Machine Translation: Translated Texts in Terms of Standards of Textuality

Audronė Daubarienė, Greta Ziezytė

crossref http://dx.doi.org/10.5755/j01.sal.0.22.4301

Abstract. The paper presents an overview of machine translation (MT), its paradigms, and the differences between two machine translation systems: *Google Translate* and *VDU Translate*. *Google Translate* applies statistical techniques to learn translation models from different translated texts, bilingual or multilingual texts. Statistical machine translation (SMT) is the application of previously used algorithm, adopted from parallel corpora, parallel texts, bilingual texts or multilingual texts, to translate other completely new translation units. *VDU Translate* is a rule-based (RB) bilingual unidirectional MT system implementing linguistic rules in text translation. For the most part a range of morphological, syntactic, and/or semantic analyses are being applied, as well as the structural text transformations.

Two texts (a scientific paper and a newspaper article) translated by *Google Translate* and *VDU Translate* have been compared in terms of 7 standards of textuality to find out their acceptability to the reader. The seven standards of textuality include cohesion, coherence, intentionality, acceptability, informativity, situationality, and intertextuality proposed by R. A. Megrab. The findings of the research demonstrated that the newspaper article and the scientific paper translated by both MT systems comply with standards of text cohesion. However, they do not comply with the standards of coherence, acceptability and informativity due to numerous semantic mistakes, therefore, translations of this quality cannot be informative or acceptable to the reader.

***Key words**: machine translation (MT), statistical machine translation, rule-based machine translation, standards of textuality, text genres.*

Introduction

The roots of translation date back to the 18th century BC, when "Epic Gilgamesh" was translated from Sumerian into Hittite. This is the first recorded written translation made on earth (Pažūsis and Armalytė, 1990, p. 7). Nowadays translations have even more significance due to continuously proceeding globalization processes which enlarge the communicative needs among speakers of different languages, because of political considerations as they require document and other text translation, because of widely used World Wide Web, and stronger integration of the European Union. The number of translations has increased, and translators are not able to impart all necessary texts into the target language. Thus, a strong linguistic shift is necessary to achieve a dynamic change. Relatively slow and rather expensive human translation, the invention and further developments of computer technologies and software, and scientific knowledge capacitated the invention of machine translation in the 1950s. Great efforts and grants led to constant improvements of different systems of machine translation, such as "BABEL FISH" or "PROMT". This article focuses on *Google Translate* and *VDU Translate* translation tools.

The aim of this article is to analyse 2 texts of different genres (an article and a scientific paper) translated using machine translation systems *Google Translate* and *VDU Translate* in terms of seven standards of textuality.

The **object** of this paper is the products of machine translation systems *Google Translate* and *VDU Translate*: translated texts of different genres.

Machine Translation Paradigms: *Google Translate* – statistical MT and *VDU Translate* – rule-based MT

Statistical MT (SMT) requires little human interference when monolingual or bilingual corpora are available, meanwhile, rule-based MT (RBMT)

> *"needs an explicit representation of linguistic information whose coding by human experts requires a considerable amount of time" (Sanchez-Cartagena et al., 2011, pp. 90–96).*

Google applies statistical techniques to learn translation models from different translated texts, bilingual or multilingual texts. The search engine *Google* was created in 1997 by Sergey Brin and Larry Page. Soon *Google* operated at the epicentre of the internet with qualified IT professionals what led *Google* not only to be a search system, but to extend its preserve (Vise, 2005, p. 6, 8).

One of *Google* services is *Google Translate*: a machine translation system. It is a cost-free MT system supporting 64 languages (multilingual translation system designed to translate among more than two languages (Hutchins and Somers, 1992, p. 70)) capacitating to translate texts, documents, or web pages. *Google's* mission is to make information accessible across language boundaries, whose approach to MT is information-driven and empirical.

SMT is the application of previously used algorithm, adopted from parallel corpora, parallel texts, bilingual texts or multilingual texts, to translate other completely new translation units. SMT enables a MT system for a new language pair to be built within a short period of time (Lopez, 2008, p. 5). Finally, SMT can be said to convey a general content quite precisely, though the translation itself can be not very logical. Therefore, the main problem of such translation is the scale of parallel corpora, due to

which lexis and grammar are limited. Though, if source and target languages have similar grammar, better SMT results are to be expected (Daudaravičius, 2006, pp. 13–15).

RBMT implement linguistic rules in text translation. For the most part a range of morphological, syntactic, structural text transformations, and/or semantic analyses are being applied (Carl and Way, 2003, p. 8).

VDU Translate is a RBMT system designed for translating texts of different formats from English into Lithuanian. It is a bilingual unidirectional system translating from one language into another in one direction (Hutchins and Somers, 1992, p. 70). *VDU Translate* was created by the researchers of the Centre of Computational Linguistics at Vytautas Magnus University (Lithuania).

Text and Standards of Textuality

According to Neubert and Shreve (1992, p. 69), a text

> *"should be linguistic enough to allow empirical approaches to the underlying cognitive system, but it should not be restricted to a registration of the formal devices of the two languages".*

"A text is the central defining issue in translation" because texts and their situations describe translation as a process, so it is not possible to determine translation as a process without mentioning peculiar texts (Neubert and Shreve, 1992, p. 5). J. McGann (2001, p. 138) maintains that a text "should be understood as a document composed of both semantical and graphical signifying parts".

On the surface, a text is a corpus of words and sentences, but it has other totality of properties which enables a text to be perceived as a text, which can be used as a means of communication (Wright and Budin, 2001, p. 774). Therefore, the meaning of a text is partially a meaning of a discourse, which rendered the origination to the specific texts. Texts are determined by the type of a discourse and particular limitations of a genre. Discourse and genre determine text linguistic expression (Marcinkevičienė, 2008, p. 15), therefore, they become concurrent. Thus, modern text linguistics practises on the principle of textuality. Textuality involves translating and word knowing with the text as a product. Textuality is a compendium of features that each text must have to be reputed as text (Neubert and Shreve, 1992, p. 69). Moreover, Neubert and Shreve (1992, p. 70) claim, that the principles of textuality can determine the conditions under which source language text and its translation can be said to be textually equivalent. These principles are described by a number of linguists, such as R. A. Megrab (1997, p. 232), S. E. Wright and G. Budin (2001, p. 774), T. Bex (2001, p. 74). All the authors suggest seven standards of textuality: **cohesion**, **coherence**, **intentionality**, **acceptability**, **informativity**, **situationality**, and **intertextuality** (Megram, 1997, pp. 232–236).

Methodology

Seven standards of textuality suggested by different linguists such as R. A. Megrab (1997, p. 232), S. E. Wright and G. Budin (2001, p. 774), T. Bex (2001, p. 74) are necessary that a text could be reputed as such, and two different text genres were introduced and described. The aim of the analysis is to ascertain if the product texts (texts translated with a translation system) of *Google Translate* and *VDU Translate* meet the requirements of all seven standards of textuality. Even though a text does not comply with one of the standards, the analysis is continued in order to identify which standards of textuality are most common and which are not, and to find out their frequency throughout all the texts. They are analysed because machine translation is a part of linguistics, and it is a rather new phenomenon. The analysis is based on descriptive – analytical and comparative methods. Two texts of different genres have been chosen: an excerpt of an article and an abstract of a scientific paper. The genres and the texts have been chosen intentionally. The first text is an excerpt of an article entitled "Libya: Rebels battle for road to Gaddafi hometown Sirte" is taken from online – "BBC News", the length of which is 21 sentence. This article is about the war and rebels in Libya. The second text is an abstract of a scientific paper called "Machine Translation of Noun Phrases from Arabic to English Using Transfer-Based Approach", it contains 17 sentences.

The steps in this part of the paper are (1) to translate the selected texts using machine translation tools *Google Translate* and *VDU Translate*, (2) to analyse the texts according to seven standards of textuality, and (3) to point out specific linguistic components that are most challenging in using MT from English into Lithuanian.

First, the analysis of article translated both with *Google Translate* and *VDU Translate* will be presented. Next, the analysis of translated scientific paper will follow.

Analysis of Article Translation with *Google Translate*

The translation of a twenty-one sentence long article about Libya taken from BBC internet webpage was analysed. First, it was analysed in terms of referent, conjunctive and semantic cohesion. Next, the analysis focused on coherence, intentionality, acceptability and informativity, and finally, situationality and intertextuality of the translated text were discussed.

Referent **cohesion** is when a word or a phrase cohere to another and make sense in that meaning, therefore, it occurs in seven sentences as pronouns, e.g. , *savo*, *ji*, *jie*, *jų*, *joms*, *tai*, *mūsų*, are used, e.g. [...] *sukilėliai pasakė, kad jie buvo užgrobę Sirtę* [...].

Conjunctive cohesion is in eleven sentences because subordinating and coordinating conjunctions, such as *bet*, *kuri*, *kol*, *ir*, *kad*, *kuris*, and junction *ankstesni* are used to make sentences and parts of it cohesive to each other, e.g. [...] *juda sparčiai į vakarus, bet atėjo pagal* [...].

Lexical cohesion including repetition and synonymy occurs in fifteen sentences as the keywords, such as *Libija* and its cases, Libyan towns, *NATO*, *Gaddafi*, *sukilėliai*, are repeated throughout the text. Two phrases are headings, and in this case they are not considered as sentences.

It is impossible to achieve **coherence** in translation if some of the words are not translated, thus, seven sentences are incoherent because a keyword *rebels*, abbreviations *Col*,

Lt, and *Gen*, and names of Libyan cities are not translated, e.g. *Uqayla, Bin Jawad, Nawfaliyah*. Thus, it is seen that the MT system cannot recognise English abbreviations of military ranks because if to consult a dictionary (see Cambridge Advanced Learner's Dictionary, Macmillan English Dictionary), these abbreviations most often are spelled with a full stop: *Col., Lt., Gen*. The reason is frequency of use. What is interesting to note about the translated and not translated names of Libyan cities, the MT tool translates *Tripoli, Benghazi*, because they are globally known names in the international arena and electronic media, while locally known names of *Uqayla, Bin Jawad, Nawfaliyah* haven't been spread before the current events in Libya.

Twelve sentences lack coherence because they contain words which are translated literary, though they do not correspond to the context, e.g. [...] *sukilėliai pabėgo atgal į šiukšlinę Jawad* [the rebels fleeting back to Bin Jawad]. These examples demonstrate that polysemy in SL is another big barrier for MT tool that is still difficult to overcome. In case of *Bin Jawad*, the MT tool does not distinguish between "bin" and a proper name *Bin Jawad*.

Two sentences are incoherent because they contain words which do not exist, e.g. *pasalas, kritikos*. 1 sentence is incoherent either because a phrase *kartotinių pasala* has a different meaning which in this case mismatches the context. In conclusion, only 3 sentences are coherent, thus, this text does not comply with coherence.

The title "*Libija: rebels kova už kelių į Kadhafi gimtajame Sirte*" ["Libya: Rebels battle for road to Gaddafi hometown Sirte"] of an article outlines the main idea of the text, that is the fight of Libyan rebels. The aim of the article is to inform the readers, so **intentionality** is achieved by introducing to a situation which is a war in Libya.

Acceptability is closely related to intentionality. Even though the sender's intention is revealed in the text, it does not fulfil acceptability, because in order to be acceptable the text has to be determined by specific textual features, that is grammar and lexis. There are several words that are not translated in six sentences, e.g. *rebels, Col, Bin Jawad, Gen*. Also, there are some words in eleven sentences which are translated literary and contravene the context, e.g. *Didžioji Britanija ir Prancūzija primygtinai ragino Pulkininkas Gaddafio šalininkų defektu", kol ji dar ne per vėlu"* [Britain and France have urged Col Gaddafi's supporters to defect "before it is too late"].

The *Google Translate* product contains a lot of grammatical and lexical mistakes, such as wrong punctuation, e.g. [...] *", kol ji dar ne per vėlu"*; *Anti-Gaddafi*, inappropriate use of noun and adjective cases, gender, and numbers, e.g. *NATO, kuri dabar veikia koalicijos veiksmų* [...]; the words which are not translated at all, e.g. *rebels.*

One of the differences between Lithuanian and English is tenses. The latter has simple, perfect, continuous tenses, and their composition is sometimes prolific of auxiliary words, meanwhile, Lithuanian almost does not contain such. So, MT system does not recognise some tenses, e.g. Present Perfect Continuous tense form *has been moving* is translated by splitting the auxiliary verb *has* been and the verb *moving* and translating the separately: *buvo juda*.

The first order **informativity** in the product of *Google Translate* is presented by functional words *bet* and *kad*. The original text almost does not contain functional words; meanwhile 2 product sentences possess these words although the original text does not, e.g. […] *sukilėliai pasakė, kad jie buvo užgrobę* [rebels said they had seized […]. However, the text does not comply with two other orders of informativity due to many grammatical and lexical mistakes and not translated words occurring in each sentence.

This article is written during afloat of a war in Libya, when everybody is interested in the situation happening there, so it complies with **situationality**, one of the standards of textuality.

Intertextuality deals with reference to other texts. It is an article and it has a title *Libija: rebels kova už kelių į Kadhafi gimtajame Sirte*, a heading *Libijos sukilėliai kovoja kontrolės kelio, vedančio į vyriausybės lojalistų širdyje*, and sub-headings *Pakartotinas pasalas* and *Rusijos kritikos*. It is intertextual to other texts because this issue is not a novelty neither to the writer, nor to readers. Thus, it complies with this standard of textuality.

To sum up, the excerpt of the article translated with *Google Translate* does not comply with coherence, acceptability, and informativity.

Analysis of Article Translation with *VDU Translate*

Six sentences in this text are cohesive as referent pronouns, such as *jo, tai, jie, jų, mūsų,* are used, e.g. *Nato, kuris dabar valdo koalicijos veiksmą* [...].

Conjunctive **cohesion** is identified in twelve sentences as junction *anksčiau* and subordinate and coordinate clauses are linked using conjunctions, e.g. *bet, kuris, ir, kad*, e.g. [...] *kariuomenė judėjo greitai į vakarus, bet patyrė* [...].

Lexical **cohesion** is in sixteen sentences because such words as *Libija*, its towns, *NATO, Gaddafi, sukilėliai*, are repeated throughout the text. Also, there are some synonyms used in the article, e.g., NATO in one sentence is replaced by a word *sąjunga* [alliance], and a word rebel becomes *kovotojas* [fighter]. Two phrases are headings, and in this case they are not considered as sentences, thus a product of *VDU Translate* is 100 % cohesive.

A lack of **coherence** occurs in five sentences because they contain not translated words, e.g. Libyan cities: *Ras Lanuf, Uqayla*; *Lt*, thus, the meaning of the sentences cannot be understood. As well as *Google Translate* this system does not recognise abbreviation *Lt* as it is written without a full stop, whereas, dictionaries suggest *Lt.* The recent news about Libya give stimulus to present more Libyan cities not only the capital. Still, some of them are not presented in the dictionaries or on the internet, so the MT system does not find any equivalents. However, the text contains fewer not translated words in comparison with *Google Translate*, as the latter has not translated *rebels* or *Col*.

It is impossible to perceive the meaning of seven sentences because of the unduly translated words, e.g. [...] *turi aprūpinti sukilėlio pažangos dangą* [are to provide cover for a rebel advance], thus the sentences are incoherent. The incorrect word translation most often occurs because of word polysemy, thus, the MT system is not capable of finding the best equivalent. In case of *zona be musių VDU Translate* does not distinguish between a small insect and going by plane.

In comparison to *Google Translate*, the product of *VDU Translate* contains more coherent sentences, altogether nine sentences, thus, 43 % of the translated sentences are coherent.

After reading the title of the article translated by *VDU Translate* it is obvious what is the main idea of the text, and further reading introduces the author's intentions: to inform readers about the situation in Libya, to announce news about the military situation, thus, the text complies with the standard called **intentionality**.

This text does not fulfil **acceptability** as its main features are compliance with grammatical and lexical norms since there are a lot of mistakes in the translation. Altogether, there are only three correct sentences which have no grammatical and lexical mistakes.

All the other sentences contain various lexical mistakes, such as incorrect meaning of the translated words, e.g. [...] *atsakinga už reikalavimą zonos be musių* [...] [responsible for enforcing the no-fly zone]. Also, there are four sentences in which the names of Libyan cities are not translated, e.g. *Ras Lanuf, Uqayla, Bin Jawad, Nawfaliyah*. Moreover, a lot of grammar mistakes occur across the article, such as incorrectly written inverted commas (not localized into Lithuanian), e.g. [...] *dezertyruoti "anksčiau, negu tai bus per vėlai"*; an incorrectly used hyphen, e.g. *Anti-Gaddafi*; irregular noun and adjective cases and numbers, e.g. [...] *sukilėlio kariuomenė* [...] [The rebel army].

The first order **informativity** in the translation is presented by different functional words, such as *bet, kad, negu* in ten sentences. Still, the text does not comply with the second and third order informativity due to lexical and grammatical mistakes occurring in 86 % of the sentences, as only three sentences are correct, and because of not translated words in sentences.

In comparison to the translation of *Google Translate*, this text also meets the standard of textuality called **situationality** as the text is written when all the news broadcasts reported the latest news about the situation in Libya.

As this text is an article it has its title, its incident heading, and sub-headings, e.g. *Pakartotinė pasala* and *Rusų kritika*. This article is related to other texts as the topic is not new, thus, it complies with **intertextuality**.

Analysis of Scientific Paper Translation with *Google Translate*

Reference is one of the major ways to achieve **cohesion** in the text. In the analysed scientific paper there are eight sentences where reference is used. Referent cohesion occurs between two sentences where a pronoun *tai* defines the meaning of the previously written sentence. Three sentences have pronouns *jie, ji, jų*, which give reference to the precedent sentences, e.g. *Jie taip pat tarnauja kaip taisyklės* [...].

Elliptical cohesion is achieved in four sentences because in these sentences the word *kalba* is anticipated but not written, e.g. *Bet iš arabų į anglų kompiuterinis vertimas* [...].

Subordinate clauses, expressed by subordinate conjunctions: *kurios, kadangi, kad, kuri, kuris*, in six sentences make the latter cohesive. Coordinating conjunctions, such as *bet, tai, ir, arba*, are used in five sentences and render them cohesive, e.g. [...] *mašininio vertimo metodų ir šiuo metu yra* [...].

Conjunctive cohesion is ascertained in three sentences as conjunctions *taip pat, kaip antai, iš tiesų* are used. As the subject of this scientific paper is Machine Translation, therefore, the keyword *vertimas* or *versti* and their different cases and persons or their English acronym *MT* are repeated in nine sentences, e.g. *Tai kelia didelių sunkumų MT dėl daugybė būdų* [...]. Another repetition occurs in six sentences as a word *kalba* recurs. Five sentences are synonymously cohesive to one sentence, as synonyms *kompiuterinis vertimas*, *mašininis vertimas* and *sistema* of a term *automatinis vertimas* are used, e.g. [...] *yra sukurti mašininio vertimo, kad verčia iš* [...].

"Abstract" is a heading, so it is not considered as a sentence. Thus, 100 % of all the sentences are cohesive; in consequence, the product of *Google Translate* fulfils cohesion.

Five sentences lack **coherence** as not all the words are translated, e.g. *MT, rulebased*. It is apparent that machine translation system does not recognise a compound word *rulebased*. This occurs because a suffix *–based* is usually written hyphened. It is interesting to note that *Google Translate* system does not translate an acronym MT. Even though dictionaries suggest translation for this acronym, machine translation tool does not translate its acronym.

Thirteen sentences are inconceivable due to the words, the meanings of which are wrongly translated and contravene the context; therefore, these sentences are incoherent, e.g. *Perkėlimas pagrįstas automatinio vertimo yra viena instancija rulebased mašininio vertimo metodų* [...] [Transfer-based machine translation is one instance of rulebased machine translation approaches]. One sentence lacks coherence as well as the previously mentioned sentences because it does not contain any logical structure necessary for a sentence to be understood. As it is seen, one more barrier for MT into Lithuanian is long sentences of complex structure containing subordinate clauses, several subjects, and attributes.

Only two sentences are coherent throughout the text, thus, 12 % of all the sentences are coherent; consequently, this text does not comply with coherence.

Intentionality in this abstract of the scientific text translated with *Google Translate* is achieved by introducing new

information about machine translation, presenting its challenges, problems, and occurring mistakes, and introduces a new machine translation system. Thus, the text complies with this standard of textuality.

The standard of **acceptability** is met, when the receiver has realized author's aims, and specific textual features (grammar and lexis) have to be determined. There is only one sentence, constituted of one word, which is lexically and grammatically correct. All the other sixteen sentences contain both, lexical and grammatical mistakes.

The most frequent mistakes are the following: mistakenly used capital letters, e.g. [...] *iš arabų į anglų Naudojant* [...]; improper noun and adjective cases, gender, and numbers, e.g. [...] *prisiderinti prie žodžių tvarka* [...], incorrectly translated tenses, e.g. [...] *sistema turėtų būti sugebėti* [...], wrong punctuation, e.g. [...] *turėtų būti kruopščiai, taikomos siekiant užtikrinti* [...], the wrong meaning of the translated words, e.g. *Tai kelia didelių sunkumų MT dėl daugybė būdų išreikšti tą pačią bausmę arabų* [This poses a significant challenge to MT due to the vast number of ways to express the same sentence in Arabic.], and the words which are not translated at all, e.g. *MT*; *rule-based*. The paper does not meet the standard of textuality: acceptability, because of numerous mistakes.

The first order **informativity** in the product of *Google Translate* is presented by functional words, such as *kadangi*, *kad*. However, the text does not meet with second and third order informativity because of a great variety of lexical and grammatical mistakes in 94 % of the sentences. So, the abstract of the scientific paper does not comply with this standard of textuality.

Machine translation is quite a new phenomenon not only among linguists and IT specialists, but also among the people who as a result of different reasons are somehow related to translating and translation. That is why the paper is relevant to nowadays world, and it complies with the standard of textuality: **situationality**.

This text fulfils **intertextuality** well as it contains qualities appropriate for a scientific text, such as terminology, objective information, and impersonal sentences. Also this abstract includes constituents intrinsic to a scientific paper, in this case (as only the abstract is being analysed), a title, an abstract, and an approach.

To sum up, the translated abstract of the scientific paper does not meet coherence, acceptability, and informativity.

Analysis of Scientific Paper Translation with *VDU Translate*

Cohesion in this text is expressed by reference in two sentences using a pronoun *jie*, e.g. *Jie taip pat tarnauja kaip taisyklės* [...]. Pronoun *jie* is used instead of *užsakymo ypatybės*, thus, making two sentences cohesive. Also one sentence contains cohesion because a demonstrative is used. The item *jie* is used in the place of a phrase *žodžio užsakymo taisyklės*.

Elliptical cohesion in two sentences is used as a word *kalba* is anticipated, e.g. [...] *nuo arabų kalbos į anglų*. Conjunctive cohesion is identified in twelve sentences as conjunctions, such as *kuri*, *ir*, *kad*, *kadangi*, *tai*, *ar*, *kuria*, *taip pat*, *kurios*, *kuris*, *iš tikrųjų*, *tokius kaip* are used, e.g. [...] *ryšį su žodžio užsakymu, kurį arabų kalba rodo*.

A term *mašininis vertimas*, verbal form of *vertimas* or an English acronym *MT* are repeated in thirteen sentences and the latter are lexically cohesive. One more cohesive repetition of a keyword *kalba* is in ten sentences. One sentence is synonymously cohesive to three because a word *sistema* is used instead of a term *mašininis vertimas*. To sum up, 88 % of all the sentences are cohesive, and 12 % are non-cohesive, thus, the text fulfils cohesion.

MT is an acronym and *rule-based* is a key word constantly repeated throughout the text, but they are not translated into the target language, thus, five sentences are **incoherent**.

It is impossible to perceive meanings of three sentences as the meanings of some of the translated words contravene the context, thus, the sentences are incoherent, e.g. *Žodžio užsakymo taisyklės yra labai svarbios kartai sakinių kalboje į kurią verčiama* [The word order rules are crucial for the generation of sentences in the target language]. One sentence is incoherent either because the ideas are stated disconnectedly; therefore, it is impossible to understand the meaning, e.g. *Naudojant pranašumus perkėlimu pagrįsto mašininio vertimo tokius kaip analizės žingsnis* […].

Quite reasonably conveyed ideas occur in three sentences, consequently, these sentences are coherent. To sum up, only 17 % of the sentences throughout are coherent, thus, the text does not comply with coherence because generally the text is formed out of incomprehensively expressed ideas. As compared to the product of *Google Translate* this text contains twice as many coherent sentences.

Usually scientific papers provide information concerned with the theme, so this paper introduces information about machine translation, so **intentionality** is achieved in both, *Google Translate* and *VDU Translate* products.

Acceptability in the text is confirmed if the sender's aims are perceived and specific textual features are determined. The author's aim is realized, but there are plenty of mistakes found in the text. The most frequent mistakes are: incorrect noun cases and numbers, e.g. [...] *Mašininį vertimą* [...] *sistema* [...], irregular verb tenses, e.g. [...] *plačiausiai panaudotų* [...] [is currently one of the most widely used]; irregular use of capital letters, e.g. [...] *vertimas Daiktavardinių grupių* [...], wrong punctuation, e.g. [...] *sakinio kalboje į kurią verčiama*, wrong word translations, e.g. *Žodžio užsakymo problema* tampa [...] [The word order problem becomes], not translated words, e.g. *MT*.

This translated abstract of a scientific paper does not meet **informativity,** due to different kinds of mistakes, starting with lexical and in conclusion to grammatical ones occurring in all the seventeen sentences of the text and making 100 % of sentences incorrect.

This translation complies with **situationality** as well as the product of *Google Translate* because machine translation is a new linguistic and IT phenomenon.

The abstract of the scientific paper translated with *VDU Translate* in comparison with *Google Translate* meets **intertextuality** too, as it interacts with the same genre texts (terminology, scientific discourse, objective information evaluation, constituents). It also requires some previous sender and receiver's knowledge about the topic.

Conclusions

The article and the scientific paper translated with *Google Translate* and *VDU Translate* comply with cohesion because sentences are externally related, comply with intentionality because the author's intentions are revealed in the translations, they fulfil situationality and intertextuality as they are relevant to a situation and interface with other texts.

However, translations do not comply with coherence, acceptability, and informativity standards as a result of numerous mistakes in the translations and they cannot be acceptable and informative to a reader.

The article translated with *Google Translate* contains 14 % of the coherent sentences; meanwhile *VDU Translate* product contains 43 % of coherent sentences.

Nonetheless, the scientific paper translated with *Google Translate* and *VDU Translate* does not fulfil coherence because the product of *Google Translate* contains 12 % of the coherent sentences, and *VDU Translate* – 24 %.

None of the translated texts complied with all the seven standards of textuality. All the translations of the article and the scientific paper contain a lot of mistakes, furthermore, they are incoherent, a lot of words are not translated, as a result, a lot of post-editing is needed.

VDU Translate is developed by Lithuanians, thus it has been expected to be more precise when translating into the Lithuanian language in comparison to *Google Translate*.

The differences between synthetic (Lithuanian) and analytical (English) languages lead to a conclusion that the products of machine translation have incorrect word inflexions, cases, and genders, and wrongly translated tenses.

One of the most common mistakes in the target text is incorrect word meaning. It occurs because of polysemy in the source and the target languages. This is a big obstacle for machine translation to overcome, because MT systems do not distinguish different meanings of the same word in different contexts.

References

1. Bex, T., 2001. *Variety in Written English. Texts in Society: Societies in Text.* London: Routledge.
2. Carl, M., Way, A., 2003. *Recent Advantages in Example-based Machine Translation* [online]. The Netherlands: Kluwer Academic publishers. Available at: http://books.google.com [accessed March 2013].
3. Daudaravičius, V., 2006. Pradžia į begalybę. *Darbai ir dienos. Vertimų pasaulyje*, No. 45. Kaunas: Vytauto Didžiojo universitetas, pp. 7–18.
4. Hutchins, W. J., Somers, H. L., 1992. *An Introduction to Machine Translation* [online]. London: Academic Press. Available at: http://www.hutchinsweb.me.uk/IntroMT-TOC.htm [accessed February 2013].
5. Lopez, A., 2008. *Machine Translation of Noun Phrases from Arabic to English Using Transfer-Based Approach* [online]. New York: ACM Computing Surveys. Accession no. 10.1145/1380584.1380586. Available at: http://portal.acm.org/citation.cfm?doid=1380584.1380586&preflayout=flat#source [accessed March 2013].
6. Marcinkevičienė, R., 2008. *Žanro ribos ir paribiai. Spaudos patirtys.* Vilnius: Versus aureus.
7. McGann, J., 2001. *Radiant Textuality: Literature after the World Wide Web* [online]. New York: Palgrave Macmillan. Available at: http://books.google.com [accessed December 2012].
8. Neubert, A., Shreve, G., 1992. *Translation as Text* [online]. Kent: Kent State University Press. Available at: http://books.google.com [accessed February 2013].
9. Pažūsis, L., Armalytė, O., 1990. *Vertimo teorijos pradmenys.* Vilnius: Vilniaus universitetas.
10. Sanchez-Cartagena, V. M., Sanchez-Martınez, F., Perez-Ortiz, J. A., 2011. *Enriching a Statistical Machine Translation System Trained on Small Parallel Corpora with Rule-Based Bilingual Phrases* [online]. Spain: University of Alicante. Available at: http://www.aclweb.org/anthology-new/R/R11/R11-1013.pdf [accessed February 2013].
11. Megrab, R. A., 1997. Standards of Textuality and the Translation of the Hadith. *Translating Sensitive Texts: Linguistic Aspects* [online]. Amsterdam: Rodopi. Available at: http://books.google.com [accessed January 2013].
12. Vise, D. A., 2005. *The Google Story*. New York: Delacorte Press.
13. Wright, S. E., Budin, G., 2001. *Handbook of Terminology Management* [online]. Amsterdam: John Benjamins Publishing Co. Available at: http://books.google.com [accessed February 2013].

List of Sources

1. Libya: Rebels Battle for Road to Gaddafi Hometown Sirte. *BBC News Online.* Available at: http://www.bbc.co.uk/news/world-africa-12885395 [accessed March 2011].
2. Shirko, O. et al., 2010. Machine Translation of Noun Phrases from Arabic to English Using Transfer-Based Approach [online]. *Science Publications.* Available at: http://www.scipub.org/fulltext/jcs/jcs63350-356.pdf [accessed March 2011].

Translation Tools

1. *Google Translate.* Available at: http://translate.google.lt/?hl=lt&tab=wT# [accessed March 2013].
2. *VDU Translate.* Available at: http://vertimas.vdu.lt/twsas/ [accessed March 2013].

Audronė Daubarienė, Greta Ziezytė

Mašininis vertimas: išverstų tekstų atitikimas tekstualumo standartams

Santrauka

Straipsnyje pateikiama mašininio vertimo (MV) apžvalga, jo paradigmos bei dviejų MV sistemų, „Google Translate“ ir „VDU Translate“, skirtumai. „Google Translate“ taikomi statistiniai metodai, siekiant įsisavinti vertimo modelius iš skirtingų verstinių dvikalbių ar daugiakalbių tekstų. „VDU Translate“ yra taisyklėmis pagrįsta vertimo technologija, tekstui versti taikomos lingvistinės taisyklės. Straipsnyje yra pateikta dviejų skirtingų žanrų tekstų (mokslinio ir publicistinio straipsnių), išverstų su minėtomis mašininio vertimo sistemomis, analizė. Darbe taikyti morfologinis, sintaksinis, semantinis bei struktūrinių teksto transformacijų tyrimo metodai. Tekstų vertimai yra analizuojami bei lyginami pagal septynis tekstualumo standartus. Jie apima R. A. Megrab bei kitų autorių pateiktus tekstualumo standartus: kohezijos, koherentiškumo, adresanto ketinimų, priimtinumo, informatyvumo,

relevantiškumo bei intertekstualumo. Analizės metu nustatyta, kad publicistinis ir mokslinis straipsniai, išversti su abiem vertimo sistemomis, atitinka teksto kohezijos standartus, tačiau neatitinka koherencijos, priimtinumo ir informatyvumo standartų dėl daugybės semantinių klaidų, todėl taip išversti straipsniai negali būti priimtini ir informatyvūs skaitytojui.

Straipsnis įteiktas 2013 03
Parengtas spaudai 2013 06

About the authors

Audronė Daubarienė, lecturer, Kaunas University of Technology, Lithuania.
Academic interests: language and technologies, computer-mediated communication and e-learning.
Address: Centre of Foreign Languages, Faculty of Humanities, Kaunas University of Technology, Gedimino Str. 43, LT-44240 Kaunas, Lithuania.
E-mail: audra.daubariene@ktu.lt

Greta Ziezytė, Kaunas University of Technology, Lithuania.
Academic interests: technical translation and localization.
Address: Centre of Foreign Languages, Faculty of Humanities, Kaunas University of Technology, Gedimino Str. 43, LT-44240 Kaunas, Lithuania.
E-mail: greta.ziezyte@ktu.lt