne visi: 3 iš 4 profesionalių vertėjų ir 2 iš 4 neprofesionalių vertėjų.

Originalo tekstas	MV	Vertėjų taisymai
The final one is two sets of things I think you should be doing .	Paskutinis yra du dalykai, kuriuos, manau, turėtumėte padaryti .	daryti

MV sistemai sunku nustatyti giminės ir parinkti tinkamus įvardžius, nes anglų kalboje žodžiai nekaitomi giminėmis taip, kaip lietuvių kalboje. Ypač neaišku, kaip versti *it*. Šiuo atveju buvo daug kalbama apie *happy pill* ir vartojamas įvardis *it*. MV sistema visur jį vertė į *tai*. Pažymėtina, kad įvardis *you* gali būti neverstinas, kai kalbama apie žmones bendrąja prasme. Lietuvių kalboje jis taip pat turi dvi reikšmes – *tu* ir *jūs*. MV sistema rinkosi visais atvejais *you* versti *jūs*, išskyrus vieną atvejį, kai vertė *tu*.

Originalo tekstas	MV	Vertėjų taisymai
It will give you better grades even though you're not supposed to care about that,	Tai duos jums geresnius pažymius, net jei jums tai neturėtų rūpėti,	Ji suteiks; Ji lems; Tai atneš; Ji padės gauti; Jūs gausite
but it will make you look better	bet tai leis jums atrodyti geriau	ji padės atrodyti geriau; tai netgi pagerins jūsų išvaizdą; jūs atrodysite gražiau
even though you're not supposed to care about your looks	nors tu neturėtum rūpintis savo išvaizda	neturėtume; jūs neturėtumėte; jums neturėtų rūpėti
but it will do that before you do.	bet tai padarys anksčiau nei jūs .	ji tuo pasirūpins; ji padarys tai; tai atsitiks anksčiau; už jus tuo pasirūpins anksčiau; ji padarys greičiau nei jūs

Pažodinio vertimo klaidų pasitaikė daugiausiai. Šios kategorijos klaidų negalima pateisinti teksto segmentavimu. Tai, kad vertimo teksto sintaksė ar vartojama leksika atitinka originalo teksto sintaksę ir leksiką, neturėtų savaime būti laikoma klaida, tačiau kai vertimo tekstas skamba nenatūraliai ir jaučiama originalo teksto įtaka, nukentčia vertimo teksto sklandumas, o kartais ir suprantamumas.

Originalo tekstas	MV	Vertėjų taisymai
And this is the point where I get to...	Ir tai yra taškas , kai aš galiu būti bausis pardavėjas kursuose	Ir tai yra esminis momentas; Ir tai yra esmė; Ir tai yra vieta kursuose; Šiuo metu jūs galite pagyvoti; Ir šioje vietoje
They tried to see whether or not they could use an exercise intervention to	Jie bandė išsiaiškinti, ar galėtų panaudoti pratimų intervenciją	fizinį aktyvumą; mankštą; pratimus
Like, what could this horrible, wonderful thing be?	Pavyzdžiui, kas gali būti šis bausis, nuostabus dalykas?	Pavyzdžiui keista į Taigi, bausis, nuostabus – bausiai nuostabus, siaubingai nuostabus; siaubingas, nuostabus

Leksiniai pasirinkimai yra didelis iššūkis MV sistemai: neturint platesnio konteksto dažniausiai parenkama pirma žodžio reikšmė ir vertimo tekstas skamba ne-sklandžiai.

Originalo tekstas	MV	Vertėjų taisymai
where I could pretend that I have a happy pill ,	kur galėčiau apsimesti, kad turiu laimingą tabletę ,	laimės tabletę; laimės stebuklingą tabletę
They both fall under the guideline of we should be seeking out not	Jie abu patenka į gairę , kad neturėtume ieškoti	Jie abu priskirtini prie sveikų pasirinkimų; Jie abu yra iš srities; Jais abiem siūloma ne siekti; Jie abu patenka į nuostatą; Jie abu kalba apie tai; Jie abu patenka į kategoriją; Jie abu patenka į grupę dalykų
And so, this is what Babyak and colleagues looked at .	Taigi Babyak ir jo kolegos pažvelgė į tai .	atkreipė dėmesį; tyrinėjo; nagrinėjo

Atskiras daugiareikšmiškumo atvejis – terminų vartojimas. Šioje ištraukoje profesorė keturis kartus paminėjo terminą *condition*, kalbėdama apie eksperimento tiriamųjų grupes. MV sistema vieną kartą pateikė vertimą *sąlygų* ir tris kartus – *būklė*. Ši klaida galėtų būti skaičiuojama kaip viena, nes tai vienas ir tas pats terminas, tačiau pasirinkta skaičiuoti visus 4 atvejus atskirai, nes MV tą patį terminą vertė skirtingai, o vertėjai skirtingai jį taisė ir nesilaikė nuoseklumo reikalavimų. Pažymėtina, kad tik 2 vertėjai spėjo postredaguoti visas keturias eilutes.

Originalo tekstas	MV	Vertėjų taisymai
And so, they assigned people to one of three conditions for six weeks	Taigi jie šešioms savaitėms paskyrė žmonėms vieną iš trijų sąlygų ,	režimų; grupių; nurodymų; metodų
a medicine condition they gave you the antidepressant	vaistinė būklė , kurią jie davė jums antidepressantų	gydymą; metodą; grupę
you get a certain number in the medicine condition maybe just over half are recovered.	Jūs gaunate tam tikrą skaičių vaistų būklėje , gal kiek daugiau nei pusė pasveiko.	grupėje
But if you look at an exercise condition,	Bet jei pažvelgsite į pratimų būklę ,	grupę

Idioma *cheesy* ištraukoje kartojasi du kartus. MV pateikia du skirtingus vertimus, tačiau abu nėra tinkami, neatsižvelgiama į kontekstą. Reikėtų paminėti, kad visi – tiek profesionalūs, tiek neprofesionalūs – vertėjai siūlė savo taisymsus, tačiau ne visi taisymai būtų laikomi teisingais.

Originalo tekstas	MV	Vertėjų taisymai
all these cheesy things that you see on infomercial that will make you happier.	visi šie sūrūs dalykai, kuriuos matote informacinėje reklamoje, padarys jus laimingesnius.	banalūs; nereikšmingi; pigius
Sounds really cheesy . You think you want a huge salary,	Skamba tikrai šauriai . Manote, kad norite didžiulio atlyginimo,	banaliai; neįtikinamai; kvailai

Sisteminės klaidos

MV tekste pasitaikė tik viena sisteminė klaida. Visi vertėjai pastebėjo šį netikslų vertimą ir pateikė pataisytus variantus. Tai, kad sisteminių klaidų beveik nėra, rodo, jog MV gana progresavęs ir tikėtina, kad greitai metu šios klaidų kategorijos bus galima atsisakyti. Kita vertus, tokias akivaizdžias klaidas lengviau pastebėti ir jas pataisyti, tad mažėjant šios kategorijos klaidų vertėjai turėtų dar atidžiau tikrinti originalo tekstą su vertimo tekstu, nes klaidų gali ir nebepastebėti.

Originalo tekstas	MV	Vertėjų taisymai
Well, it turns out that it is two things that are	Pasirodo, kad palyginimai	Pasirodo, kad ši „tablete“ yra du mums visuomet prieinami dalykai; Pasirodo, kad tai yra du dalykai, kurie yra; Pasirodo, kad tai yra du dalykai; Pasirodo, kad tai yra du dalykai, kurie; Pasirodo, kad yra du visada jums prieinami dalykai; Pasirodo, tai yra du dalykai

Laisvieji taisymai

Dalyviai atliko ir laisvųjų taisymsų, kurie nebuvo būtini, nes MV variantas atitiko originalo teksto prasmę.

Originalo tekstas	MV	Vertėjų taisymai
The final one is two sets of things I think you should be doing.	manau	man regis
It will also have no side effects and it's legal .	teisėta	legali; tikra

Lietuvių kalboje užsienietiškos pavardės gali būti gramatinamos arba paliekamos originalios rašybos. Vienas profesionalus vertėjas pridėjo prie tyrėjo pavardės lietuvišką galūnę. Tikėtina, jog kiti dalyviai vadovavosi taisykle, kad mokslinėje literatūroje autorių pavardes galima palikti nepakeistas, nes tai – akademinio pobūdžio informacija.

Originalo tekstas	MV	Vertėjų pataisymai
And so, this is what Babyak and colleagues looked at.	Taigi Babyak ir jo kolegų pažvelgė į tai.	Babyak'as

Kitas įdomus laisvųjų pasirinkimų atvejis – termino *infomercial* vertimas. Du profesionalūs vertėjai pasirinko ištaisyti MV siūlytą terminą *informacinė reklama* į tiesiog *reklama*, tačiau to daryti nereikėjo, nes terminas *informacinė reklama* vartojamas srities specialistų ir apibrėžtas Lietuvos komunikacijos agentūrų asociacijos KOMAA⁵.

Originalo tekstas	MV	Vertėjų pataisymai
all these cheesy things that you see on infomercial that will make you happier.	visi šie sūrūs dalykai, kuriuos matote informacinėje reklamoje , padarys jus laimingesnius.	reklamoje

Anglų kalboje *Present Continuous* laikas turi keletą reikšmių: veiksmas šiuo metu, tendencijos ir veiksmas netolimoje ateityje. Profesorė kalba apie eksperimento dalyvių savijautą praėjus 16 savaitių, todėl galima pasirinkti ateities reikšmę, nors esamasis laikas taip pat puikiai perteikia kalbėtojos mintį. Šią vietą taisė 2 iš 3 neprofesionalių vertėjų, postredagavusių šį segmentą.

Originalo tekstas	MV	Vertėjų pataisymai
... are you still suffering from depression?	...ar vis dar sergate depresija?	sirgsite

⁵ <https://zodynas.vz.lt/informacine-reklama>

Apklaustos duomenų analizė

Atlikę postredagavimo užduotį, dalyviai atsakė į klausimyno klausimus (1 priedas). Jų atsakymai pateikiami dviejose grupėse: profesionalių ir neprofesionalių vertėjų. Vėliau atsakymai palyginti ir išskirti panašumai bei skirtumai.

Įvertinti sistemą paprašius profesionalių vertėjų, trys respondentai pasirinko atsakymą „patenkinamai“, vienas – „blogai“. Įvardydami sunkumus, su kuriais susidūrė postredaguodami MV, dalyviai nurodė sintaksę, sakinio struktūrą. Profesionalūs vertėjai teigė, kad vertimą apsunkino subtitravimas, nes reikėjo versti tekstą, kuris pateiktas konkrečioje eilutėje, todėl sunku keisti sakinio struktūrą, o vertimas ne tik turi atitikti originalą, bet ir tilpti į nustatyto ženklų skaičiaus eilutę. Taip pat buvo sudėtinga aptikti klaidas, nes neaišku, kurioje vietoje sistema klaidingai išvertė, ir tuomet spręsti, ar tą klaidą taisyti ir kaip taisyti, o gal geriau versti originalo tekstą iš naujo.

Kalbėdami apie sistemos privalumus, respondentai paminėjo, kad pateikiamas vertimo juodraštis, rėmai, nuo kurio galima atsispirti ir nereikia daug galvoti. Taip pat nereikia visko rašyti, nes dalis sakinio yra tinkama. MV sistemos pateiktas vertimas yra suprantamas.

Atsakydami į klausimą apie sistemos trūkumus, profesionalūs vertėjai paminėjo, kad MV yra pažodinis, vargina didelės apimties nereikalingi elementai, tai, kad tas pats žodis verčiamas skirtingai, teksto segmentavimas, sklandumo ir vientisumo trūkumas verčiant tą patį sakinį, išdėstytą per kelis segmentus. Įterpiniai, pavyzdžiui, *like*, verčiami kaip prasminiai žodžiai, trūksta tinkamų žodžių prasmių parinkimo.

Paklausti, ar naudotą sistemą versdami platformos „Coursera“ turinį, du vertėjai atsakė neigiamai, vienas abejojo, o trys pateikė teigiamus atsakymus. Vertėjai, kurie ne-naudotą MV sistemos, savo atsakymą grindė tuo, kad tai yra akademinės srities tekstas, kuriam būtinas tikslumas.

Du neprofesionalūs vertėjai sistemą įvertino patenkinamai, vienas – blogai, vienas – gerai. Sunkiausia dalyviams buvo nustatyti stilių, taisyti pažodinį vertimą, atkurti prasmę, surasti sakinio centrą ir taisyti nelietuviškus junginius.

Respondentai įžvelgė sistemos privalumą, kad tekstas jau surinktas ir reikia tik atlikti taisymus. Kaip sistemos trūkumai įvardyti pažodinis vertimas ir netikslūs žodžių seka.

Paklausti, ar naudotą sistemą versdami tokio pobūdžio mokomuosius vaizdo įrašus, trys iš keturių dalyvių atsakė teigiamai, nes pateikiamas vertimo pagrindas. Vienas dalyvis atsakė, kad gerai pagalvotų, ar naudotųsi.

Lyginant profesionalių ir neprofesionalių vertėjų sistemos įvertinimus, galima išskirti šiuos panašumus:

1. Dauguma dalyvių sistemą įvertino patenkinamai: 7 iš 10.
2. Didžiausias kliuvinys postredaguojant – sintaksė.

3. Didžiausias sistemos trūkumas – ji nepritaikyta lietuvių kalbai.
4. Didžiausias sistemos privalumas – didelė dalis teksto yra išversta ir jį reikia tik redaguoti.
5. 6 respondentai sakė, kad naudotų „Smartling“ pateiktą MV, jei jiems reikėtų versti daugiau tokių vaizdo įrašų į lietuvių kalbą; 2 – nesinaudotų, 2 nežino, ar naudotųsi.
6. Profesionalūs vertėjai į MV sistemą žvelgė įtariau ir ją įvertino blogiau nei neprofesionalūs vertėjai.

Išvados

Atlikus postredagavimo eksperimentą su 6 profesionaliais ir 4 neprofesionaliais vertėjais, nustatytos MV klaidos verčiant platformos „Coursera“ turinį iš anglų kalbos į lietuvių kalbą vertimo sistema „Smartling“. Dažniausiai išryškėjo tokie trūkumai kaip pažodinis vertimas, netinkamas daugiareikšmių žodžių vertimas ir netaisyklingas įvardžių vartojimas. Tarp lingvistinių klaidų gramatinių rasta nedaug: dvigubai dažniau pasitaikė leksinių klaidų, nustatyta tik viena sisteminė klaida. Sisteminių klaidų nebuvimas rodo, kad MV sistemos jau gana progresavusios ir šios kategorijos klaidų beveik nebedaro.

Vienas iš svarbiausių MV sistemos „Smartling“ postredagavimo ypatumų yra teksto skaidymas į subtiltrų eilutes. Viena vertus, dėl šio skaidymo MV sistema daro gramatinių klaidų, nes nesieja verčiamo teksto su tekstu, kuris yra prieš tai esančioje eilutėje, o dėl to sakinytis išskaidomas į atskirus segmentus, kurie nėra tarpusavyje susiję. Kita vertus, kaip pažymi profesionalūs vertėjai, tai sunkina vertėjų darbą, nes sunku keisti sakinio struktūrą, spręsti, kur prasideda ir baigiasi konkretus sakinio segmentas, ir užtikrinti, kad išverstas tekstas tilptų į nustatyto ženklų skaičiaus eilutę. Tikėtina, kad šie trūkumai būdingi ne tik anglų ir lietuvių kalbų porai.

Vertėjų nustatytos lingvistinės klaidos rodo, kad vertimo sistema „Smartling“ verčia dar nepakankamai gerai, ypač į lietuvių kalbą. Tai rodo daug pažodiškumo atvejų, netaisyklingai vartojami įvardžiai, netinkamai parenkamos žodžių reikšmės. Šie trūkumai galėtų būti pašalinti, jei MV sistema būtų išmokyta atsižvelgti į platesnį kontekstą, pavyzdžiui, versti visą sakinį, o ne atskirus segmentus. Tyrimo rezultatai rodo, kad tokiu būdu didžiosios dalies gramatinių klaidų būtų galima išvengti.

Profesionalūs vertėjai postredaguodami dėjo daugiau kognityvinių pastangų ir išžvelgė daugiau sistemos trūkumų nei privalumų. O neprofesionalių vertėjų nuomonė apie sistemą palankesnė: jie entuziastingiau vertino privalumus ir galimybę pasinaudoti MV. Šis skirtingas požiūris svarbus, nes vertimo sistema „Smartling“ naudojami savanoriai, dažniausiai – neprofesionalūs vertėjai.

Vienintelis sistemos privalumas, kurį nurodė tiek profesionalūs, tiek neprofesionalūs vertėjai, yra tas, kad išverstu tekstu galima naudotis kaip juodraščiui: tekstas jau parašytas, tad jo nereikia spausdinti. Nepaisant išsakytos kritikos, tyrimo dalyvių apklausa rodo, kad šis privalumas gana didelis, nes tik 2 iš 10 tyrimo dalyvių versdami platformos „Coursera“ turinį iš anglų kalbos į lietuvių kalbą MV sistema „Smartling“ nesinaudotų.

Literatūra

- Almeida, Giselle de. 2013. *Translating the Post-Editor: An Investigation of Post-Editing Changes and Correlations with Professional Experience across Two Romance Languages*. Unpublished thesis. Dublin City University.
- Arfaoui, Fathi. 2022. Percentage of Internet Content for Users by Language. *Trustiko*. [žiūrėta 2022-03-16]: <https://trustiko.com/percentage-of-internet-content-for-users-by-language/>.
- Callison-Burch, Chris, Cameron Fordyce, Philipp Koehn, Christof Monz and Josh Schroeder. 2007. (Meta-) Evaluation of Machine Translation. *Proceedings of the Second Workshop on Statistical Machine Translation*, 136–58. Prague, Czech Republic: Association for Computational Linguistics. [žiūrėta 2022-01-30]: <https://aclanthology.org/W07-0718>.
- Carl, Michael, Barbara Dragsted, Jakob Elming, Daniel Hardt and Arnt Lykke Jakobsen. 2011. The Process of Post-Editing: A Pilot Study. *Copenhagen Studies in Language* 41. 131–42.
- Chan, Yee Seng and Hwee Tou Ng. 2008. MAXSIM: A Maximum Similarity Metric for Machine Translation Evaluation. *Proceedings of ACL-08: HLT*, 55–62. Columbus, Ohio: Association for Computational Linguistics. [žiūrėta 2022-02-14]: <https://aclanthology.org/P08-1007>.
- Daems, Joke, Sonia Vandepitte, Robert J. Hartsuiker and Lieve Macken. 2017. Identifying the Machine Translation Error Types with the Greatest Impact on Post-Editing Effort. *Frontiers in Psychology* 8. [žiūrėta 2021-12-17]: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01282>.
- Doddington, George. 2002. Automatic Evaluation of Machine Translation Quality Using N-Gram CoOccurrence Statistics. *Proceedings of the Second International Conference on Human Language Technology Research -*, 138. San Diego, California: Association for Computational Linguistics. [žiūrėta 2021-12-17]: <https://doi.org/10.3115/1289189.1289273>.
- Farrús, Mireia, Marta R. Costa-jussà, José B. Mariño and José A. R. Fonollosa. 2010. Linguistic-Based Evaluation Criteria to Identify Statistical Machine Translation Errors. *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*. Saint Raphaël, France: European Association for Machine Translation. [žiūrėta 2022-02-14]: <https://aclanthology.org/2010.eamt-1.12>.
- Federico, Marcello, Matteo Negri, Luisa Bentivogli and Marco Turchi. 2014. Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-Effects Models. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1643–53. Doha, Qatar: Association for Computational Linguistics. [žiūrėta 2022-03-08]: <https://doi.org/10.3115/v1/D14-1172>.
- Font Llitjós, Ariadna, Jaime G. Carbonell and Alon Lavie. 2005. A Framework for Interactive and Automatic Refinement of Transfer-Based Machine Translation. *Proceedings of the 10th EAMT Conference: Practical Applications of Machine Translation*. Budapest, Hungary: European Association for Machine Translation. [žiūrėta 2022-01-16]: <https://aclanthology.org/2005.eamt-1.13>.
- Giménez, Jesús, and Lluís Màrquez. 2008. Heterogeneous Automatic MT Evaluation Through NonParametric Metric Combinations. *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*. [žiūrėta 2022-01-15]: <https://aclanthology.org/I08-1042>.

- Guzmán, Francisco, Ahmed Abdelali, Irina Temnikova, Hassan Sajjad and Stephan Vogel. 2015. How Do Humans Evaluate Machine Translation. *Proceedings of the Tenth Workshop on Statistical Machine Translation*, 457–66. Lisbon, Portugal: Association for Computational Linguistics. [žiūrėta 2021-12-18]: <https://doi.org/10.18653/v1/W15-3059>.
- Hu, Ke, Sharon O'Brien, and Dorothy Kenny. 2020. A Reception Study of Machine Translated Subtitles for MOOCs. *Perspectives* 28 (4). 521–38. [žiūrėta 2022-02-06]: <https://doi.org/10.1080/0907676X.2019.1595069>.
- Jakobsen, Arnt Lykke. 2011. Tracking Translators' Keystrokes and Eye Movements with Translog. *Methods and Strategies of Process Research*, July. 37–55.
- Jiménez-Crespo, Miguel A. 2017. *Crowdsourcing and Online Collaborative Translations: Expanding the Limits of Translation Studies*. John Benjamins. [žiūrėta 2021-11-30]: <https://doi.org/10.1075/btl.131>.
- Kelly, Nataly, Rebecca Ray and Donald A. DePalma. 2021. From Crawling to Sprinting: Community Translation Goes Mainstream. *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 10 (October). [žiūrėta 2022-03-04]: <https://doi.org/10.52034/lanstts.v10i.278>.
- Kirchhoff, Katrin, Daniel Capurro, and Anne Turner. 2012. Evaluating User Preferences in Machine Translation Using Conjoint Analysis. *Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, 119–26. Trento, Italy: European Association for Machine Translation. [žiūrėta 2021-12-18]: <https://aclanthology.org/2012.eamt-1.35>.
- Koehn, Philipp. 2009. *Statistical Machine Translation*. Cambridge: Cambridge University Press. [žiūrėta 2021-11-04]: <https://doi.org/10.1017/CBO9780511815829>.
- Koponen, Maarit, Wilker Aziz, Luciana Ramos and Lucia Specia. 2012. Post-Editing Time as a Measure of Cognitive Effort. *Workshop on Post-Editing Technology and Practice*. San Diego, California, USA: Association for Machine Translation in the Americas. [žiūrėta 2021-10-31]: <https://aclanthology.org/2012.amtawtp.2>.
- Koponen, Maarit, Leena Salmi and Markku Nikulin. 2019. A Product and Process Analysis of PostEditor Corrections on Neural, Statistical and Rule-Based Machine Translation Output. *Machine Translation* 33 (1). 61–90. [žiūrėta 2022-01-14]: <https://doi.org/10.1007/s10590-019-09228-7>.
- Krings, Hans P. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. Edited by Geoffrey S. Koby. 1st edition. Kent, Ohio: The Kent State University Press.
- Lavie, Alon and Abhaya Agarwal. 2007. METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments. *Proceedings of the Second Workshop on Statistical Machine Translation*, 228–31. Prague, Czech Republic: Association for Computational Linguistics. [žiūrėta 2022-01-14]: <https://aclanthology.org/W07-0734>.
- Lommel, Arle, Aljoscha Burchardt, Maja Popovic, Kim Harris, Eleftherios Avramidis and Hans Uszkoreit. 2014. Using a New Analytic Measure for the Annotation and Analysis of MT Errors on Real Data. *Proceedings of the 17th Annual Conference of the European Association for Machine Translation*, 165–72. Dubrovnik, Croatia: European Association for Machine Translation.
- McDonough Dolmaya, Julie. 2012. Analyzing the Crowdsourcing Model and Its Impact on Public Perceptions of Translation. *The Translator* 18 (2): 167–91. [žiūrėta 2022-02-11]: <https://doi.org/10.1080/13556509.2012.10799507>.
- Moorkens, Joss and O'Brien. 2017. Assessing User Interface Needs of Post-Editors of Machine Translation. *Human Issues in Translation Technology*. Edited by: Dorothy Kenny. London: Routledge, 109130.
- Nitzke, Jean and Anne-Kathrin Gros. 2020. Preferential Changes in Revision and Post-Editing. *Translation Revision and Post-Editing*. Edited By Maarit Koponen, Brian Mossop, Isabelle S. Robert, Giovanna Scocchera. London: Routledge. 21–34.

- O'Brien, Sharon. 2011. Towards predicting post-editing productivity. *Machine Translation* 25(3). 197–215.
- Padó, Sebastian, Daniel Cer, Michel Galley, Dan Jurafsky and Christopher D. Manning. 2009. Measuring Machine Translation Quality as Semantic Equivalence: A Metric Based on Entailment Features. *Machine Translation* 23 (2), 181–93. [žiūrėta 2022-01-25]: <https://doi.org/10.1007/s10590-009-9060-y>.
- Papineni, Kishore, Salim Roukos, Todd Ward and Wei-Jing Zhu. 2002. Bleu: A Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–18. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics. [žiūrėta 2022-03-13]: <https://doi.org/10.3115/1073083.1073135>.
- Petkevičiūtė, Inga and Bronius Tamulynas. 2011. Kompiuterinis vertimas į lietuvių kalbą: alternatyvos ir jų lingvistinis vertinimas. *Kalbų studijos* 18. 38–45.
- Popović, Maja. 2015. ChrF: Character n-Gram F-Score for Automatic MT Evaluation. *Proceedings of the Tenth Workshop on Statistical Machine Translation*, 392–95. Lisbon, Portugal: Association for Computational Linguistics. [žiūrėta 2022-02-02]: <https://doi.org/10.18653/v1/W15-3049>.
- Popović, Maja. 2018. Error Classification and Analysis for Machine Translation Quality Assessment. *Translation Quality Assessment: From Principles to Practice*, edited by Joss Moorkens, Sheila Castilho, Federico Gaspari and Stephen Doherty, 129–158. Machine Translation: Technologies and Applications. Cham: Springer International Publishing. [žiūrėta 2021-12-06]: https://doi.org/10.1007/978-3319-91241-7_7.
- Popović, Maja and Hermann Ney. 2009. Syntax-Oriented Evaluation Measures for Machine Translation Output. *Proceedings of the Fourth Workshop on Statistical Machine Translation*, 29–32. Athens, Greece: Association for Computational Linguistics. [žiūrėta 2022-04-12]: <https://aclanthology.org/W09-0402>.
- Snover, Matthew, Bonnie Dorr, Rich Schwartz, Linnea Micciulla and John Makhoul. 2006. A Study of Translation Edit Rate with Targeted Human Annotation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, 223–31. Cambridge, Massachusetts, USA: Association for Machine Translation in the Americas. [žiūrėta 2022-04-01]: <https://aclanthology.org/2006.amta-papers.25>.
- Temnikova, Irina. 2010. Cognitive Evaluation Approach for a Controlled Language Post-Editing Experiment. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA). [žiūrėta 2022-01-12]: http://www.lrec-conf.org/proceedings/lrec2010/pdf/437_Paper.pdf.
- Turian, Joseph P., Luke Shen and I. Dan Melamed. 2003. Evaluation of Machine Translation and Its Evaluation. *Proceedings of Machine Translation Summit IX: Papers*. New Orleans, USA. [žiūrėta 2022-04-12]: <https://aclanthology.org/2003.mtsummit-papers.51>.
- Vandenbosch, Betty. 2020. More than 1.6 Million Learners around the World Benefit from Partner Contributions in Coursera's Response to the Pandemic. Coursera Blog. 31 July 2020. [žiūrėta 2021-06-12]: <https://blog.coursera.org/more-than-1-6-million-learners-around-the-world-benefit-from-partnercontributions-in-courseras-response-to-the-pandemic/>.
- Vilar, David, Jia Xu, Luis Fernando D'Haro and Hermann Ney. 2006. Error Analysis of Statistical Machine Translation Output. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy: European Language Resources Association (ELRA). [žiūrėta 2022-01-15]: http://www.lrec-conf.org/proceedings/lrec2006/pdf/413_pdf.pdf.

PRIEDAI

1 priedas. Postredagavimo užduotis

	Coursera - 445 žodžiai	Mašininis vertimas
1	The final one is two sets of things I think you should be doing.	Paskutinis yra du dalykai, kuriuos, manau, turėtumėte padaryti.
2	They both fall under the guideline of we should be seeking out not	Jie abu patenka į gairę, kad neturėtume ieškoti
3	good grades and not big salary but we should be seeking out healthier practices.	geri pažymiai ir ne didelis atlyginimas, bet turėtume ieškoti sveikesnės praktikos.
4	And this is the point where I get to be a terrible salesman in the course,	Ir tai yra taškas, kai aš galiu būti baisus pardavėjas kursuose,
5	where I could pretend that I have a happy pill,	kur galėčiau apsimesti, kad turiu laimingą tabletę,
6	like the panacea pill,	kaip panacėjos tabletė,
7	to just make you happier over time and have	kad tik laikui bėgant taptumėte laimingesni ir turėtumėte
8	all these cheesy things that you see on infomercial that will make you happier.	visi šie sūrūs dalykai, kuriuos matote informacinėje reklamoje, padarys jus laimingesnius.
9	It will give you better grades even though you're not supposed to care about that,	Tai duos jums geresnius pažymius, net jei jums tai neturėtų rūpėti,
10	but it will make you look better	bet tai leis jums atrodyti geriau
11	even though you're not supposed to care about your looks	nors tu neturėtum rūpintis savo išvaizda
12	but it will do that before you do.	bet tai padarys anksčiau nei jūs.
13	It will also have no side effects and it's legal	Tai taip pat neturės jokio šalutinio poveikio ir yra teisėta.
14	It's not like some sketchy drug and it's even free.	Tai ne kaip koks nors eskizinis narkotikas ir netgi nemokamas.
15	Like, what could this horrible, wonderful thing be?	Pavyzdžiui, kas gali būti šis baisus, nuostabus dalykas?
16	Well, it turns out that it is two things that are	Pasirodo, kad palyginimai

	Coursera - 445 žodžiai	Mašininis vertimas
17	completely available to you all the time but we forget the importance of.	visą laiką jums prieinama, bet pamirštame apie tai svarbą.
18	The first one of these is just simply exercise.	Pirmasis iš jų yra tiesiog mankšta.
19	Sounds really cheesy. You think you want a huge salary,	Skamba tikrai šauniai. Manote, kad norite didžiulio atlyginimo,
20	but really simple increases in your physical activity all	bet tikrai paprastas jūsų fizinio aktyvumo padidėjimas
21	the time can have huge, huge psychological benefits.	laikas gali turėti didžiulės, didžiulės psichologinės naudos.
22	Even huger benefits than the other mechanisms we have for boosting happiness chemically.	Dar didesnė nauda nei kiti mechanizmai, kuriuos turime, kad chemiškai padidintume laimę.
23	And so, this is what Babyak and colleagues looked at.	Taigi Babyak ir jo kolegos pažvelgė į tai.
24	They tried to see whether or not they could use an exercise intervention to	Jie bandė išsiaiškinti, ar galėtų panaudoti pratimų intervenciją
25	really increase happiness in a population	tikrai padidina gyventojų laimę
26	that was really suffering from happiness deficits,	kuris tikrai kentėjo nuo laimės trūkumo,
27	namely people with major depression.	tai yra žmonės, sergantys pasikartojančia depresija.
28	And so, they assigned people to one of three conditions for six weeks,	Taigi jie šešioms savaitėms paskyrė žmonėms vieną iš trijų sąlygų,
29	either exercise which meant you just three times a week did	bet kurį pratimą darėte tik tris kartus per savaitę
30	30 minutes of your own choice of exercise or	30 minučių jūsų pasirinktas pratimas arba
31	a medicine condition they gave you the antidepressant	vaistinė būklė, kurią jie davė jums antidepressantų
32	Zoloft or they did both or they kind of combined both of these together.	Zoloft arba jie padarė abu, arba jie sujungė abu šiuos dalykus.
33	And the question is, when you come back after 16 weeks,	Ir kyla klausimas, kai grįšite po 16 savaičių,
34	can I look ten weeks after those 16 weeks of	Ar galiu atrodyti dešimt savaičių po tų 16 savaičių
35	your relapse rate and are you still suffering from depression?	Jūsų atkryčių dažnis ir ar vis dar sergate depresija?
36	Have you relapsed or are you happy?	Ar atsigavote ar esate laimingas?

	Coursera - 445 žodžiai	Mašininis vertimas
37	And here's what they find. It's pretty incredible data.	Ir štai ką jie randa. Tai gana neįtikėtini duomenys.
38	So, if you look at the percentage of people who are fully recovered,	Taigi, jei pažvelgtumėte į visiškai pasveikusių žmonių procentą,
39	you get a certain number in the medicine condition maybe just over half are recovered.	Jūs gaunate tam tikrą skaičių vaistų būklėje, gal kiek daugiau nei pusė pasveiko.
40	But if you look at an exercise condition,	Bet jei pažvelgsite į pratimų būklę,
41	about 90 percent of the people 10 weeks on are actually recovered from this.	apie 90 procentų žmonių po 10 savaičių iš tikrųjų pasveiksta nuo to
42	Oddly enough, it's actually even better than	Kaip bebūtų keista, tai iš tikrųjų net geriau nei
43	the condition where I'm giving you Zoloft and exercise.	būsena, kai aš tau duodu Zoloftą ir mankštinausi.

2 priedas. Klausimynas

- Kaip vertinate „Smartling“ platformos mašininį vertimą:
 - gerai
 - blogai
 - patenkinamai
- Kas buvo sunkiausia postredaguojant?
- Kokius pastebėjote „Smartling“ mašininio vertimo sistemos privalumus?
- Kokius pastebėjote „Smartling“ mašininio vertimo sistemos trūkumus?
- Ar naudotumėtės „Smartling“ mašininio vertimu verčiant šį vaizdo įrašą?