Maumevičienė, Dainora. 2023. Integration of Machine Translation Tools in Software Localization: mission (im)possible? In: L10N Journal 2(2), pp. 62–79.

Integration of Machine Translation Tools in Software Localization: mission (im)possible?

Dainora Maumevičienė

Kaunas University of Technology

dainora.maumeviciene@ktu.lt

Abstract

The use of machine translation tools in various translation related activities has exponentially increased, not only because of the pandemic and post-pandemic context, the growth of the language industry, but also because of the current technological development of machine translation tools. Researchers, translators and developers actively discuss machine translation capabilities, its impact on translation and language, and challenges it brings. However, when it comes to software localization, possibilities of machine translation tools seem overestimated and limited, not only because of the quality of the output the machine translation tools produce, but also because of technical capacities to recognize programming language and handle difficult scenarios. This article aims to introduce possibilities and limitations of integrating machine translation tools in the process of software localization into the Lithuanian language. Here the term "integration" is used as the application of machine translation tools in the workflow of localization so as to speed up the process and help translators, but not as the technological integration when machine translation tools are integrated with computer aided translation (CAT) tools. The article presents an experiment during which several machine translation tools, such as *Google Translate*, *DeepL*, *Vilnius University machine translation tool* and *Tildė machine translation tool*, were tested with the Lithuanian language as the low-resourced language. The four machine translation tools were selected due to their popularity and current development in and for the Lithuanian language (Utka et al. 2020). The machine translation tools were given to translate .rc2 or .txt software-related resource files. The output quality produced in the Lithuanian language was compared in terms of text cohesion, term accuracy, identification of segments to be localized and damaged programming code. Moreover, the machine translation outputs were compared with the output of professional translation and localization CAT tools such as Passolo and Trados. The results showed that none of the machine translation tools used, despite the current integration of artificial intelligence solution, can produce high-quality translated text in the Lithuanian language due to the assumption that the Lithuanian language (with around 3 million speakers) is not commercially attractive. The output produced cannot be applied to speed up or ease localization in terms of the output text quality.

Keywords: software localization, machine translation tools, integration, output, quality

Maumevičienė, Dainora. 2023. Integration of Machine Translation Tools in Software Localization: mission (im)possible? In: L10N Journal 2(2), pp. 62–79.

Introduction

Machine translation tools, their current development in terms of implementing different artificial intelligence (AI) solutions, are much discussed in various research papers (Skadiņš et al. 2014; Dew et al. 2018; Pituxcoosuvarn, Ishida, 2018; Rossi, Chevrot, 2019; Hoi, 2020; Rossetti, O'Brien, Cadwell, 2020; Kasperė et al. 2021; Sadadany et al. 2021) that outline possibilities and challenges of applying machine translation tools to ease multilingual and cross-cultural communication, and help users understand specific linguistic content and context in the language that the user is not proficient or does not speak at all. Other researchers (Gaspari et al. 2015; Doherty, 2016; Das et al. 2019; Kenny et al. 2020) examine the development of machine translation tools that have also changed the profile of translation as a profession and have also modified the activities translators-professionals undertake. Instead of actual translation from scratch, post-editing skills of the output, produced by machine translation tools, are identified as one of the core competencies a translator must have, in addition to other technology-related competences. This, in turn, has also changed the competences of translators, as indicated in the translator's competence framework 2022, as introduced by the European Master's in Translation Board at the European Commission (*European Master's in Translation Competence Framework*, 2022). Due to its simple use and easily gained accessibility to different types of information by means of translation, machine translation tools are integrated in various settings such as healthcare, enterprises, the European Commission. Yet, the use of machine translation tools poses some risks in terms of its output quality and impact on human perception. The approach to investigate the impact of machine translation use might be different due to the aim of research. Some publications and research papers (Rimkutė, Kovalevskaitė, 2007; Gaspari et al. 2015; Doherty, 2016; Das et al. 2019; Kenny et al. 2020; Rossetti et al. 2020; Utka et al. 2020; Kasperė et al. 2021) aim to examine the impact of machine translation tools on translator training, the profession in general, and translation didactics. Whereas other research papers (Yamashita, Ishida, 2006; Cvilikaitė, 2008; Yasouka, Bjorn, 2011; Castilho, 2016; Doherty, 2016; Canfora, Ottman, 2020; Kasperavičienė et al. 2020) focus on the quality of machine translated output, usability, mistakes, and machine translation impact on user perception. If translation-didactics related approach emphasises functional aspects of machine translation output (the quality of the output, post-editing efforts) as a product by focusing different types of text, the research on how machine translation tools might be used in software localization is limited. Though there are many posts on social media networks such as *LinkedIn* or other on how machine translation might help in software localization by speeding up the process, saving localization costs, and aiding translators-localizers, the research that would focus on productivity, usability and efficient use of machine translation in localization is scarce, though several research papers have been observed by *Tildė* company researchers such as Skadiņš et al. 2014, Skadiņa, 2019 and some

Lithuanian researchers Utka et al. 2016; Utka et al. 2020. Taking it into account, this article aims to examine the impact of machine translation and its use in software localization to localize resources and help files into the Lithuanian language as a low-resourced language. This article presents an experiment carried out at Kaunas University of Technology to examine if machine translation tools can be used for more complicated tasks of software localization, i.e., localization of resource and help files. During the research and experiment, an assumption was made that due to constant development and technological advancement, machine translation tools can already provide output texts of reasonable and acceptable quality, since deep learning is embedded as a solution. Moreover, cases when translators, especially novice translators use machine translation for translatable segments and copy them to professional software for translation or localization with an aim to translate texts faster have been also observed, thus, it is interesting to find out if that aids in the process of localization.

Literature review

The research on machine translation tools, their application challenges, advantages, disadvantages and various possibilities of their use has been increasing due to the constant development of machine translation tools, their popularity in terms of user-friendly usability, possibilities to understand information in a language that a user does not speak as well as use machine translation in work related settings, i.e., tourism, medicine, shopping, etc. In addition, many research papers (Kenny, Moorkens, do Carmo, 2020) tackle specific aspects of machine translation use in relation to translation, translator training, ethical and sustainable machine translation, technological solutions, and other factors connected to socio-economic, educational, organizational and technological aspects of machine translation use (Hoi, 2020). The analysis of scientific literature on machine translation-related research papers has demonstrated that many articles and research works focus on the use of machine translation tools in various multilingual and cross-cultural communication related settings, where machine translation is used to aid personal and professional communication and collaboration (Yamashita, Ishida, 2006; Yasouka, Bjorn, 2011; Dew et al. 2018; Pituxcoosuvarn, 2018; Rosi, Chevrot, 2019; Saanady et al. 2021). Another type of research and publications on machine translation focuses on translator training practices, new translation profession related perspectives interlinked with challenges and possibilities of machine translation application in translation activities and practices, or examine machine translation output and its quality in terms of human versus machine translation mistakes and their evaluation (Gaspari et al. 2015; Doherty, 2016; Castilho, 2016; Kenny, Morkens, do Carmo, 2020; Rossetti et al. 2020). Risks and dangers that end-users might face when exhausting the possibilities of machine translation tools also receive the attention of researchers due to the misunderstandings

and pitfalls of machine translated output in various settings (Das et al. 2019; Canfora and Ottmann, 2020).

Machine translation tools and their usability in software localization are also analyzed, but the research is rather limited and even more limited in the investigation of machine translation and its integration in software localization processes with low-resourced languages, which includes the Lithuanian language (Skadiņš et al. 2014; Utka et al. 2016). It must be mentioned that not much is discussed about how machine translation might be applied to localize help and resource files to complete a full scale software localization. One of the first articles to examine the application of machine translation tools in software localization for low-resourced languages was published by a group of researchers at *Tildė* company (Skadiņš et al. 2014) and aimed at examining the usability and productivity of machine translation tools in software localization. The research results indicated that segments with formatting tags were rendered incorrectly, and the output produced contained many mistakes in improper terminology and phrase use. The researchers admitted the pitfalls of machine translation integration in difficult scenarios, but the use of the tools allowed for an increase in the productivity of translation and localization. Though the research findings were promising and hopeful, no other research work has been carried out regarding further investigations on how machine translation tools could be productively and efficiently applied in software localization into the Lithuanian language. Thus, it might be stated that the scientific investigations in the area of machine translation usability to localize, i.e., to Lithuanise (the terms "*Lithuanise / Lithuanisation*" are frequently applied in the Lithuanian language and context as a Lithuanian counterpart to replace the term "*localize*" and describe the process of software or webpage localization, but not translation) software texts, are rather limited.

However, it must be mentioned that the research on different aspects of machine translation use and its application to translate various texts from English into the Lithuanian and Lithuanian-English languages, and the examination of the output text quality started several decades ago. The first attempts to examine the potential of machine translation in the Lithuanian language in general, classify the main criteria for quality assessment, and give an overview of the most typical mistakes were provided in the research papers of Labutis (2005), Daudaravičius (2006), Rimkutė and Kovalevskaitė (2007, 2008). Cvilikaitė (2008) focused on the analysis of lexical mistakes, the translation of non-equivalent lexical items, by identifying that machine translation quality and non-equivalent word translation do not depend on the discourse and register of the input text. These papers and research works might be described as pre-neural machine translation research and the starting point. The results of the research demonstrated that ordinary nomination conditions machine translation mistakes, i.e., polysemy, homonymy (Cvilikaitė, 2008). Scientists Petkevičiūtė and Tamulynas (2011) provided research insights and described the main

indicators that could be applied to examine machine translation quality in terms of the English-Lithuanian language pair, and enumerated challenges that translators might face with and machine translation tools that could trigger when translating into the Lithuanian language. Further on, more advanced research in terms of machine translation tool use for the Lithuanian language was carried out by Kasperavičienė et al. (2020). A group of researchers at Kaunas University of Technology carried out an extensive examination of machine translation quality assessment by means of an eye-tracking experiment that was applied to determine types of mistakes that would cause understanding difficulties when reading machine-translated text from English into the Lithuanian language. The research results indicated that erroneous machine-translated texts require additional cognitive effort in comparison to error-free texts. Moreover, the perception of lexical errors requires more time and effort to understand a machine-translated text from English into Lithuanian.

In addition, Kasperė et al. (2021) aimed to examine different attitudes towards the usability and quality of machine translated output from English into the Lithuanian language from the perspective of the end-user. The research focused on how Lithuanian end-users apply machine translation tools and view their produced output. The research findings were grounded by the results of a substantial survey carried out in Lithuania in 2021, and involved 402 participants. Both the survey and the research results indicated that *Google Translate* is the best known and most frequently used machine translation tool by Lithuanian users. It is frequently applied for various purposes, since 96,6 percent of respondents identified it as the main machine translation tool. *DeepL, Tildė, Microsoft* and other were enumerated among other less frequently used machine translation tools (Kasperė et al. 2021). It is important to note that the investigation of the researchers has laid the foundation for further research in terms of machine translation usability, quality inspection from the end-user perspective due to the fact that among the respondents the opinions of translation professionals were included. The research did not focus on the error analysis; yet, the results mentioned above might serve as a starting point to further investigate machine translation application in software localization and this article is one of the first attempts to dwell deeper on machine translation powered software localization and move forward with the research on machine translation integration for the Lithuanian language.

Research methodology

The research methodology applied in this particular paper focuses on the systematic literature review method, so as to provide the grounds for this type of analysis. The aim of the research is to evaluate the possibilities and limitations of machine translation tool integration in the process of software localization, i.e., Lithuanisation. Possibilities of four machine translation tools, such as *Google Translate, DeepL*, *Vilnius University*

machine translation tool, and *Tildė machine translation tool,* were examined following the criteria adapted from Multidimensional Quality Metrics (MQM) proposed as a framework (Lommel et al. 2014). The main aspects that participants of the experiment focused on are related to the following MQM error hierarchy, which has been adapted by adding a graphical layout of the source texts. This is mainly related to the ability of machine translation tools to retain the same graphical structure in both source and target or input and output texts when using .txt, .pdf, .hhc, and similar extensions of various resource and help files. Therefore, possible modifications and changes of the input and output texts were examined by considering the following aspects:

- Accuracy
- Fluency (with a focus on correct grammar and punctuation use)
- Terminology (incorrect and inconsistent use of terms)
- Style
- Graphical layout.

It needs to be noted that in this research and several experiments, the productivity of translators in the process of localization was not examined and addressed, due to the primary aim of the research to focus on the possibility of integrating machine translation tools in more complicated scenarios of software Lithuanisation. The research was carried out using two scenarios.

Scenario 1 was related to the localization of resource files from English into Lithuanian using no specialized software of translation technology tools, and opening the file in *Notepad.* The resource files had to be localized and translated using the program. When the localization of the resource file was completed, the localized file was uploaded to *SLD Passolo Translator 2018* or *Passolo 2022* and aligned with the source file so that to perform the analysis of the usability and quality of the produced output.

Scenario 2. The resource file (.rc2 file extension) was uploaded to the four machine translation tools (*Google Translate, DeepL, Vilnius University machine translation tool, Tildė machine translation tool*) to test the possibilities of using machine translation tools in software localization and compare their output quality. Since some of the machine translation tools (*Google Translate, DeepL, Vilnius University machine translation tool*) do not support .txt, .rc2 and similar help and content-related files, the source text was uploaded as a .pdf extension file. During the experiment, the idea that .pdf extension files might cause incorrect segmentation was taken into consideration as well. Participants of the experiments were instructed that .doc, or .docx file extensions should not be used due to the fact that when a .txt file is opened as a Word document, the programming code might be damaged, segments are truncated, and the file cannot be used in subsequent processes of software localization. When the output texts were obtained, they were qualitatively examined and compared in terms of their possible usability for localization processes, as well as the output quality was inspected. In both scenarios, text-to-text segments were tested out and examined. During the first stages

of carrying out scenario 2 in 2021 it was not initially planned to try out generative tools such as *ChatGPT*; however, due to its popularity and the observations that translators start using such generative tools for translation tasks, it was decided to test what output quality the tool would produce when localizing a file from English into the Lithuanian language. Therefore, the trial was completed in 2023.

The experiment design

For the experiment and testing out if neural machine translation tools can be applied to resource and help file localization into the Lithuanian language, an .rc2 file in English was selected. The file was opened in *Notepad* program to perform the analysis of the file to be localized and provide a possibility for the participants of the experiment (translators-localizers) to localize and test out the use of *Notepad* tool. The example of the resource file is presented in the picture below:

Picture 1. ***The example of the resource file as a source text***

File Edit View

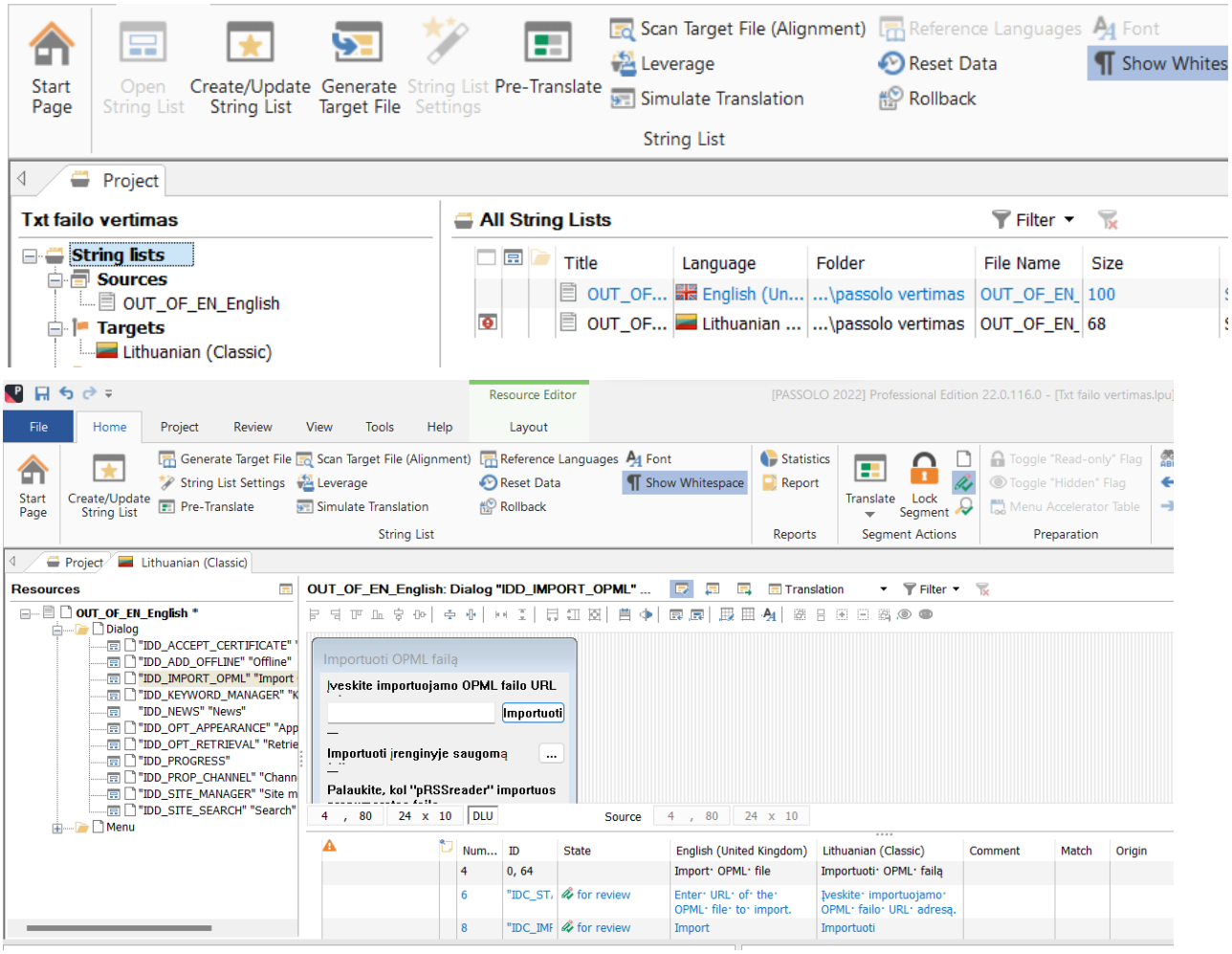
```
//////////////////////////////////////////////////////////////////////////////
// Resources

//////////////////////////////////////////////////////////////////////////////
//
// Dialog
//

IDD_NEWS DIALOG DISCARDABLE  0, 0, 139, 154
STYLE DS_MODALFRAME | WS_POPUP | WS_VISIBLE | WS_CAPTION | WS_SYSMENU
CAPTION "News"
FONT 8, "System"
BEGIN
    LTEXT           "",IDC_SITE_NAME,0,0,138,15,SS_CENTERIMAGE
END

IDD_IMPORT_OPML DIALOG DISCARDABLE  0, 0, 136, 151
STYLE DS_MODALFRAME | WS_POPUP | WS_CAPTION | WS_SYSMENU
CAPTION "Import OPML file"
FONT 8, "System"
BEGIN
    LTEXT           "Enter URL of the OPML file to import.",IDC_STATIC1,4,3,
                    130,10
    EDITTEXT        IDC_URL,4,16,92,12,ES_AUTOHSCROLL
    DEFPUSHBUTTON   "Import",IDC_IMPORT,99,16,35,12
    LTEXT           "---",IDC_HRULE1,4,29,130,8
    LTEXT           "Import file stored on your device",IDC_STATIC2,4,40,111,
                    10
    PUSHBUTTON      "...",IDC_BROWSE,119,38,15,12
    LTEXT           "Please wait a moment while pRSSreader imports the subscription file.",
                    IDC_IMPORT_LBL,4,60,130,18,NOT WS_VISIBLE
    LTEXT           "Status:",IDC_STATUS_LBL,4,80,24,10
```

After the localization of the .rc2 file using *Notepad* was completed, the localized target file was opened and aligned in *SDL Passolo 2018* and *SDL Passolo 2022* software, which is typically used in software localization processes. The localized .rc2 file was opened in *Passolo* to examine if translatable segments are not truncated, all Lithuanian diacritics are correctly marked, segments are not too long and fit the indicated number of characters as tab labels and the programming code is not destroyed. The participants

of the experiment were also asked to try out and localize the provided source file with the help of *Passolo* and compare the difference of using both programs. The rationale for carrying out the experiment that direction is related to the fact that *SDL Passolo 2018* and *SDL Passolo 2022* software is expensive and not so many translators purchase it for localization. Moreover, in cases when localization tasks are outsourced and translators work as freelancers, word-only translation environments are numerous. Therefore, if no contextual information is provided, programming and coding language, available in .txt files and readable in *Notepad* becomes a valuable source of information for translators in further decision-making processes on how segments could be rendered in the target language.

In addition, *SDL Passolo 2018* and *SDL Passolo 2022* software's analysis function was applied to identify the number of strings to be localized into the Lithuanian language. All in all, the localizable file contained 100 strings, where a string could be composed of just one lexical item, such as a word (a term), and the longest one consisting of several sentences, composed of 21 words. It must be noted that *Passolo* recognised both the original .rc2 file and the aligned target file in the Lithuanian language. The picture of .rc2 file as uploaded and aligned in *Passolo* software is presented in the picture below:

Picture 2. ***The target .rc2 resource file in the Lithuanian in Passolo 2022***



With an aim to test out machine translation tools in the process of software Lithuanisation, the following machine translation tools were selected: *Google Translate, DeepL*, *Vilnius University machine translation tool,* and *Tildė machine translation tool.* The machine translation tools were purposefully selected, due to the fact that some of them, i.e., *Google Translate* and *DeepL* were indicated in the research of Kasperė et al. (2021) as the most popular machine translation tools. *Google Translate* was selected because it is considered to be the most popular and best-known machine translation tool that is widely used for different purposes. Another factor that was important in the selection of *Google Translate* is the fact that it is powered by AI solutions and it is a neural machine translation tool. *DeepL* was applied in the research due to the fact that it is easily accessible and quite often used in comparison to other machine translation tools by professional translators.

Vilnius University machine translation tool was selected for the research and analysis due to the fact that the tool was developed as part of the European Union funded project in 2012–2014. Though the machine translation tool was initially designed as a statistical machine translation tool (www.tilde.lt), it incorporates AI neural network models to solve various translation tasks. Moreover, the tool was developed for the Lithuanian language, selecting English-Lithuanian-English and French-Lithuanian-French language combinations. *Tildė machine translation tool* was chosen since the tool was created to accommodate the needs of the Lithuanian language, in addition to other Baltic languages, and also considers the peculiarities of the Lithuanian language. Furthermore, this machine translation tool is also AI-technology-based, and systemic research has been carried out since the start of the machine translation tool development. Thus, results of the research can be embedded or compared with other forthcoming research works.

Though initially other tools were not considered to be applied in the experiment, in the last stages of the experiment, in 2023, the translation tool provided by a generative tool, i.e., *ChatGPT*, was tested. Yet, the results obtained cannot be further examined, as the experiment with the generative tool mentioned above was tried only several times and still needs to be replicated so that valid data could be produced and examined. The main reasons to try how *ChatGPT translator* reacts to the uploaded English .rc2 file and translates the strings into the Lithuanian language were related to the discussions and possibilities of various generative tools to produce the target file in the Lithuanian language and the fact that novice translators start applying such generative tools in the process of translation as a new translation practice. The author of the research took into consideration the fact that the results received in the process of localizing the source file into the Lithuanian language by means of *ChatGPT translator* could not be identical, yet, it was interesting to compare the quality of the output produced in terms of its readability and usability.

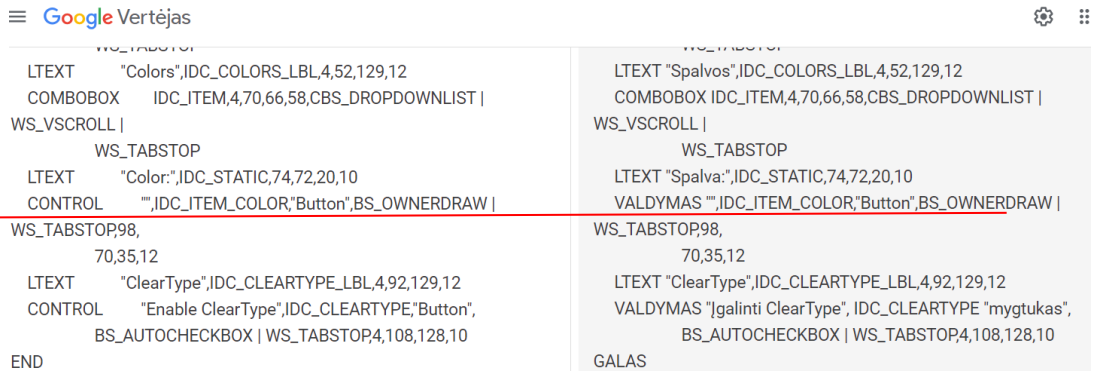
Participants of the experiment and their instruction

The participants of the experiment were professional translators with a Bachelor's degree in translation or linguistics. Most of them work in several translation agencies or carry out editing and proof-reading tasks in private translation companies and publishing houses. The total number of participants who took part in the experiment is 25 translators. 19 participants were women, and 6 of them were men. In 2021, 10 participants took part in the experiment, in 2022 – 6 and in 2023 – 9. Prior to the experiment and realisation of scenario 1 and 2, all participants were instructed about the procedures and process of how both scenarios should take place. No timing limitations were imposed in terms of duration or specific time setting during which participants of the experiment had to localize the resource file. They were allowed to carry out the Lithuanisation of the file at home. In addition to this, if they had no *Passolo* software, they were provided with it at Kaunas University of Technology. During the first stages of the experiment participants were asked to localize the selected resource file with *Notepad* and after that, upload and align the localized resource file in *Passolo*. Then they were asked to localize the resource file into the Lithuanian language with the four machine translation tools. *ChatGPT translator* was used by the last cohort of participants in 2023.

Results of the research

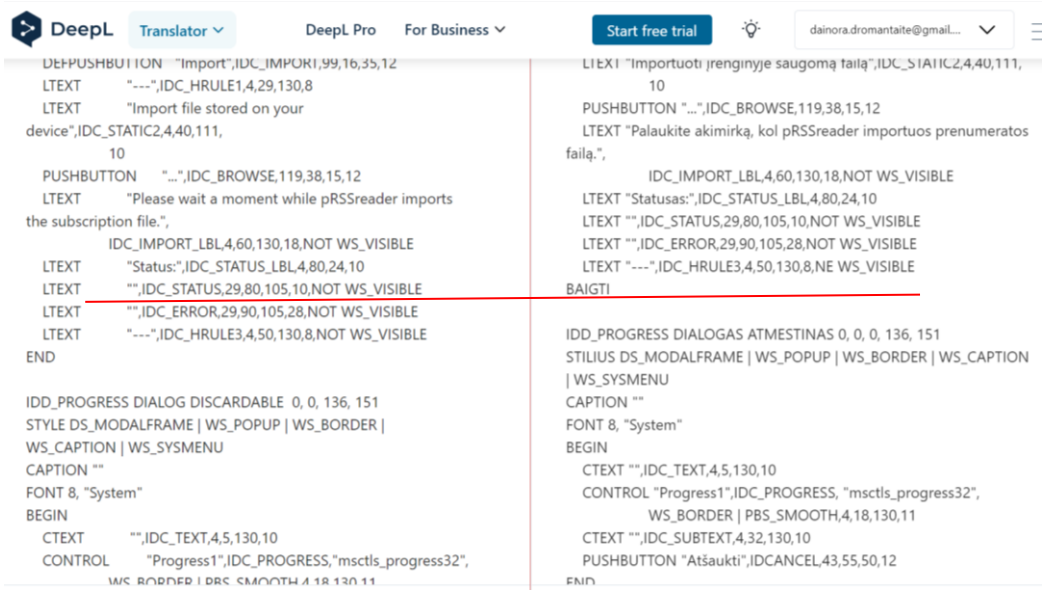
The experiment carried out returned interesting research findings in terms of comparing the accuracy, fluency, terminology, style and graphical layout of the output file in the Lithuanian language. Comparing the graphical layout of all machine translation tools, *Google Translate, Vilnius University machine translation tools, Tildė* and *DeepL,* it has been noticed that *Google Translate* and *DeepL* machine translation tools did not retain the graphical structure of the resource file in the Lithuanian language. For example, the structure of the target file was changed, different indentation of segments, spacing or no spacing were maintained in the target file. This might be related to the idea that both machine tools do not support .rc2 or .txt files and the .pdf file extension of the source file might have also caused incorrect segmentation of the localized file. The examples below demonstrate mistakes of machine translation tools in correctly extracting segments to be localized in the Lithuanian language:

Picture 3. ***Google Translate screen from English into the Lithuanian language***



The underlined example of *Google Translate* illustrates the example of segment truncation when no spacing is used in between of the programming language and the segments to be localized. Visual comparison on the left and the right also demonstrates that the graphical layout of the resource file in the source and target languages was changed. Moreover, the programming element "CONTROL" is rendered as the Lithuanian equivalent "VALDYMAS", though the segment should not be translated into the Lithuanian language at all.

Picture 4. ***DeepL translate screen into the Lithuanian language***



The picture above presents the example from *DeepL*. It showcases that the machine translation tool does not support .txt or .rc2 extension files and having uploaded the .txt file as a .pdf file, the machine translated output file was similar to *Google Translate* file in terms of its graphical outlay. The localized resource file into the Lithuanian language contained different spacing, indentation and segmentation in comparison to the source file.

However, *Vilnius University machine translation* tool recognized the .rc2 resource file but returned the output file with elements that should not be translated into the Lithuanian language. For instance, the item "8 FONT", "CAPTION", "BEGIN", "END" and others were rendered into Lithuanian language as "8 ŠRIFTAS", "KADRAS", (Engl. *Picture/Shot/Frame)* "PRADĖTI" (Engl. *Start*) or "PABAIGA" (Engl. *End*) and such cases were numerous, as demonstrated in the picture below:

Picture 5. ***VU machine translation tool translated output in the Lithuanian language***

```
//////////////////////////////////////////////////////////////////////////
// Ištekliai

//////////////////////////////////////////////////////////////////////////
//
// Dialogas
//

IDD_NEWS DIALOGO DISCARDABLE 0, 0, 139, 154
STILIAUS DS_MODALFRAME | WS_POPUP | WS_VISIBLE | WS_SYSMENU | WS_CAPTION
KADRAS „Naujienos“
8 ŠRIFTAS, „Sistema“
PRADĖTI
LTEXT „, IDC_SITE_NAME, 0,0, 138,15, SS_CENTERIMAGE
PABAIGA

IDD_IMPORT_OPML DIALOGO DISCARDABLE 0, 0, 136, 151
STILIAUS DS_MODALFRAME | WS_POPUP | WS_SYSMENU | WS_CAPTION
ANTRAŠTĖ „Importuoti OPML failą“
8 ŠRIFTAS, „Sistema“
PRADĖTI
LTEXT „Įveskite importuojamo OPML failo URL.“, IDC_STATIC1, 4,3,
130,10
EDITTEXT IDC_URL, 4,16,92,12, ES_AUTOHSCROLL
DEFPUSHBUTTON „Importas“, IDC_IMPORT, 99,16,35,12
LTEXT „---“, IDC_HRULE1, 4,29,130,8
LTEXT „Importuoti jūsų prietaise saugomą failą“, IDC_STATIC2, 40,111,
10
MYGTUKAS „...“, IDC_BROWSE, 119,38,15,12
LTEXT „Palaukite, kol pRSSreader importuos prenumeratos failą.“,
IDC_IMPORT_LBL, 4,60,130,18, NE WS_VISIBLE
LTEXT „Būsena:“, IDC_STATUS_LBL, 4,80,24,10
LTEXT „, IDC_STATUS, 29,80,105,10, NE WS_VISIBLE
LTEXT „, IDC_ERROR, 29,90,105,28, NE WS_VISIBLE
LTEXT „---“, IDC_HRULE3, 4.50.130.8, NE WS_VISIBLE
PABAIGA
```

The examination of the output file also revealed the fact that all the translatable segments were correctly detected by the machine translation tool. Yet, it failed to recognize some of the elements of the programming code. Still, it needs to be mentioned that the graphical layout, spacing, segmentation, and indentation of the target file were identical to the source file. The results might be related to the current technological improvements of the machine translation tool and the fact that it is further developed as a tool for webpage and software localization.

In terms of *Tildė machine translation tool,* an assumption was made that the machine translation tool should provide the best results due to the constant research and developments of the tool. The research and experiment demonstrated that the text's layout was retained, but some programming elements were rendered into the Lithuanian language, damaging the programming code of the segments. The segments that were Lithuanised were the same segments that were also localized by other machine translation tools. For instance, "CAPTION" was translated as "ANTRAŠTĖ" (Engl. *Headline*), the item "8 FONT" as "8 ŠRIFTAS", "BEGIN", "END" as "PRADĖTI"

(Engl. *Start*) or "PABAIGA" (Engl. *End*), "PUSHBUTTON" as "MYGTUKAS" (Engl. *Button*) and other. Yet, the output file contained no truncated segments and cases of different spacing, indentation or segmentation.

The comparison of the results of the four machine translation tools with the generative tool *ChatGPT* revealed the fact that the latter identified the translatable content and separated the programming language, however, the quality of the output file in terms of acceptability and usability was poor due to numerous lexical and grammatical mistakes or wrong terms. This might be related to the fact that the generative tool can generate both programming code and linguistic content. Yet, in terms of translation quality into the Lithuanian language, many improvements are needed so as to make the text usable and reader-friendly. Still, it is interesting to note that participants of the experiment copied separate segments that were produced by the tools to either *Notepad* or *Passolo* so as to speed up the process of localization. Since when the segments were copied from the machine translation outputs and pasted into the .txt file on *Notepad,* programming mistakes occurred. And when the files were aligned in *Passolo,* the segments were either truncated or presented incorrectly.

Picture 6. Tildė ***machine translated output in the Lithuanian language***

```
//////////////////////////////////////////////////////////////////////////// /
// ištekliai

//////////////////////////////////////////////////////////////////////////// /
//
// dialogas
//

IDD_NEWS DIALOGO DISCARDABLE 0, 0, 139, 154
STILIAUS DS_MODALFRAME | WS_KONTEKSTINIS MENIU | WS_VISIBLE | WS_CAPTION | WS_SYSMENU
ANTRAŠTĖ „Naujienos“
8 ŠRIFTAS, „System“
PRADĖTI
LTEXT „“, IDC_SITE_NAME, 0,0, 138,15, SS_CENTERIMAGE
BAIGTI

IDD_IMPORT_OPML DIALOGO DISCARDABLE 0, 0, 136, 151
STILIAUS DS_MODALFRAME | WS_IŠŠOKANTIS LANGAS | WS_SUBTITRAI | WS_SYSMENU
ANTRAŠTĖ „Importuoti OPML failą“
8 ŠRIFTAS, „System“
PRADĖTI
LTEXT „Įveskite importuotino OPML failo URL.“, IDC_STATIC1, 4,3,
130,10
EDITTEXT IDC_URL, 4,16 92,12, ES_AUTOHSCROLL
DEFPUSHBUTTON „Importuoti“, IDC_IMPORT, 99,16,35,12
LTEXT „---“, IDC_HRULE1, 4,29,130,8
LTEXT „Importuoti failą, saugomą jūsų prietaise“, IDC_STATIC2, 4,40,111,
10
MYGTUKAS „...“, IDC_BROWSE, 119,38,15,12
LTEXT „Palaukite, kol pRSSreader importuos prenumeratos failą.“,
IDC_IMPORT_LBL, 4,60,130,18, WS_VISIBLE
LTEXT „būsena:“, IDC_STATUS_LBL, 4,80,24,10
LTEXT „“, IDC_STATUS, 29,80,105,10, WS_VISIBLE
LTEXT „“, IDC_ERROR, 29,90,105,28, WS_VISIBLE
LTEXT „---“, IDC_HRULE3, 4,50,130,8, WS_VISIBLE
BAIGTI
```

Considering the application of consistent terminology and the use of standardized Lithuanian terms, it was observed that *Tildė's* output was the most consistent and fluent in terms of the correct and approved terminology use. Yet, all machine translation tools failed to use the recently standardized terms, as approved in the *Encyclopaedic Dictionary of Computer Terms* (2022), and their translation from English into the Lithuanian language. For instance, none of the machine translated outputs

rendered such words as “Manager” into the correct Lithuanian equivalent as “Tvarkytuvė“, “Reader” as the Lithuanian counterpart “Skaitytuvė”, “User” as “Naudotojas “ instead of “Vartotojas” (Engl. *Consumer*), or the button “Cancel” as “Atsisakyti” (Engl. *Refuse*) instead of another semantically close Lithuanian synonym “Atšaukti” (Engl. *Withdraw/Recall*), which is frequently confused with the latter. Here an assumption could be made that the machine translation tools failed to provide the right translation of the above-mentioned terms due to the fact that non-standardized terms are more often used in information and communication technologies related discourse in comparison with the approved ones. Programmers, computer specialists and software providers would apply non-standardized terms in the descriptions of their localized products, for instance, *Microsoft.* Moreover, it was interesting to observe that the participants of the experiment failed to recognize the improper use of the terms, mentioned above. The same mistakes of incorrect terminology or phrase use were observed in their final localizations of the resource file either in *Notepad* or *Passolo*. This might be related to the fact that they used the machine translation tools to translate the segments fast and since all of the four tools used the same and wrong Lithuanian counterparts, they did not question the correctness and validity of the terms used. Therefore, the same errors of incorrectly used terms were repetitive in the resource files that were localized using *Notepad* and *Passolo* programs.

The analysis of the machine-translated outputs also demonstrated inconsistencies in the correct language grammar use and style. All the machine translation tools failed to render the non-translatable items, such as “pRSSreader” or “ClearType”, into the Lithuanian language. The segments should have been Lithuanised by calque and partial localization when the item “pRSSreader” should be explained in terms of adding up the Lithuanian counterparts “Skaitytuvė” (Engl. *Reader*) next to the English item “pRSSreader” or “Mygtukas” (Engl. *Button*) next to the item “ClearType”. Moreover, the English words “pRSSreader” or “ClearType” in the segment “Enable ClearType” should have rendered into the Lithuanian language by placing them within the quotation marks. For instance, in the following segment “Palaukite, kol „pRSSreader“ importuos prenumeratos failą“ (Engl. *Please wait a moment while pRSSreader imports a subscription file*) the term “pRRSreader” should have been placed within the Lithuanian quotation marks due to the fact that the non-Lithuanian words have to be identified in a Lithuanian sentence.

In addition, cases where the output text does not correspond to the punctuation rules in terms of the fluency in the Lithuanian language are numerous. For example, the segment as produced by *VU translation tool* “Įveskite raktinį žodį (- US), JEI norite ieškoti:” (Engl. “*Enter keyword(s) to search for:”)* applies capital letters in the middle of the sentence, which is inappropriate for the Lithuanian language, since the interjection “Jei” (Engl., *If*) does not stand as a proper noun or an acronym that could be written using capital letters in the Lithuanian sentence. Moreover, the term “keywords” is

translated as "raktinis žodis" which is a literal translation from English into Lithuanian, yet, it should be replaced with another Lithuanian term "reikšminis žodis" (Engl. *notional/meaningful word*). And it is interesting to observe that the majority of participants of the experiment used inadequate translation of the English segment "keywords" and none of them selected the approved term "reikšminis žodis".

Finally, the examination of accuracy and style of the outputs, produced by the machine translation tools, revealed the fact that all the machine translation tools provide rather accurate translations in terms of the message rendered in the Lithuanian language and recognise the genre of such technical texts. The style of the outputs produced was appropriate for this type of text, despite the fact that in some cases it could have been more coherent. For instance, the English segment "Import file stored on your device" was produced by *Tildė machine translation tool* applying the strategy of literal translation with slight modifications, i.e., "Importuoti failą, saugoma jūsų prietaise". Though the Lithuanian counterpart, produced by the machine translation tool is correct, the style of the sentence and the word order could be replaced by the following sentence, i.e., "Importuoti įrenginyje saugomą failą" (Engl. *Import the file on the device*). Yet, the machine-translated segment is accurate and the translator could edit it by eliminating the Lithuanian pronoun "jūsų" (Engl. *Your*) and making its style more coherent. Such corrections would not require much of a cognitive effort of the translator in comparison to editing of lexical mistakes which requires more of the attention, precision, particularity and critical thinking of the translator.

Conclusions

The interest in machine translation tools, their usability, and different research on various aspects of machine translation tools is constantly increasing and revving up. Yet, the paper indicates that the research on the application of machine translation tools and their integration in processes of software localization for Lithuanian as the low-resourced language is rather limited and scarce during the period of this paper publication. Moreover, the investigation on how software localization into the Lithuanian language can be powered by machine translation tools is rather limited, despite the fact that some research has been carried out into the output quality of machine translated texts, examination of possible and practical challenges, and machine translation acceptance and usability from the perspective of Lithuanian users. Therefore, this paper is one of the first pilot studies to examine the usability of several machine translation tools in software Lithuanisation.

The experiment did not aim to provide specific statistical calculations of machine-made mistakes, but rather focused on the usability of the tools with 25 users who had to localize the resource file by means of *Notepad, Passolo* and the four machine translation tools. The results of the experiment demonstrated the fact that the participants, though

informed about the processes of the experiment, did use the machine translation tools in scenario 1, prior to the completion of scenario 2. This resulted in rather fast production of target files, yet the mistakes that occurred in the machine translated outputs were repetitively observed in all translations. This leads to a conclusion that when similarly translated segments are provided by different machine translation tools, translators do not necessarily critically evaluate the suggestions of the machine translation tools or double-check terminology, grammar, syntax, accuracy, and word use and instead trust the quality of the output produced.

The results of the research also demonstrated that in terms of the resource file recognition, *Tildė and VU machine translation* tools can handle and render .rc2 or .txt format files, while *Google Translate* and *DeepL* can deal with the mentioned format when the file is uploaded as a .pdf file. In terms of term consistency, accuracy, fluency, and style, *Tildė* translated machine translation output was the most cohesive and fluent. This might be related to the fact that the machine translation tool takes into account the aspects and peculiarities of the Lithuanian language. Furthermore, the tool is constantly technologically improved. Whereas *VU machine translation tool*, though developed mainly for the Lithuanian language, contains more mistakes, incorrect use of terminology and lacks text cohesion. However, further research needs to be carried out to statistically evaluate the percentage of mistakes and output quality, to provide more reliable and statistically grounded conclusions in terms of machine translation usability and integration in software localization for the Lithuanian language.

Furthermore, it must be mentioned that all four machine translation tools can be used in the process and workflow of resource file localization at the segment or string level. If rendered coherently and accurately, the segments could be edited and pasted into the screens of *Passolo* translators. Yet, attention should be paid when copying separate segments and making sure that when the machine-translated segment is copied, it is not truncated or presented incorrectly. In addition to that, the output of machine-translated segment and string should be double-checked to make sure that it is appropriate for the target language. This would require more time and effort; therefore, further research is needed to determine if such multitasking increases productivity of translators and localizers.

Bibliography

Canfora, Carmen; Ottmann, Angelika. 2020. Risks in neural machine translation. Transl. Spaces. 9: 58–77.

Castilho, Sheilla. 2016. Measuring Acceptability of Machine Translated Enterprise Content. Ph.D. Thesis, Dublin City University, Dublin, Ireland.

Cvilikaitė, Jurgita. 2008. Leksinės mašininio vertimo klaidos: beekvivalenčių žodžių vertimas. Filologija 18. 27–38. https://talpykla.elaba.lt/elabafedora/objects/elaba:6138715/datastreams/MAIN/content. Accessed on: September 15, 2023.

Maumevičienė, Dainora. 2023. Integration of Machine Translation Tools in Software Localization: mission (im)possible? In: L10N Journal 2(2), pp. 62–79.

Daudaravičius, Vidas. 2006. Pradžia į begalybę. Mašininis vertimas ir lietuvių kalba. Darbai ir dienos, 45. 7–18. https://www.ceeol.com/search/article-detail?id=209872. Accessed on: August 23, 2023.

Das, P.; Kuznetsova, A.; Zhu, M.; Milanaik, R. 2019. Dangers of Machine Translation: The Need for Professionally Translated Anticipatory Guidance Resources for Limited English Proficiency Caregivers. Clin. Pediatr.58: 247–249. PMID: 30392383.

Dew, K.N.; Turner, A.M.; Choi, Y.K.; Bosold, A.; Kirchhoff, K. 2018. Development of machine translation technology for assisting healthcommunication: A systematic review. J. Biomed. Inform., 85, 56–67.

Doherty, Stephen. 2016. The Impact of Translation Technologies on the Process and Product of Translation. https://ijoc.org/index.php/ijoc/article/viewFile/3499/1573. Accessed on: September 15, 2023.

Enciklopedinis kompiuterijos terminų žodynas //Encyclopaedic Dictionary of Computer Terms. 2022. http://www.ims.mii.lt/ALK%C5%BD/angl.html.

European Commission. 2022. European Master's in Translation Competence Framework. https://commission.europa.eu/system/files/2022-11/emt_competence_fwk_2022_en.pdf. Accessed on: October 15, 2023.

Gaspari, F.; Almaghout, H.; Doherty, S. 2015. A survey of machine translation competences: Insights for translation technologyeducators and practitioners. Perspectives. 23: 333–358.

Hoi, Huynh, Tan. 2020. Machine Translation And Its Impact In Our Modern Society. Int. J. Sci. Technol. Res. 9, 1918–1921. https://www.researchgate.net/publication/341650157_Machine_Translation_And_Its_Impact_In_Our_Modern_Society. Accessed on: September 20, 2023.

Kasperavičienė, Ramunė; Motiejūnienė, Jurgita, Patašienė, Irena. 2020. Quality Assessment of Machine Translation Output: Cognitive Evaluation Approach in An Eye-Tracking Experiment. Texto Livre: Linguagem e Tecnologia, 13(2). 271–285. https://doi.org/10.35699/1983-3652.2020.24399.

Kasperė, Ramunė; Horbačauskienė, Jolita; Motiejūnienė, Jurgita; Liubinienė, Vilmantė; Patašienė, Irena; Patašius, Martynas. 2021. Towards sustainable use of machine translation: usability and perceived quality from the end-user perspective // Sustainability. Basel : MDPI. ISSN 2071-1050. 2021, vol. 13, iss. 23, art. no. 13430, p. 1-17. DOI: 10.3390/su132313430.

Kenny, Dorothy; Moorkens, Joos; do Carmo, Felix. 2020. Fair MT: Towards ethical, sustainable machine translation. Transl. Spaces 2020, 9, 1–11.

Labutis, Vitas. 2005. Išaugusi vertėjų paklausa–nauji pavojai lietuvių kalbai. Bendrinė kalba(iki 2014 metų–Kalbos kultūra), 78. 205–209. https://www.ceeol.com/search/arti-cle-detail?id=93919. Accessed on: August 23, 2023.

Lommel, Arle; Burchardt, Aljoscha; Uszkoreit, Hans. 2014. Multidimensional Quality Metrics (MQM): A framework for declaring and describing translation quality metrics. Tradumàtica: tecnologies de la traducció, 0:455–463.

Petkevičiūtė, Inga, Tamulynas, Bronius. 2011. Kompiuterinis vertimas į lietuvių kalbą: alternatyvos ir jų lingvistinis vertinimas. Kalbų studijos 18. 38–45. https://etalpykla.lituanistikadb.lt/object/LT-LDB-0001:J.04~2011~1367174892606/. Accessed on: August 23, 2023.

Pituxcoosuvarn, Mondheera; Ishida, Toru. 2018. Multilingual communication via best-balanced machine translation. New Gener. Comput.36, 349–364.

Rimkutė, Erika; Kovalevskaitė, Jolanta. 2007. Mašininis vertimas–greitoji pagalba globalėjančiam pasauliui. Gimtoji kalba, 9. 3–10. http://donelaitis.vdu.lt/lkk/pdf/MV2.pdf. Accessed on: June, 2023.

Rimkutė, Erika; Kovalevskaitė, Jolanta. 2008. 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.

Maumevičienė, Dainora. 2023. Integration of Machine Translation Tools in Software Localization: mission (im)possible? In: L10N Journal 2(2), pp. 62–79.

Rossetti, A.; O'Brien, S.; Cadwell, P.2020. Comprehension and Trust in Crises: Investigating the Impact of Machine Translation and Post-Editing. In Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, Lisbon, Portugal, 3–5 November 2020; European Association for Machine Translation: Lisboa, Portugal, 2020; pp. 9–18.

Rossi, Caroline; Chevrot, Jean-Pierre. 2019. Uses and perceptions of Machine Translation at the European Commission. J. Spec. Transl. (JoSTrans).31: 117-200.

Saadany, H.; Orasan, C.; Quintana, R.C.; do Carmo, F.; Zilio, L.2021. Challenges in Translation of Emotions in Multilingual User-Generated Content: Twitter as a Case Study. arXiv, arXiv:2106.10719.

Skadiņa, Iguna. 2019. Some Highlights of Human Language Technology in Baltic Countries. In Databases and Information Systems X: Selected Papers from the Thirteenth International Baltic Conference, DB&IS 2018. Vol. 315. ISOS Press: pp.18-32.

Skadiņš, Raivis; Pinnis, Marcis; Vasiļjevs, Andrejs; Skadiņa, Iguna; Hudik, Toma. 2014. Application of Machine Translation in Localization into Low-Resourced Languages. http://aclanthology.lst.uni-saarland.de/2014.eamt-1.43.pdf. Accessed on: September 1, 2023.

Utka, Andrius; Amilevičius, Darius; Krilavičius, Tomas; Vitkutė-Adžgauskienė, Daiva. 2016. Overview of the Development of Language Resources and Technologies in Lithuania (2012–2015). In Human language technologies - the Baltic perspective: proceedings of the 7th international conference, Baltic HLT 2016, Riga; pp. 12–19.

Utka, Andrius; Vaičenonienė, Jurgita; Briedienė, Monika; Krilavičius, Tomas. 2020. Development and Research in Lithuanian Language Technologies (2016–2020). In Human language technologies - the Baltic perspective: proceedings of the 9th international conference, Baltic HLT 2020, Kaunas; Amsterdam : IOS Press; pp.215–222.

Yamashita, Naomi; Ishida, Toru. 2006. Effects of Machine Translation on CollaborativeWork. In Proceedings of the 2006 20th Anniversary Conference on Computer Supported CooperativeWork (CSCW '06), Banff, AB, Canada, 4–8 November 2006; Association for Computing Machinery: New York, NY, USA; pp. 515–524.

Yasuoka, M.; Bjorn, P. 2011. Machine Translation Effect on Communication: What Makes It Difficult to Communicate through Machine Translation? In Proceedings of the 2011 Second International Conference on Culture and Computing, Kyoto, Japan, 20–22 October 2011; pp. 110–115.