



DE GRUYTER
OPEN

10.1515/llce-2017-0005

Issues in Machine Translation

A case of mobile apps in the Lithuanian and English language pair

Gilvilė Stankevičiūtė

Kaunas University of Technology, Lithuania
gilvile.stankeviciute@ktu.edu

Ramunė Kasperavičienė

Kaunas University of Technology, Lithuania
ramune.kasperaviciene@ktu.lt

Jolita Horbačauskienė

Kaunas University of Technology, Lithuania
jolita.horbacauskiene@ktu.lt

Abstract

Machine translation (MT) is still a huge challenge for both IT developers and users. From the beginning of machine translation, problems at the syntactic and semantic levels have been faced. Today despite progress in the development of MT, its systems still fail to recognise which synonym, collocation or word meaning should be used. Although mobile apps are very popular among users, errors in their translation output create misunderstandings. The paper deals with the analysis of machine translation of general everyday language in Lithuanian to English and English to Lithuanian language pairs. The results of the analysis show that more than two thirds of all the sentences were translated incorrectly, which means that there is a relatively small possibility that a mobile app will translate sentences correctly. The results are disappointing, because even after almost 70 years of MT research and improvement, researchers still cannot offer a system that would be able to translate with at least 50% correctness.

Keywords

Machine translation, mobile apps, types of errors, the Lithuanian and English language pair

1 Overview of machine translation

At the beginning of the 20th century, it was already obvious that fully automated high quality translation was not possible and both pre- and post-editing would be needed (Hutchins, 2010). Dostert claimed that there were 2 main issues in MT: a unit of words in the source language may have more than one possible equivalent in the target language; and the order of source language sentences or units may not be suitable for the output in the target language (1957). Post-editing was considered to be unavoidable, because an MT system would have to deal with flexible syntax, vocabulary and a variety of subjects and different genres of discourse (Vasconcellos, Marjorie, 1988). Due to this, it has been very difficult to transfer a sense from one language to another for MT systems. Nevertheless, Stefcik (2015) points out that “MT is used with caution where a high degree of quality is required, but in translation of certain types of texts MT may be successfully applied” (p. 140).

Some of the most popular MT systems are SYSTRANet (includes 35 pairs of languages), Babelfish (includes 35 pairs of languages), Google Translator (includes 25 pairs of languages), InterTran (includes 29 pairs of languages), and Wordlingo (includes 14 pairs of languages), etc. (Rimkutė & Kovalevskaitė, 2007a). The majority of these MT systems translate single words, sentences, texts, files, websites, emails, etc. in the most popular languages such as English, German, Russian, Polish, French, Czech, Italian, Spanish, Chinese, Japanese, Korean, Arabic, and Swedish (ibid). However, MT development is an ongoing and continuous process because of structural and lexical differences of source and target languages and polysemy.

A technology for intercultural oral communication integrates automatic speech recognition and MT speech translation. A speech translation system consists of an automatic speech recogniser (ASR), a machine translator, and text to speech system. The majority of ASR systems are developed and structured in order to minimise the word error rate (WER). However, WER counts word errors only at the surface level, not considering context and syntax, which are often critical for MT (He; Deng, 2008). Over the last decade, a lot of publications and research projects have demonstrated that ASR and MT could be combined for direct translation of spoken language (Grazina, 2010).

Zheng et al. (2010, p. 113) claim that “large-vocabulary S2S translation system typically uses multiple computing intensive components and requires large memory availability to store models in dynamic data structures, therefore posing significant challenges to developers.” In order to adapt speech recognition system to the purposes of small languages, more advanced techniques have to be developed.

A number of mobile applications have been developed including text-to-text, text-to-speech, and speech-to-text. They are helpful in providing, for example, travellers with a quick and simple way to communicate in an unfamiliar language. The first mobile translation apps have been developed recently (Google Translate App, Duolingo,

Busuu, iTranslate, etc.). Miguel A. Jimenez-Crespo (2016) has made research into mobile apps and translation crowdsourcing, which revealed that smartphones have occupied the place of desktop computers. The goal of mobile translation apps “is to deliver flexible, dynamic and quick translations of varying degrees of quality” (Jiménez-Crespo, 2016). This degree of translation quality is still low and requires more or less of improvement depending on the type of text.

2 Classification of machine translation errors

MT receives a lot of criticism (Labutis, 2005; Rimkutė; Kovalevskaitė, 2007, 2007a). However, cost, speed, and size of dictionaries are the factors that largely determine the quality of machine-generated translations. Flanagan (1994) identifies a few reasons why machine translation quality is difficult to evaluate: a text can have more than one correct translation; errors can involve not only single words but also phrases, discontinuous expressions, word order or relationships across sentence boundaries, and as noted by Vilar et al. (2006) it makes simple counting of the number of wrong words in the translation pointless; one error can lead to another; and the cause of errors in MT output is not always clear, as the evaluator usually does not have ability to trace the tests and actions in the software, making it difficult to identify what went wrong in the text translation.

Flanagan’s well-known classification scheme based on common MT errors was created for the English-French and English-German languages pairs (Hsu, 2014). The taxonomy consists of 19 categories: spelling, not found words, inflection, rearrangement, category, content and function words, agreement, cause boundary, word selection, and expression. However, this classification is applicable in the 21st machine translation as a framework for error identification (Stymne, Ahrenberg, 2012; Elliott, Hartley, and Atwel, 2004; Valotkaitė and Asadullah, 2012, Climent et al., 2003). This paper employs the taxonomy proposed by Vilar, Xu, D’Haro, and Ney (2006) for error identification in machine translation (see Figure 1).

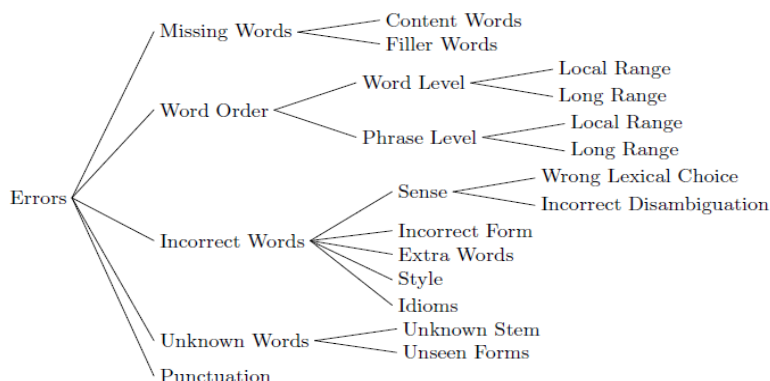


Figure 1. Classification of translation errors by Vilar, Xu, D'Haro and Ney (2006)

Five subcategories are distinguished: sense, incorrect form, extra words, style, and idioms. For highly inflected languages like Lithuanian, where a variety of inflections of the content word classes produces a problem for machine translation, the MT system is not able to generate the correct word form, even though the translation of the base form is correct. Errors of unknown words (further subdivided into unknown words or stems) and unseen forms of known stems are common for the pairs of the majority of European languages, because unknown words can be rendered by copying the input word in order to produce a sentence. Punctuation errors are only important when they interfere with the logical structure of a sentence. This taxonomy has been employed by a number of researchers (Popovich; Ney, 2011; Avramidis; Koehn, 2008; Stymne, 2011) as a framework for error analysis in machine translation.

3 Methodology

The paper analyses 400 sentences (100 compound and 100 simple sentences in English, 100 compound and 100 simple sentences in Lithuanian) recorded into 2 different mobile translation (speech to speech) apps. The sentences were taken from English-to-Lithuanian and Lithuanian-to-English phrase books published for tourists, as they contain the most useful and common sentences, which become of vital importance for a person without basic knowledge of the language spoken in a particular country, if he or she wants to be understood. For the purposes of the research, it was considered that the automatic speech recognition system recognised sentences correctly. Therefore, only MT errors were analysed and identified according to the taxonomy proposed by Vilar, Xu, D'Haro and Ney (2006). Punctuation and style mistakes were excluded, as only individual sentences were used and style errors were covered by the incorrect word category. The word-for-word translation subcategory was added since

some MT output sentences were non-editable due to a large number of mistakes. The analysis was performed in Lithuanian to English and English to Lithuanian language pairs. The apps chosen were developed by international (App 1) and Lithuanian IT developers (App 2).

4 Results

Error rate analysis in simple and compound sentences. The analysis of recorded and translated sentences demonstrated that almost 70% of all the sentences (553 sentences out of 800) were translated incorrectly (see Figure 2 and Figure 3).

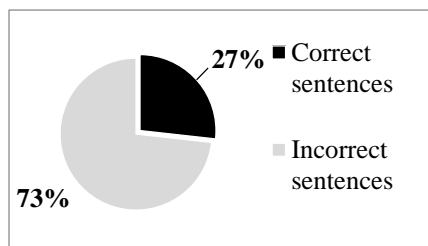


Figure 2. Correct and incorrect sentences translated by Application 1

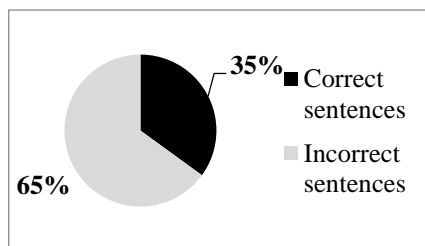


Figure 3. Correct and incorrect sentences translated by Application 2

More than two-thirds of all the sentences were translated incorrectly, which means that there is only a 33% possibility that an MT system will translate sentences correctly, which is too disappointing as such high number of mistakes most probably will lead to misunderstandings. In order to understand what types of sentences and which language pair translations cause the major issues, the results of simple and compound sentence translations were analysed separately. They also were distributed into English to Lithuanian and Lithuanian to English categories (see Figure 4 and Figure 5). It was discovered that the apps made fewer mistakes in translation from Lithuanian into English. It may be assumed that MT systems struggle with the Lithuanian language because it is a synthetic and morphologically rich language. The system fails to understand how relations between words in the sentence work.

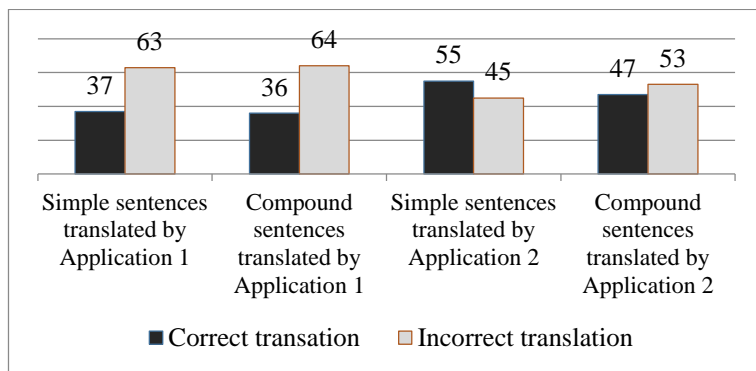


Figure 4. Correctly and incorrectly translated sentences from Lithuanian to English

Figure 4 shows that in translation of simple sentences from Lithuanian into English by Application 2 the number of correct sentences (55 of 100) was higher than the number of sentences translated incorrectly (45 of 100). Meanwhile, Application 1 succeeded to translate only 37 instances of simple sentences correctly, producing incorrect translation in 63 instances. The highest rate of errors in translation from Lithuanian into English was made in compound sentences translated by Application 1 (64 incorrect vs 36 correct), while Application 2 provided almost equal results in translating compound sentences (47 correct vs 53 incorrect).

Figure 5 represents the results of simple and compound sentence translation from English to Lithuanian. Less than 30% of translated sentences were correct (see Figure 5).

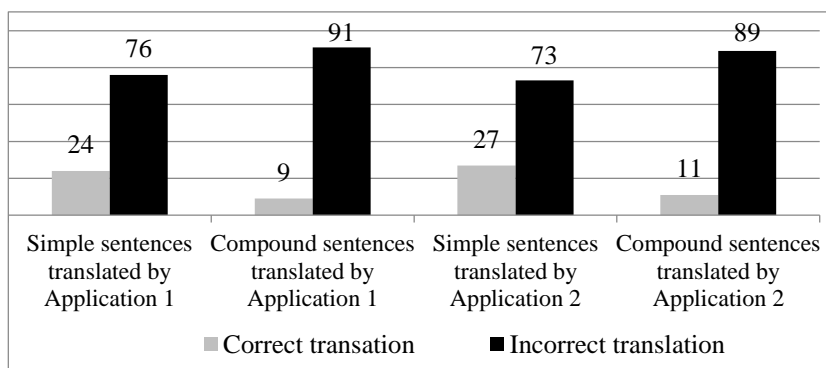


Figure 5. Correctly and incorrectly translated sentences from English to Lithuanian

Like in the Lithuanian to English translation, Application 2 showed slightly better results: only 27 simple and 11 compound sentences were accurate when translating from

English to Lithuanian. Nevertheless, the number of errors made by both applications was approximately equal when translating from English to Lithuanian. However, the Lithuanian to English translations displayed better translation quality, providing the best output by Application 2.

Analysis of missing word errors. All the missing word errors were divided into missing content and filler word errors in simple sentences and missing content and filler word errors in compound sentences. In Lithuanian to English translation, both mobile applications provided similar results. However, slightly better results were observed in the translation by Application 2, e.g.

- (1) Lithuanian: *Atleisk, kad taip užtrukau atrašyti.*
 forgive-IMP that so take-1P-SG-PAST write-INF
 App 1: *Sorry that it took me to write back.*
 App 2: *Sorry it's taken me so long to write.*
- (2) Lithuanian: *Man rodos, Jūs važiuojate ne ta*
 me-DAT seem-3P-PRES you-PL go-2P-PL-PRES not that
kryptimi.
 direction-INTR
 App 1: *I am afraid you are going in the wrong direction.*
 App 2: *It seems to me, to go in the wrong direction.*

In Example 1, the missing content word *taip* (adverb) may not have other meanings in application memories, that being a cause for error. However, Application 2 recognised the word and rendered it correctly. In Example 2, the mistake made by MT was the missing filler word *jūs*. The pronoun was present in the Lithuanian sentence, and Application 2 failed to render it in the translated version of the sentence.

In the translation from English to Lithuanian, missing content and filler word errors were also present, e.g.

- (3) English sentence: *Sorry to keep you waiting.*
 App 1: *Gaila, kad jus laukia.*
 regrettably that you-PL-ACC wait-3P-PRES
 App 2: *Atsiprašome jus laukia.*
 apologise-1P-PL-PRES you-PL-ACC wait-3P-PRES
 Correct: *Atsiprašau, kad priverčiau laukti.*
 apologise-1P-SG-PRES that force-1P-SG-PRES wait-INF

In Example 3, the verb *priverčiau* (force-SUBJ-1P-SG) is missing. The MT system did not recognise and identify the meaning of the phrase *keep someone waiting* in the source sentence, which is why the MT system failed to translate the sentence correctly.

(4) English: *I would like something **for** insomnia.*

App 1: *Norėčiau ko nors nemiga.*

want-SUBJ-1P-SG something insomnia-NOM-SG

App 2: *Norėčiau kažką nemigos.*

want-SUBJ-1P-SG something-ACC insomnia-GEN-SG

Correct: *Norėčiau ko nors **nuo** nemigos.*

want-SUBJ-1P-SG something for insomnia-GEN-SG

As the example above demonstrates, both applications failed to render the preposition in the target sentence due to failure of MT systems to identify the meaning of the phrase *something for*.

The errors of missing words in the target sentences most probably appear due to the inability of MT systems to recognise the relations between words (i.e. subject and predicate). The size of the database may also play a role since the developers of mobile applications normally reduce the size of the database in order to ensure good performance of applications.

Analysis of word order errors. The majority of the mistakes occurred at the phrase level in both Lithuanian to English and English to Lithuanian translations, e.g.

(5) Lithuanian: *Ar galėtumėte man **tai** suvynioti?*

QPT can-SUBJ-2P-PL me-DAT that wrap-INF

App 1: *Would you mind **me wrap** this?*

App 2: *Can you tell **me it wrapped**?*

Correct: *Could you **wrap it for me**, please?*

(6) English: *Thank you for everything you **have done for me**.*

App 1: *Ačiū jums už viską, ką tu padarei*

thank you-DAT for everything-ACC that you-SG PREF-do-SG-2P-PAST
man.

me-DAT

App 2: *Dėkojame už viską, ką padarei*

thank-1P-PL-PRES for everything-ACC that PREF-do-SG-2P-PAST
man

me-DAT

Correct: *Dėkoju už viską, ką dël manęs*

thank-1P-PL-PRES for everything-ACC that for me-GEN
padarėte.

PREF-do-PL-2P-PAST

Examples 5 and 6 indicate that the mistakes of possessor order were made. Both applications failed in translation. This happened because MT systems followed the word order of the source language and not the target language. It might be assumed that MT systems face two challenges: different word order in different languages and sentence sense recognition.

Analysis of unknown words. If the MT system cannot recognise the word in a source sentence, the word is left untranslated in the target sentence. Application 2 was observed to have a richer vocabulary and contained more information about words and word formation, making it more suitable for users. Application 1 in the Lithuanian to English translations produced a greater number of mistakes than in the English to Lithuanian translation; meanwhile, Application 2 failed in fewer cases, regardless of the translation direction.

In a few cases, the MT system did not translate core words, e.g.

- (7) Lithuanian: *Likime* stay-IMP-1P-PL *draugais.* friend-PL-INSTR

App 1: *Likime* friends.

App 2: *Let's stay* friends.

- (8) Lithuanian:

Paskambinkite man kitą kartą kai norėsite,
call-IMP-2P-PL me-DAT next-SG-ACC time-SG-ACC when want-2P-PL-FUT
kad jūs kur nors nuvežčiau.

that you-ACC-PL where some take-SUBJ-1P-SG

App 1: *Call me next time when you want to make you a nuvežčiau.*

App 2: *Call me the next time when you want to keep you somewhere nuvežčiau.*

Correct: *Call me next time you need a lift anywhere.*

The length of sentences, in fact, does not influence the error rate. Sometimes MT systems are not able to recognise one word, i.e. *nuvežčiau* (take-SUBJ-1P-SG; Example 7) or *likime* (stay-FUT-1P-PL; Example 8). Application 2 managed to recognise the word; however, other mistakes occurred in the target sentence. It is interesting to notice that Application 1 identified the verb *nuvežčiau* (subjunctive mood) as a noun, adding an indefinite article.

In the translation from English to Lithuanian, untranslated words were also frequent in both applications, e.g.

- (9) English: *If you'd like to meet up sometime, let me know!*

App 1: *Jei norite susitikti kažkada, let me know!*

if want-2P-PL meet-INF sometime, let me know

App 2: *Jei norėtųėte susitikti kada, let me know!*

if want-2P-PL meet-INF when, let me know

Correct: *Jei norėtum kada susitikti, pranešk!*

if want-SUBJ-2P-SG when meet-INF tell-INF

The phrase let me know in Example 9 is a collocation, but it was not recognised and was left untranslated in both applications.

MT systems may not be able to identify single words and lexical units. Sometimes MT systems translate phrases word-for-word. However, the length and the complexity of sentences has no influence on the frequency of this type of errors. Poor vocabulary and insufficient databases may be considered the main factors for appearance of such errors.

Analysis of incorrect word errors. MT error analysis showed that incorrect word errors (wrong lexical choice, disambiguation, incorrect form, extra word, idiom and word-for-word translation) were the most common mistakes in both applications. The correct lexical choice was the greatest challenge for MT translation systems, accounting for the majority of incorrect word errors, e.g.

(10) Lithuanian:

Greitojo	traukinio	išvykimas	į
fast-SG-M-GEN-PRONOM	train-SG-M-GEN	departure-NOM-M-SG	to
Londoną	numatomas	keturioliktą	
London-ACC	schedule-PAST-PTC-SG-NOM	fourteen-ACC-SG-FM	
valandą.			
hour-ACC-SG-FM			

App 1: *Speed train departure to London expected the fourteen hour.*

App 2: *High-speed train departure to London expected the fourteenth hour.*

Correct: *The departure of the London express is scheduled for 2 p.m.*

(11) English: *The ferry to Klaipėda is anchored at pier number three.*

App 1:

Kelto	į Klaipėdą	yra	įtvirtinta
ferry-SG-M-GEN	to Klaipėda-ACC	be-3P-PRES	PREF-fix-SG-F-PAST-PTC
priplaukos	skaičius	numeris 3.	
prier-SG-F-GEN	digit-NOM-SG-M	number-SG-M-NOM	3

App 2:

Keltas	į Klaipėdą	yra	įtvirtinti
ferry-SG-M-GEN	to Klaipėda-ACC	be-3P-PRES	PREF-fix-PL-M-PAST-
molo	numeris	trys.	
PTC prier-SG-M-GEN	number-SG-M-NOM	three-NOM.	

Correct:

Keltas *į Klaipėdą* *prišvartuotas*
 ferry-SG-M-GEN to Klaipėda-ACC be-3P-PRES PREF-wharf-SG-M-PAST-PTC
prieplaukoje *numerys* *trys*.
 pier-SG-F-GEN number-SG-M-NOM three-NOM

Examples 10 and 11 illustrate that MT systems may choose a wrong synonym due to which the target sentences gain a different or unclear meaning. The users may still guess the meaning, but because of an incorrect word the sentence does not sound natural.

Incorrect form mistakes also posed a greater challenge in translation into Lithuanian than into English. That is because the words and connections between the sentences are managed by word endings in the Lithuanian language. In translation into Lithuanian, MT systems failed to select the correct inflection as they were not able to identify the relations between words accurately and to link them, e.g.

(12) English: I

App 1 and 2: *Ar turite kopūstų sriuba?*
 QPT have-PL-2P cabbage-PL-M-GEN soup-SG-F-NOM
 Correct: *Ar turite kopūstų sriubos?*
 QPT have-PL-2P cabbage-PL-M-GEN soup-SG-F-GEN

In Example 12, the MT-provided translations are incorrect due to disagreement in the inflections required by the Lithuanian language grammatical rules.

Languages under investigation are different in their structure, Lithuanian being highly inflectional. Hence, both applications struggled to choose correct words or word meanings, and forms, due to which target sentences appeared to have unclear meanings or sound unnatural.

5 Final implications

A lot of MT issues still remain unsolved. The analysis of the results of translations in Lithuanian to English and English to Lithuania indicate that MT apps face deep syntactical and lexical issues. Translation from English to Lithuanian is much worse than the one from Lithuanian to English due to a few reasons: Lithuanian is a relatively small language also being a synthetic, highly inflectional language where the relations between elements of the sentence are defined by word endings.

Both apps struggled most in choosing a correct word or word meaning, and a word form. Unknown word, missing word and word order errors were also frequent. The results are disappointing, because even after almost 70 years of MT research, researchers still cannot offer a system that would be able to translate sentences with at least 50% correctness. One of the factors influencing mobile translation apps results is

the type of the chosen text. The analysis and comparison of mobile translation apps output in a variety of texts may be the next step for highlighting the points for improvement of MT apps.

ABBREVIATIONS

ACC – Accusative	P – person
DAT – dative	PAST – past
F –feminine	PL – plural
FUT – future	PREF – prefix
GEN – genitive	PRES – present
IMP – imperative	PRONOM - pronominal
INF – infinitive	PTC – participle
INSTR – instrumental	QPT – question particle
M – masculine	SG – singular
NOM – nominative	SUBJ - subjunctive

References

- Avramidis, E., & Koehn, P. (2008). Enriching Morphologically Poor Languages for Statistical Machine Translation. *Proceedings of ACL 08HLT*: 763-770.
- Climent, S., Moré, J., Oliver, A., Salvatierra, M., Sánchez, I., Taulé, M., & Vallmanya, L. (2003). Bilingual newsgroups in Catalonia: A challenge for machine translation. *Journal of Computer-Mediated Communication* 9(1).
<http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2003.tb00360.x/full>
- Dostert, L. (1957). *Research in machine translation: Report of the 8th Annual Round table meeting on Linguistics and language studies*. Washington, D.C.: Georgetown University Press.
- Elliott, D., Hartley, A., & Atwell, E. (2004). A fluency error categorization scheme to guide automated machine translation evaluation. *Machine Translation: From Real Users to Research*. Berlin Heidelberg: Springer, pp. 64-73.
- Flanagan, M. (1994). Error classification for MT evaluation. *Technology Partnerships for Crossing the Language Barrier: Proceedings of the First Conference of the Association for Machine Translation in the Americas*. pp. 65-72.
- Grazina, N. M. M. (2010). Automatic Speech Translation: MA thesis. Universidade Técnica de Lisboa. Retrieved from <http://www.inesc-id.pt/pt/indicadores/Ficheiros/5512.pdf>
- He, X.; Deng, L. (2008). *Discriminative Learning for Speech Recognition: Theory and Practice*. Morgan & Claypool Publishers.
- Hsu, J. A. 2014. Error Classification of Machine Translation A Corpus-based Study on Chinese-English Patent Translation. *Translation Studies Quarterly* 18:121-136.

- Hutchins, W. H. (2010). Machine translation: a concise history. *Journal of Translation Studies* 13(1-2): 29-70.
- Jimenez-Crespo, Miguel A. (2016). Mobile apps and translation crowdsourcing: the next frontier in the evolution of translation. *Revista Tradumàtica: tecnologies de la traducció* 14:75-84.
- Labutis, Vitas. (2005). Išaugusi vertėjų paklausa – nauji pavojai lietuvių kalbai. *Kalbos kultūra* 78: 205–209.
- Popović, M. & Ney, H. (2011). Towards automatic error analysis of machine translation output. *Computational Linguistics* 37(4): 657-688.
- Rimkutė, E. & Kovalevskaitė, J. (2007a). Mašininis vertimas – greitoji pagalba globalėjančiam pasauliui. *Gimtoji kalba* 9: 3-10.
- Rimkutė, E. & Kovaliauskaitė, J. 2007b. Pastabos apie žmogaus ir mašininį vertimą. *Gimtoji kalba* 10: 11–19.
- Štefčík, J. (2015). Evaluating machine translation quality: a case study of a translation of a verbatim transcription from Slovak into German. *Vertimo Studijos* 8: 139-153.
- Stymne, S. (2011). Blast: A tool for error analysis of machine translation output. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations*. Association for Computational Linguistics, pp. 56-61.
- Stymne, S. & Ahrenberg, L. (2012). On the practice of error analysis for machine translation evaluation. *Proceedings of 8th International Conference on Language Resources and Evaluation LREC 2012: 1785–1790*.
- Valotkaite, J. & Asadullah, M. (2012). Error detection for post-editing rule-based machine translation. *Proceedings of the AMTA - WPTP2*. San Diego: USA
- Vasconcellos, M. & Marjorie L. (1988). SPANAM and ENGSPAN: Machine translation at the Pan American Health Organization. In: Jonathan Slocum (ed.) *Machine translation systems*. Cambridge, New York: Cambridge University Press, pp. 187-235.
- Vilar, D., Xu, J., d'Haro, L. F., & Ney, H. (2006). Error analysis of statistical machine translation output. *Proceedings of the 5th international Conference of Language resources and Evaluation (LREC)*. Genoa, pp. 697-702.
- Zheng, J., Mandal, A., Lei, X., Frandsen, M., Ayan, N. F., Vergyri, D., Wang, W., Akbacak, M. & Precoda, K. (2010). Implementing SRIs Pashto speech-to speech translation system on a smart phone. *IEEE Workshop on Spoken Language Technology (SLT 2010)*. Berkeley: CA.

Contact

dr. Ramunė Kasperavičienė
Faculty of Social Sciences, Arts and Humanities
Kaunas University of Technology
Mickevičiaus st. 37-1309

Kaunas, LT-44244, Lithuania
ramune.kasperaviciene@ktu.lt

dr. Jolita Horbačauskienė
Faculty of Social Sciences, Arts and Humanities
Kaunas University of Technology
Mickevičiaus st. 37-1309
Kaunas, LT-44244, Lithuania
jolita.horbacauskiene@ktu.lt