МЕЖКУЛЬТУРНАЯ КОММУНИКАЦИЯ И СОПОСТАВИТЕЛЬНОЕ ИЗУЧЕНИЕ ЯЗЫКОВ

www.volsu.ru

LOST IN MACHINE TRANSLATION: CONTEXTUAL LINGUISTIC UNCERTAINTY

Anton V. Sukhoverkhov

Kuban State Agrarian University, Krasnodar, Russia

Dorothy DeWitt

University of Malaya, Kuala Lumpur, Malaysia

Ioannis I. Manasidi

Kuban State Agrarian University, Krasnodar, Russia

Keiko Nitta

College of Arts, Rikkyo University, Toshima City, Tokyo, Japan

Vladimir Krstić

University of Auckland, Auckland, New Zealand

Abstract. The article considers the issues related to the semantic, grammatical, stylistic and technical difficulties currently present in machine translation and compares its four main approaches: *Rule-based* (*RBMT*), *Corpora-based* (*CBMT*), *Neural* (*NMT*), and *Hybrid* (*HMT*). It also examines some "open systems", which allow the correction or augmentation of content by the users themselves ("crowdsourced translation"). The authors of the article, native speakers presenting different countries (Russia, Greece, Malaysia, Japan and Serbia), tested the translation quality of the most representative phrases from the English, Russian, Greek, Malay and Japanese languages by using different machine translation systems: PROMT (RBMT), Yandex.Translate (HMT) and Google Translate (NMT). The test results presented by the authors show low "comprehension level" of semantic, linguistic and pragmatic *contexts* of translated texts, mistranslations of rare and culture-specific words, unnecessary translation of proper names, as well as a low rate of idiomatic phrase and metaphor recognition. It is argued that the development of machine translation requires *incorporation* of literal, conceptual, and content-and-contextual forms of meaning processing into text translation *expansion* of metaphor corpora and contextological dictionaries, and *implementation* of different types and styles of translation, which take into account gender peculiarities, specific dialects and idiolects of users. The problem of untranslatability ('linguistic relativity') of the concepts, unique to a particular culture, has been reviewed from the perspective of machine translation. It has also been shown, that the translation of booming Internet slang, where national languages merge with English, is almost impossible without human correction.

Key words: machine translation, untranslatability, contextual translation, linguistic relativity, lexical ambiguity, syntactic ambiguity.

МЕЖКУЛЬТУРНАЯ КОММУНИКАЦИЯ

МЕЖКУЛЬТУРНАЯ КОММУНИКАЦИЯ is a header.

МЕЖКУЛЬТУРНАЯ КОММУНИКАЦИЯ

ТРУДНОСТИ МАШИННОГО ПЕРЕВОДА: КОНТЕКСТНАЯ ЯЗЫКОВАЯ НЕОПРЕДЕЛЕННОСТЬ

Антон Владимирович Суховерхов

Кубанский государственный аграрный университет, г. Краснодар, Россия

Дороти де Витт

Малайский университет, г. Куала-Лумпур, Малайзия

Иоаннис Игоревич Манасиди

Кубанский государственный аграрный университет, г. Краснодар, Россия

Кейко Нитта

Колледж искусств, Университет Риккё, Тошима, Токио, Япония

Владимир Крстич

Университет Окленда, Окленд, Новая Зеландия

Аннотация. В статье изучаются актуальные проблемы, связанные с семантическими, грамматическими, стилистическими и техническими трудностями машинного перевода, сравниваются 4 основных метода такого перевода: 1) на основе правил (RBMT); 2) на основе корпусов текстов (CBMT); 3) нейронный (NMT); 4) гибридный (HMT). Описываются некоторые «открытые системы» перевода, которые позволяют самим пользователям исправлять или дополнять содержание перевода («краудсорсинговый», или «коллективный, перевод»). Коллективом авторов статьи, носителями языка разных стран (России, Греции, Малайзии, Японии и Сербии), проведено тестирование качества перевода наиболее показательных фраз на английском, русском, греческом, малайском и японском языках с использованием различных систем машинного перевода: PROMT (RBMT), Яндекс.Переводчик (HMT) и Google Translate (NMT). В результате тестирования выявлен недостаток учета семантического, лингвистического и прагматического контекстов переводимого текста (А. Суховерхов), неверный перевод редкой или лингвоспецифичной лексики (К. Нитта), смысловой перевод имен собственных (И. Манасиди), низкое распознавание идиоматических выражений и метафор (Д. де Витт). Авторами статьи показано, что для совершенствования современных систем машинного перевода требуется объединение буквальной, концептуальной и контентно-контекстной форм обработки смыслов текста, улучшение корпусов метафор и контекстологических словарей (Д. де Витт), разработка различных типов и стилей перевода, включающих специфические диалекты и идиолекты пользователей, а также гендерные особенности языка (К. Нитта). На материале сербского языка В. Крстичем переосмыслена с точки зрения машинного перевода проблема непереводимости («языковой относительности») понятий, уникальных для определенной культуры. И. Манасиди показано, что без участия человека невозможен перевод бурно развивающегося интернет-сленга, характеризующегося смешением национальных языков с английским.

Ключевые слова: машинный перевод, непереводимость, контекстуальный перевод, лингвистическая относительность, лексическая многозначность, синтаксическая многозначность.

The language barrier
and machine translation

Natural languages *per se* are hybrid, dynamic, context-sensitive and eco-logical [Sukhoverkhov, 2014; 2015; Steffensen, Fill, 2014; Sukhoverkhov, Fowler, 2015]. Each has its own syntax, multiple word meanings, idioms, innuendos, intertextualities, ecological and cultural embeddedness that sometimes do and sometimes do not coincide with each other. Although analytic philosophy, the generative-linguistic theory, Russian formalists, French structuralists and others have all contributed to language formalisation in recent years, the ecological, process and system approaches to language nature questioned the possibility and effectiveness of such formalisation. For example, ecolinguistics, the distributed language theory, the dynamic and adaptive systems approaches to language, systemic functional linguistics, and cognitive linguistics show that the same language has potentially an infinite variety of meanings and structures and that, by its nature, it is dynamic, interactive, situated, and ecologically / culturally embedded [De Bot, Lowie, Verspoor, 2007; Fowler, Hodges, 2011; Verspoor, De Bot, Lowie, eds, 2011]. As natural languages being developed, distributed, and situated within various systems of activities cannot be completely formalised, the process of translation *sui generis* is also approximate and constantly developing.

Machine translation programs can effectively produce "verbum pro verbo" translations but the metaphorical, metonymic, and idiomatic expressions are not captured in most cases [Abd Rahman, Md Norwawi, 2013; Yusoff, Jamaludin, Yusoff, 2016]. However, the process of *human* translation is not based on a simple rendering according to denotation *per se*; it requires capturing the concept of a word, phrase (sentence), and the general idea of the whole message (text). The results are even more distinguishable when those languages are seldom used or when they belong to a different language family. For instance, the metaphorical expression "*The wheels are falling apart*" or idiomatic phrase "*let's call it a day*!" cannot be translated literally, because they express a problem of a human relationship or need to rest. Knowledge of the relevant culture is also crucial for correct translations. For instance, Bahasa Malaysia or Malay (the national language of Malaysia) has significant varieties of idioms that have been used as a tool of socialization and have contributed to transferring the values and thoughts of the Malay culture [Muhammad, 2006]. A simple idiom such as "*hitam manis*" in Malay (directly translated as "*black sweet*") would be used to refer to a pretty lady (but never to a boy) with a dark complexion. Context sensitivity is another problem for human and machine translation. For example, the Malay word "*geram*", when used in reference to an adorable child, conveys "love and fondness" such as in the expression "*Geram melihat anak comel ini*", but the same word, when used in another context, can denote anger or disappointment.

Furthermore, in some languages, including Japanese, homophones cause an analogous problem. Numerous homophones in Japanese may be fairly distinguished one another when they are spelled in correct Chinese characters (i.e., ideograms) or pronounced with conventional intonations. For instance, one within the system of Japanese would never be seriously complicated between *kuma* ("bear") and *kuma* ("dark circles under eyes") even without a strict context. However, when either human or machine translators lack sufficient knowledge of Chinese characters, they cannot translate even the simple sentence "*Tsukare te kuma ga deki-ta*", meaning 'I've got dark circles under my eyes due to fatigue' correctly. Indeed, an actual trial of translating the sentence by Google Translate, which uncontrollably detects the Japanese sentence merely in alphabetical syllables, ends up with 'I got tired and made a bear'. Obviously, there are obstacles to even performing an elementary "literal translation" in a general sense. The case demonstrates the multiple layers of both polysemy as well as literariness as a critical issue of translation.

The complexity and multidimensionality of translation presupposes an interaction between the understanding of the general content / context of an utterance and its particular concepts or components. Yu. Marchuk shows that modern machine translation systems cannot even solve the basic task of making the correct choice between variants of polysemantic words in one phrase. For example, nowadays, the above-mentioned Google Translate, which is based on

the latest and the most advanced Neural Machine Translation system, correctly translates the phrase "*technical support system*" to Russian [Marchuk, 2016, p. 29], yet fails to identify the meaning of the Sydney Trains announcement "*Doors closing, please stand clear*" or the sceptical response "*I don't buy it!*".

In contrast to a human translator, a machine does not possess the language mastery and cultural background needed to create a trustworthy translation without having a set of rules explicitly predefined in it. These rules have been seen as the result of linguistic formalisation and are based on both cultural idiosyncratic and universal aspects [Wierzbicka, 1992, p. 26]. In comparison with previous years [Kotov, Marchuk, Nelyubin, 1983; Novozhilova, 2014], we see that translation methods and technologies have been greatly improved and diversified thus effectively diminishing the language barrier between speakers of different languages. However, many problems, issues and technical challenges related to machine translation still remain. In this paper, we revise and test the latest machine translation systems for translation accuracy of idioms, rare words, proper names, phrasal verbs and the general content of phrases [Marchuk, 2016; Nguyen, Chiang, 2017] and for the ability to keep track of the fast developing and chaotic online communication, the so-called "netspeak" [O'Curran, 2014; Lim, Cosley., Fussell, 2018; Lohar, Afli, Way, 2018].

To reach the purpose, we review existing translation algorithms, comparing outcomes of the most popular machine translation systems (PROMT, Yandex and Google) with, respectively, Rule-based (RBMT), Hybrid (HMT) and Neural (NMT) algorithms. By translating between various languages (English, Greek, Russian, Malay and Japanese), we test the ability of these systems to understand concepts, metaphorical expressions and structure of a sentence and suggest possible linguistic and technical solutions to detected problems. Therefore, another purpose of this article is to identify the unavoidable limitations of machine translation, show how these limits are predetermined by and correlated with the systemic and dynamic nature of languages, and propose some solutions for coping with this linguistic dynamics and fuzziness.

The main approaches to machine translation

Machine Translation, as a subfield of computational linguistics that investigates the use of computer software for translation of text or speech, has four main approaches on its current stage of development: Rule-based (RBMT), Corpora-based (CBMT), Neural (NTM) and Hybrid (HMT). In this chapter, the theoretical and technical premises of these approaches are reviewed. For the comparison of these methods and for the analysis of their effectiveness, we evaluate their translation quality by using popular machine translation systems: PROMT (RBMT), Yandex.Translate (HMT) and Google Translate (NMT). Received results are used for examination of properties and complexions of tested languages that have yet to be handled by these systems.

Rule-based Machine Translation (RBMT). Rule-based Machine Translation is a translation approach which uses dictionaries to determine the corresponding words, syntax and grammar between the target and the source language. After receiving the message, the machine uses the dictionaries to construct an equivalent message in the target language, which it then outputs. Examples of such systems are Apertium, GramTrans and PROMT, while new systems are being elaborated for Uralic languages [Riahovskaya, 2017; Wiechetek, 2008; Johnson et al., 2017b].

Even though Rule-based Machine Translation seems like a neat solution to our problem, it comes with various inflexibilities that can make it unsuitable in a variety of situations. To begin with, in order to create an accurate RBMT system, all grammatical rules from both languages, as well as the relations between them, have to be explicitly defined in a programmatical way, including grammar exceptions. This greatly increases the time, effort and funds needed to construct such a system. In addition, the word dictionaries (lexicons) are hard to manufacture as, on the one hand, the number of total existing words is different in each language and, on the other hand, this number is constantly increasing by leaps and bounds. For instance, the Global Language Monitor shows that the English language has 1,052,010.5 words (on March 2019)

and a new word is being created every 98 minutes, averaging to about 14.7 words per day (http://www.languagemonitor.com/global-english/no-of-words) However, such a word difference may be compensated for, as example, by adapting or borrowing words from another language, without any translation [Koltan, 2017; Cui, 2012].

Therefore, having only a set of strictly defined rules and a list of corresponding words may lead to false and untrustworthy results, especially when idioms or literary texts are involved [Riahovskaya, 2017]. Furthermore, because natural languages are constantly evolving, with new meanings being added quite frequently, keeping the corpora up to date can be just as inefficient as creating them, especially if a grand change in a language system takes place. Take, for example, the transition from Katharevousa to Demotic Greek which took place in the 1980s, putting an end to the diglossia between written text (Katharevousa) and spoken language (Demotic) in favour of the latter. Were a change like that to happen to a modern language in Rule-based Translation, its dictionaries would be instantly rendered obsolete.

Corpora-based Machine Translation (CBMT). Corpora-based Machine Translation, contrary to RBMT, does not strictly depend on defined lexicons and grammar rules but instead bases its acquisition of "language knowledge" (training) on the analysis of parallel corpora between two languages. This way, the task of manually creating and maintaining rules or word correspondences is delegated to an algorithm, solving RBMT's inflexible dictionary problem.

With the help of information and probability theories comes one of the most popular and effective CBMT's methods: Statistical Machine Translation (SMT), which, as its name suggests, translates texts based on probability values between the source and the target language. In its essence lies the fact that every word in the target language is a suitable translation of a word in the source language and has a certain probability of being correct. The word with the highest probability value is then selected and the source word is substituted by it. For metaphors or idioms, SMT systems can use phrases instead of words to deliver results. The probability values can be determined in a number of ways of which we list two: 1) by analysing the provided parallel corpora and calculating probabilities based on word or phrase equivalence between the source and target languages; or 2) by identifying the words that are more likely to appear after other words [Wang et al., 2017; Babhulgaonkar, Bharad, 2017].

The main disadvantage of Corpora-based Machine Translation, however, is its ineffectiveness when presented with text that it was not trained for. If, for example, the parallel corpora were based on distinct terminology (for a specific brand or domain), then it will struggle to translate text that is written in everyday, casual style. Moreover, if casual texts are added in the specialised training set, then some specific translations could be overridden by casual ones, as their probabilities of appearing would be higher. Consequently, it is important to exercise caution when selecting the parallel corpora, depending on the material that is going to be translated.

In recent years, various **Hybrid Approaches** have been actively developed [Costa-Jussa, Fonollosa, 2015]. Some of them combine the statistical method and the rule-based approach and are applied to popular and rare languages alike [Oladosu et al., 2016]. According to such research, this approach competes with base machine translation methods and provides the best translated output in each language. In keeping with translation quality metrics, for this approach, the National Institute of Standards and Technology (NIST) method displayed a score of 0.8963, while the Bilingual Evaluation Understudy (BLEU) algorithm output a score of 0.7923, with a value close to one indicating high similarity of the machine translation to a reference text, usually a human translation [Oladosu et al., 2016, p. 123]. An example of a Hybrid Machine Translation system is Yandex.Translate, which combines the Neural (Russian to English) and Statistical (all languages) methods, using another system for selecting the best result out of the two (CatBoost) (https://www.bbc.com/russian/features-41086998).

Neural machine translation (NMT). As of 2016-2017, Google, Yandex, Omniscien Technologies, SDL and many others have announced the deployment of neural machine translation. Generally, a neural translation system is based on encoder-decoder architecture. The encoder takes in a sentence in the source language and formalizes its semantics, outputting a sequence of numbers that represent its meaning.

This technology, in contrast to other methods of machine translation, does not "memorize" phrase-to-phrase correspondences or rules between languages, but instead tries to encode the semantics of a sentence and saves them for future reference. In order to represent linguistic (sequentially dependent) information, a more complex type of neural network is used: for example, recurrent neural networks, along with their specific architectures that can "remember" the words used in a sentence (LSTM, GRU) [Zaremba, Sutskever, Vinyals, 2014].

Neural networks represent each word and the whole meaning of a sentence through numerical values. These values are then passed through different mathematical functions and get influenced by other coefficients that hold the "language knowledge" of the system, making a prediction of what the translated text should be like. The coefficients are usually represented through $(N \times M)$ – dimensional matrices and are adjusted with the goal of minimizing the system's error value; i.e. how wrong the system was in its translations (e.g. backpropagation algorithm).

In the meantime, Google has developed an approach that allows an NMT system to generalize each language's accumulated semantics. This allows for "zero-shot translation", meaning that the system can translate between language pairs (correspondences) that were not explicitly included in the training set [Johnson et al., 2017a].

Despite its increased accuracy, NMT also has its problems; it is comparatively quite computationally expensive to train and, in translation inference, encounters difficulty with rare words. It can "over-translate" or "under-translate" (overfitting/underfitting data) and may provide wrong results where the meaning of the source sentence is ambiguous [Wu et al., 2016; Wang et al., 2017]. Taking into account these NMT translation flaws, developers and researchers have proposed hybrid models based on the integrity of the statistical and neural machine translation technologies [Wang et al., 2017].

Accuracy testing of main machine translation approaches. In order to examine the different types of machine translation methods, we used a number of phrases from Russian, English, Greek, Malay, and Japanese. The result was that Google (NMT) and Yandex (HMT) translation services showed the highest degree of accuracy, compared to PROMT (RBMT). However, PROMT had several results better than Google and Yandex.

Below are the samples of the most illustrative results of our online machine translation tests (see Tables 1-5).

Table 1

Russian language

	Phrase: *Он на седьмом небе от счастья*		
	Human translation: *He's on cloud **nine**.*		
Language	PROMT, RBMT	Yandex.Translate, Hybrid (NMT)	Google Translate, NMT
To English:	He is in the seventh heaven	He's over the moon.	He is in seventh heaven
To Greek	Αυτός είναι στον έβδομο ουρανό (He is on the seventh sky).	Είναι στον έβδομο ουρανό από ευτυχία (He* is on seventh the sky from happiness). *Pronoun with no gender difference.	Είναι στον έβδομο ουρανό (He* is on the seventh sky). *Pronoun with no gender difference.
To Malay	Not available.	Dia ke Bulan (He has gone to the moon).	Dia berada di langit ketujuh (He is in the seventh Sky).
To Japanese	*Kare wa, mujo no kofuku de imasu* (He is in the supreme happiness).	*Kare wa tsuki no ue da* (He is on the moon).	*Kare wa dai-nana tengoku ni iru* (He is in the seventh heaven).
Comments: 1) Malay translation "*he has gone to the moon*" shows that Yandex renders from Russian to English and only afterwards to Malay. 2) Malaysia also has the tradition of 7 layers of heaven. 3) Hybrid MT and NMT translated Greek in gender-neutral form, while RBMT explicitly used a male pronoun (*Αυτός*), like in the original text. 4) There is no idea of the seventh heaven in Japanese culture either religious or secular. Yet, PROMT presents the descriptive accuracy of the phrase, even though its syntax is somewhat awkward for the choice of preposition *de* instead of *ni*.			

Table 2

English language

▪ **Phrase: This is, straight up, not my cup of tea.**			
▪ Human translation: Это, честно, не в моих интересах (*Honestly, I am not interested in this*).			
Language	PROMT, RBMT:	Yandex.Translate, Hybrid (CBMT):	Google Translate, NMT:
To Russian:	Это, прямо, не моя чашка чая (*This, straight, not my cup of tea*).	Это, прямо вверх, не моя чашка чая (*This is, straight upward, not my cup of tea*).	Это, прямо, а не моя чашка чая (*This, straight, but not my cup of tea*).
To Greek	Αυτό είναι, στρέιτ*, δεν μου φλιτζάνι του τσαγιού (This is, straight*, not to me cup of tea). *Word not translated, written in Greek letters.	Αυτό δεν είναι το τσάι μου (This is not my tea).	Αυτό είναι, κατ 'ευθείαν, δεν είναι το φλιτζάνι τσάι μου (This is, right away, is not my tea cup).
To Malay	Not available.	Ini adalah, yang lurus ke atas, bukan cangkir teh saya (*This is the straight upwards, not my tea cup*).	Ini, lurus, bukan cawan teh saya (*This, straight, is not my tea cup*).
To Japanese	*Kore wa, massugu ni ue e, o-cha no watashi no kappu de wa arima-sen* (This is, straight upwards, not my tea cup).	*Kore wa, massugu de wa naku, o-cha no watashi no kappu desu* (This is, not straight, and my tea cup).	*Kore wa massugu de, watashi no o-cha de wa arima-sen* (This is straight, and not my tea).
Comments: 1) Malay translation of *cangkir* and *cawan* mean the same, meaning a 'cup'. 2) Hybrid MT failed to convey the meaning to Greek, i.e. even if the user knew both English and Greek, it would be impossible to manually translate the Greek output back to English. 3) All the three translations to Japanese fail to both convert the meaning and compose natural phrases with conventional collocations. In particular, *o-cha no watashi no kappu* means literally 'tea's my cup', even though it can be guessed as 'my teacup'. '[M]y tea' in Google Translate can be considered to connote 'my cup of tea' only by omitting the container in accordance with the grammatical convention of Japanese.			

Table 3

Greek language

▪ **Phrase: Περίμενε με ανυπομονησία το μέλλον**			
▪ Human translation: He was looking forward to the future.			
Language	PROMT, RBMT:	Yandex.Translate, Hybrid (CBMT):	Google Translate, NMT:
To Russian:	Он с нетерпением ожидали в будущем (He with *impatience* [they] awaited in the future).	Ждал с нетерпением будущее (He waited with *impatience* for the future).	Он с нетерпением ждал будущего (He with *impatience* waited for the future).
To English	He waited impatiently for the future.	Wait, looking forward to the future.	He was looking forward to the future.
To Malay	Not available.	Tunggu, sabar untuk masa depan (Wait, patience for the future).	Dia menanti masa depan (He waits for the future).
To Japanese	Not available.	*Mirai wo tanoshimi ni shite matte* (Looking forward to the future).	*Watashi-tachi no te no todoku tokoro ni* (Within our reach).
Comments: 1) RBMT provided an inaccurate translation with both invalid grammar and meaning. 2) Hybrid MT provided a Russian gender-neutral translation, but confused the gender-neutral verb with its imperative form when translating to English ('Wait,…'). The other two translation systems used a male pronoun in the Russian translation. 3) The imperative form can still be used '[Then,] look forward to the future', but is secondary and less common for the user's needs. Furthermore, the translation results for this variant are not grammatically correct. 4) Available two translations to Japanese are both incomplete as sentences: Hybrid MT fails to translate the subject/noun, while Google Translate provides the sentence lacking both verb and object besides its mistranslation of *he* as 'we'.			

Table 4

Malay language

	PROMT, RBMT:	Yandex.Translate, Hybrid (CBMT):	Google Translate, NMT:
▪ **Phrase: Timah*, dengan warna kulit kuning langsat, dikenali dengan kejelitaannya.**			
▪ Human translation: Timah, with her fair skin, is renowned for her beauty.			
Language	PROMT, RBMT:	Yandex.Translate, Hybrid (CBMT):	Google Translate, NMT:
To Russian:	Not available	Свинец, цвета кожи желтая кожа светлая, выявленных kejelitaannya (Lead, skin color yellow skin light, identified kejelitaannya).	Жесть с желтым цветом кожи известна своей красотой (The tin, with a yellow skin color, is known for its beauty).
To English	Not available.	Tin, with the color of the skin yellow complexioned, known by the kejelitaannya	The tin, with a yellow skin color, is known for its beauty.
To Greek	Not available.	Μόλυβδος, το χρώμα του δέρματος κίτρινο complexioned, που προσδιορίζονται από kejelitaannya (Lead, the colour of the skin yellow complexioned, that [are] defined from kejelitaannya).	Ο κασσίτερος, με κίτρινο χρώμα δέρματος, είναι γνωστός για την ομορφιά του (The tin*, with yellow skin colour, is known for its beauty) *The periodic element (Sn).
To Japanese	Not available.	*Hifu no iro to suzu,* kejelitaannya *ni yotte shirarete-iru, kaoiro no kiiro* (Skin color and tin, known by kejelitaannya, yellow in the complexion).	*Kiiroi hada-iro no suzu wa, sono utsukushi-sa de shirarete imasu* (Yellow complexion is known for its beauty).

Comments: 1) *Kuning langsat* is used in Malay to refer to fair-skinned or fair complexion.
2) *Timah is a colloquial shortened name for Fatimah, and is translated erroneously as 'tin'.
3) Both Hybrid MT and NMT output an illogical, absurd translation in all target languages, failing to provide the proper context in their result and using source words without any changes. 4) Two available translations to Japanese also fail to recognize the proper noun *Timah* and automatically drop the word from translation, while Hybrid MT leaves *kejelitaannya* as it is.

Table 5

Japanese language

	PROMT, RBMT:	Yandex.Translate, Hybrid (CBMT):	Google Translate, NMT:
Example 5. Japanese language			
▪ **Phrase: あのかふぇのあまざけはおいしいね (Ano kafe no amazake wa oishii-ne.)**			
▪ Human translation: *Amazake* [sweet non-alcohol rice drink] at that cafe is tasty, isn't it?			
Language	PROMT, RBMT:	Yandex.Translate, Hybrid (CBMT):	Google Translate, NMT:
To Russian:	No translation	Вот именно (It is true).	Вкус Акафу восхитителен (The taste of Akafu is delicious)
To Greek	Not available.	Σωστά (Right).	Η γεύση του Akafu είναι υπέροχη (The taste of Akafu is wonderful)
To Malay	Not available.	Itu benar (It is true)	Rasanya Akafu lazat (Akafu tastes delicious)
To English	No translation	That's right.	The taste of Akafu is delicious

Comments: 1) *[K]afe* is a loan word meaning 'cafe'. No system can identify the word, though humans perhaps can guess its sense by analogy with the original word.
2) *-ne* at the tail of the sentence is a unique binding particle used to ask for confirmation rhetorically. The function of the construction is thus analogous with the tag question in English. Again, no system succeeds in translating the structure.
3) *Amazake* is a culturally specific soft drink, produced intermediately in the process of brewing *sake*, Japanese rice wine.

These online translation examples illustrate the high degree of literal translation persistent in machine translation systems. When the machine translation is juxtaposed to human translation, it can be seen that the online translation procedure lacks conceptual meaning even though the semantic and syntactic systems are integrated into it the online translation procedure. Surprisingly, our testing indicated that all translation systems encounter problems with the recognition of both full and shortened names. In Table 4 *Timah*, a shortened name, was translated as 'tin' and Greek name *Ειρήνη* (*Irene* in English) also was not recognized as a name and got translated literally by its meaning ('peace') in another tested phrase "*Σάμπως να είχε δίκιο η Ειρήνη τελικά;*".

In most cases, idiomatic phrases were not detected and were translated literally by the systems (see Tables 1, 2). Also, this pair of examples demonstrates that the accuracy of a translation depends on the syntactical complexity of the source phrase. Whereas the outcomes in

the first example basically maintain minimum readability, those in the second example are broken in terms of the sentence structure. All three systems are obviously weak in translating adverbial syntax; the adverbial clause *straight up* seems to make them particularly confused and this results in poor performance. Likewise, additional information in the fourth example "*with her fair skin*" and simple modifier *today* in Table 5 respectively cause the same type of mistranslation. Colloquialisms are also misrepresented: the phrase "*Geram melihat anak comel ini*" should convey a feeling of affection when seeing the child, and means 'What a cute child!'. However, with Google Translate it loses its meaning in the given context, becoming 'Greedy saw this cute kid'.

The ambiguity problems, which can be easily resolved by a human, largely contribute to wrong output, as is the case with *ανυπομονησία*: the Greek word describing the feelings of impatience and excitement caused by an unknown situation (Table 3). In the same example, the translation systems fail to identify the gender of the subject and even change the verb into an imperative form. Cases regarding Japanese are even more complicated: the third-person singular pronoun is replaced by the first-person plural *we* in one case, and in the other case, the subject is dropped as seen in the participial construction. This instance suggests that ambiguity of the action can result in the misplacement of the verb causing a wrong form as well as a fragmentary phrase never aligned in a complete sentence.

As mentioned above, the Malay language is full of idiomatic expressions which reflect the various cultural aspects of the language. This feature of the language has led to many inconsistencies in machine translations. As has been shown by previous research in an accuracy analysis during the translation of 200 Malay sentences containing proverbs into English, more than half (55.0 %) of them were wrongly translated by Google Translate, and 34.0 % were correct while only 9.6 % were translated accurately into similar idioms [Abd Rahman, 2013]. The challenges encountered during machine translation were mostly rooted in the use of affixes in words and the additional stopwords in phrases during translation (both of which were used to reflect the grammatical structure of the language), as well as the use of different words

with the same meaning [Abd Rahman, 2013]. In the first issue, the example *memilih kasih* can be translated when the affixes are removed: "pilih kasih". Hence, stemming, which is the detection and filtering of the proverb to exclude affixes, needs to be done before translation [Abd Rahman, 2013]. Secondly, proverbs may have stopwords such as in the following phrases: "*sedikit-sedikit lama-lama jadi bukit*" and "*sedikit-sedikit lama-lama akan jadi bukit*". Hence, the removal of superfluous words such as *akan* would enhance accuracy of the translation [Kwee, Tsai, Tang, 2009; Abd Rahman, 2013]. Thirdly, there may be different words used to represent the same proverbial meaning – "*Ada angin, ada pokoknya*" is similar to "*Ada angin, ada pohonnya*" (meaning anything that happen has a cause) [Abd Rahman, 2013]. In the case "*bagai kera mendapat bunga*" for someone who does not appreciate the value of a gift, *beruk* and *monyet* can replace *kera* to mean the same thing.

Contemporary machine translations have yet to solve the above-mentioned problems [Yusoff, Jamaludin, Yusoff, 2016]. As language does not exist in isolation but is part of a society and culture, one would need to be familiar with Malay to be able to translate the rich and colourful cultural contexts of the language [Chan, DeWitt, Chin, 2018]. Nevertheless, there are studies of Semantic-based Translation using N-Grams that could deal with ambiguous sentences by identifying words with multiple (ambiguous) meanings [Yusoff, Jamaludin, Yusoff, 2016].

Therefore, our analyses and results of previous works in this field show that for machine translation to be effective, it first needs to incorporate three levels of meanings processing: the literal, the conceptual, and contextual. Secondly, a corpus of metaphors, idiomatic phrases and proverbs with equivalences from different cultures need to be constructed. Finally, the recognition of proper names and their shortened versions should be improved.

Machine translation and linguistic relativity theory

The act of translating from one language to another, apart from being a fairly complex problem when implemented by machines, can pose difficulties even for a human translator. A good

example of this fact is the book "English As She Is Spoke: the new guide of the conversation in Portuguese and English" by Pedro Carolino [Da Fonseca, Carolino, 2002] which is full of grammatical and stylistic mistakes that are surprisingly similar to the ones made by the machine translation systems.

The linguistic relativity theory and theories similar to it explain some of these translational difficulties. They show that many languages differ in the amount of words they have, some words describe unique feelings, a person's characteristics, professional jargon, and many other realities specific to a culture [Whorf, 1956; Kövecses, 2005; Deutscher, 2010]. The surrounding environment, natural resources, and specific activities of a region also bootstrap vocabularies or slangs that have no equivalents in other languages [Wierzbicka, 1992; Durdureanu, 2011; Sanders, 2014]. Because of this cultural and geographical specificity, many of them can be translated only with the help of a contextual explanation, rather than with a distinct word. A good example is a word *sevdah*, whose root comes from Turkish language, and is commonly used by people living in Bosnia. The standardly offered translations – "melancholy", "lovesickness", "yearning for love" – do not really capture the essence of *sevdah*. A more precise translation would be 'enjoying your state of sorrow as a very special (sorrow-ish) kind of "pleasure"'. Perhaps, the English term "wallowing in your sorrow" would not be a bad way to understand *sevdah* but only because a better concept does not exist. This specific kind of "emotion" simply seems to be "reserved" for people from the Balkans who would find sevdah-kind of happiness in singing songs describing and glorifying the heroic death of their most loved ones. A possible explanation is that the people from the Balkans in the most difficult times in which they had little to look forward to simply evolved to learn how to enjoy in their sorrow. Today, mainly due to the changed historic circumstances, even many young people from the Balkans would struggle to understand word *sevdah* itself and the state of being in sevdah.

Such types of words do not have their counterparts in other languages and this kind of untranslatability very often leads to their borrowing: we adopt words from one language (the donor language) and incorporate them into another language without or with minor modifications. Sometimes this happens on a critical cultural level because borrowings tend to overflood national languages around the globe (e.g. economic and computer terms such as *market*, *poster*, *billboard*, *slogan*, *hashtag*, *etc.*). In France, this has even caused cultural resistance [Styblo, 2007; Caruso, 2012]. However, the question whether simply introducing a foreign word into a language entails introducing the relevant concept in that language remains discussable. Suppose we incorporate word *sevdah* into English in a way in which we can incorporate *poster* into Serbo-Croatian: it is unlikely that the former borrowing will yield the same result as the latter.

Difficulties in translations of unique words also could be overcome by "adaptation" or "free translation" wherein the social or cultural reality (idea) in the source language is replaced with new realities that are closer and more natural to the audience in the target language. Such "domestication" [Lawrence, 1995] of the source text is very artistic and vulnerable to criticism, and for the moment cannot be implemented by machine translation because of its creative complexity. The opposite side in the art of translation – "foreignization" – strives to save the source language and culture, and translate text into the target language with minimal changes using, for instance, comments and explanations about original realities. However, many researchers, as in the case of verbally expressed humour based on wordplay, agree that such methods are sometimes ridiculous because additional comments and explanations destroy the amusement [Low, 2011; Hoffman, 2012]. Humour comprehension requires implicit and explicit knowledge of specific cultural and linguistic realities, and their explanations could be too long or inappropriate for translation. For example, the joke "*A priest, a rabbi, and a nun all walk into a bar, and the bartender says, 'What is this, some kind of joke?'*" requires knowledge about jokes that begin with "*A, B and C walk into a bar...*".

In this regard, we can see that the problem "how to translate" is a unique and disputable task that can be solved differently by different translators and with a variety of methods. This

complexity / relativity of languages, cultures and methods makes current machine translation systems just an auxiliary tool. However, the more successful, socially accepted or standardised examples of human translations we will have, the more data could be borrowed, formalised and used by specialists and technical systems for "neutral translation" [Razlogova, 2017]. Thus, despite variabilities in the rendering of the same text, some *formal invariants* or *typical examples* can be extracted and be practically used in the machine and human translation.

Lost in Internet translations

The on-going technology boom has further created an additional problem in terms of language translation. Internet slangs, acronyms, hidden meanings, letter and number combinations in words and intentional mistakes are, generally, an accepted way of communicating online.

Additionally, foreign languages are heavily influenced by English in this field with many hybrid words being coined as a result of mixing two languages together. For instance, incorporating English in German to create Denglish, or English in Malay for Manglish, or writing Greek characters in English (Greeklish) as a means to not constantly switch between keyboard layouts. In this case, translations need to consider processing the literal, the conceptual, and the content/context meanings, referring to contextological dictionaries for styles and idiolects of the users.

Combined with the many abbreviations used online, it can be hard for speakers who are learning the language to understand these foreign compressed messages even if they might understand colloquial speech. For instance, the numbers *55* sounds as *'go go'* in Japanese and are used to convey the English meaning. In Malaysia, *fuyoh* is commonly used online and may be equivalent to the *OMG* in the internet slang. An example in Manglish: "*Fuyoh, so cheap!*"

The difficulty and novelty of this language style or idiolect is high enough that a reader outside a given network community and culture would be frustrated to understand anything. Even YouTube comments are becoming more and more outlandish for "strangers". These are the reasons why the origins of "netspeak" and "digital natives" – competent communicators in cyber contexts – are being postulated by researchers in modern culture or in so-called "generation Z" [Crystal, 2001; Pasfield-Neofitou, 2012; Sharifian, 2017, p. 108].

Therefore, "netspeak" is a refined example of an untranslatable language system via machine translation processes because the online translation tool requires to be trained to recognise style, deliberate typos, abbreviations and acronyms. In addition, the community-specific vocabulary predominantly depends on the contexts within which it is utilised. This social reality in interlocutor exchange situations directly affects the translation programs. It is essential to have machine training based on corpora that include words, concept, and content. However, such a problem was partly solved by emergence of a new branch in translation called "crowdsourced translation". Its two main forms are: 1) non-professional community-based systems such as the Google Translate community that corrects automated translations or Luis von Ahn's "Duolingo" language approach that uses a learning platform where people translate websites as a part of the learning process and 2) crowdsourced translation service platforms such as TM-Town, Gengo, Smartling and others with professional translators providing their services (https://www.morningtrans.com/crowdsourced-translation-does-it-work/).

Because of the international boom of social networks and the need for translating user-generated content with its slang variations and informal language, Facebook and Twitter have also launched crowdsourced translation platforms in order to create multilingual posts. It must be noted, however, that crowdsource methods do not restrict themselves to translation problems only. For instance, project reCAPTCHA uses the input words or selected images to bring old books to the digital realm and to gather data for artificial intelligence, improving the accuracy of maps. This project is realised in expecting users to decipher distorted words or to identify particular pictures online to reach a successful registration. Crowdsourced translation for verification of machine translations may be the solution to accurate translation for languages in which words and phrases are heavily reliant on culture and context, such as in the Malay language. However,

the difficulty may be in getting a sufficiently dedicated and sufficiently informed crowd to contribute to the translations.

Conclusion

In retrospect, looking at the problems discussed in previous works [Kotov Marchuk, Nelyubin, 1983; Novozhilova, 2014; Arestova, 2015; Dulov, Shmeleva, Boronkinova, 2017], we still see that:

1) At the present stage and in the near future, it is impossible to exclude the human editor from machine translation. The latest and most advanced Neural Machine Translation so far needs human corrections even with simple tasks [Marchuk, 2016; Nguyen, Chiang, 2017]. Therefore, linguists and other scholars still have to contribute to the further development of such systems, for example, through the construction of metaphors corpora and contextological dictionaries that could be used for translation (interpretation) of the most difficult literary texts. For the further development of machine translation, it is also crucial to include the various pragmatics aspects of language into the process of computer-based translation. For the moment, "crowdsourced translation" could be a solution of that problem. It may provide resources for the current translation issue of fast-growing "netspeak". An attempt was also made by a recent project called "SenseTrans", a tool that adds contextual information to posts in social media using AI-analytics [Lim, 2018; Lim Cosley, Fussell, 2018]. The idea of mobilizing a crowd of translators of a variety of texts – both amateurs and professionals – into a sort of collective wisdom for trans-linguistic communications seems also to become a computational materialization of the "translation norms" or "the social reality of correctness notions" [Bartsch, 1987, p. xii]. According to the two pioneer contributors to the theory, G. Toury and Th. Herman, each individual's interaction with the multi-layered socio-cultural norms circumscribing her/his verbal operations informs the sense of accuracy and quality of a translation result [Toury, 1995; Hermans, 1996]. Crowdsourcing processes in tandem with platforms of professional translators able to verify them can potentially construct such a norm that helps make translation machine a usable apparatus.

2) We still do not have an integrated typology of various types and styles of translation. Indeed, translations of technical texts differ from translations of newspaper texts or informal online conversations. Among the systems tested in our research, only PROMT, with some limits, can translate with pre-set writing styles but still cannot detect such styles or eliminate stylistic mistakes. However, some grammar checker software (like Grammarly, StyleWriter, WhiteSmoke, etc.) have addressed this problem, revealing, for instance, colloquialisms in documents written in an academic style.

Many studies show that there are cultural and aesthetic differences between men and women in the use of vocabulary, syntax, and communication [Na, 2016; Okamoto, 2013]. In the Japanese language, "onnarashii hanashikata" (feminine ways of speech) and "otokorashii kotobazukai" (masculine ways of speech) is common. This kind of speech reflects specific forms of politeness and cultural norms represented by language. Although there is no masculine or feminine way of speaking in the Malay language, specific words may be used to describe a feminine or masculine trait. "*Hitam manis*" is always used for women but for men, they are "*berkulit gelap*" or dark complexion. In that context, wrong word choice could lead to ridiculous or impolite results in machine translation.

Nowadays, Google Search considers the original location of a search query, as well as the user's previous requests. Probably in the future, machine translation will also become less abstract and universal, and more personalised and situated, by constantly learning from the interests and idiolects of its users.

3) Modern ecolinguistics, the distributed language theory, the dynamic and adaptive systems approaches to language, systemic functional linguistics, and cognitive linguistics have shown that language (or process of "languaging") is dynamic, interactive, situated, ecologically and culturally embedded. All these aspects of language complicate its formalisation and machine translation. However, it may be still possible to find a common, formalizable core ("universal grammar") for a group of languages by means of theoretical (linguistic, mathematical) and machine analysis of languages. Furthermore, dynamic, learning and evolving models (software) can be

designed by programmers and linguists that could adapt the machine translation to the dynamic nature or the evolvability of natural language and to dialects/idiolects of their users.

REFERENCES

Abd Rahman K., Md Norwawi N., 2013. The Challenges of Handling Proverbs in Malay-English Machine Translation. *14th International Conference on Translation 2013*. Penang, University Sains Malaysia, pp. 27-29. URL: www.scribd.com/doc/163669571/Khirulnizam-The-Challenges-of-Automated-Detection-and-Translation-of-Malay-Proverb.

Arestova A.A., 2015. Sravnitelnyy analiz sistem mashinnogo perevoda [Comparative Analysis of Machine Translation Systems]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 9, Issledovaniya molodykh uchenykh* [Science Journal of Volgograd State University. Young Scientists' Research], no. 13, pp. 105-109.

Babhulgaonkar A.R., Bharad S.V., 2017. Statistical Machine Translation. *Intelligent Systems and Information Management. 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, Aurangabad, pp. 62-67.

Bartsch R., 1987. Norms of Language: Theoretical and Practical Aspects. London, New York, Longman. 348 p.

Caruso G., 2012. French Language Legislation in the Digital Age: The Use of Borrowed English Telecommunication Terms and Their Official French Replacements on Twitter and in the American Foreign Language Classroom. *Theses, Dissertations, and Other Capstone Projects.* URL: https://cornerstone.lib.mnsu.edu/cgi/viewcontent.cgi?article=1119&context=etds.

Chan S.F., DeWitt D., Chin H.L., 2018. The Analysis of Cultural and Intercultural Elements in Mandarin as a Foreign Language Textbooks from Selected Malaysian Public Higher Education Institutions. *MOJES: Malaysian Online Journal of Educational Sciences*, vol. 6 (1), pp. 66-90. URL: https://mojes.um.edu.my/article/view/12512.

Costa-Jussa M.R., Fonollosa J.A., 2015. Latest Trends in Hybrid Machine Translation and Its Applications. *Computer Speech & Language*, vol. 32 (1), pp. 3-10.

Crystal D., 2001. *Language and the Internet.* Cambridge, Cambridge University Press. 272 p.

Cui J., 2012. Untranslatability and the Method of Compensation. *Theory and Practice in Language Studies*, vol. 2 (4), pp. 826-830.

Da Fonseca J., Carolino P., 2002. *O novo guia de conversação em portuguez e inglez.* Casa da Palavra.

De Bot K., Lowie, W., Verspoor M., 2007. A Dynamic Systems Theory Approach to Second Language Acquisition. *Bilingualism: Language and Cognition*, vol. 10 (1), pp. 7-21.

Deutscher G., 2010. *Through the Language Glass: Why the World Looks Different in Other Languages.* New York, Metropolitan Books. 306 p.

Dulov S.Yu., Shmeleva A.G., Boronkinova N.T., 2017. Praktika mashinnogo perevoda i iskusstvennye yazyki v oblasti perevoda [Practice of Machine Translation and Man-Made Languages in Translation]. *Uspekhi v khimii i khimicheskoy tekhnologii* [Advances in Chemistry and Chemical Technology], vol. 31, no. 14 (195), pp. 62-64.

Durdureanu I.I., 2011. Translation of Cultural Terms: Possible or Impossible. *The Journal of Linguistic and Intercultural Education*, vol. 4, pp. 51-63.

Fowler C.A., Hodges B.H., 2011. Dynamics and Languaging: Toward an Ecology of Language. *Ecological Psychology*, vol. 23, pp. 147-156.

Hermans T., 1996. Norms and the Determination of Translation. Alvarez R., Vidal A., eds. *Translation, Power, Subversion.* Clevedon, Multilingual Matters, pp. 25-51.

Hoffman J., 2012. Me Translate Funny One Day. *The New York Times*.

Johnson M., Schuster M., Le Q.V., Krikun M., Wu Y., Chen Z., Hughes M., 2017a. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 339-351.

Johnson R., Pirinen T.A., Puolakainen T., Tyers F., Trosterud T., Unhammer K., 2017b. North-Sámi to Finnish Rule-Based Machine Translation System. *Proceedings of the 21st Nordic Conference on Computational Linguistics, NoDaLiDa, 22–24 May 2017.* Gothenburg, Sweden Linköping University Electronic Press, vol. 131, pp. 115-122.

Koltan O.A., 2017. Osobennosti ispolzovaniya angliyskikh zaimstvovaniy v sovremennom yazyke SMI i povsednevnoy zhizni [Specifics of the Use of English Borrowings in the Modern Language of the Media and Everyday Life]. *Mir yazykov: rakurs i perspektivy: sb. materialov VIII Mezhdunar. nauch.-prakt. konferentsii* [The World of Languages: Foreshortening and Perspectives. Collection of Materials from the 8th International Scientific and Practical Conferences]. Minsk, Izd-vo BGU, pp. 95-100.

Kotov R.G., Marchuk Yu.N., Nelyubin L.L., 1983. Mashinnyy perevod v nachale 80-kh godov [Machine Translation in the Early 80s]. *Voprosy yazykoznaniya* [Topics in the Study of Language], vol. 1, pp. 31-38.

Kövecses Z., 2005. *Metaphor in Culture: Universality and Variation*. Cambridge, Cambridge University Press. 314 p.

Kwee A.T., Tsai F.S., Tang W., 2009. Sentence-Level Novelty Detection in English and Malay. Theeramunkong T., Kijsirikul B., Cercone N., Ho T.B., eds. *Advances in Knowledge Discovery and Data Mining. PAKDD 2009. Lecture Notes in Computer Science*. Berlin, Springer, vol. 5476, pp. 40-51.

Lawrence V., 1995. *The Translator's Invisibility*. New York, Routledge. 353 p.

Lim H., 2018. Design for Computer-Mediated Multilingual Communication with AI Support. *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing ACM*. New York, pp. 93-96.

Lim H., Cosley D., Fussell S.R., 2018. Beyond Translation: Design and Evaluation of an Emotional and Contextual Knowledge Interface for Foreign Language Social Media Posts. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems ACM*. New York, ACM, vol. 217.

Lohar P., Afli H., Way A., 2018. Balancing Translation Quality and Sentiment Preservation (Non-Archival Extended Abstract). *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas*, vol. 1, pp. 81-88.

Low P.A., 2011. Translating Jokes and Puns. *Perspectives: Studies in Translatology*, vol. 19 (1), pp. 59-70.

Marchuk Yu.N., 2016. Kontekstnoe razreshenie leksicheskoy mnogoznachnosti [Resolution of Polysemy in Context]. *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Lingvistika* [Bulletin of the Moscow Region State University. Series: Linguistics], no. 1, pp. 26-32.

Muhammad H.S., 2006. Dalam Daun Ada Bicara: Falsafah Alam Pantun Melayu. Rogayah A.H., Jumaah I., eds. *Pandangan Semesta Melayu Pantun*. Kuala Lumpur, Dewan Bahasa dan Pustaka, pp. 1-32.

Na W.E.I., 2016. Gender Differences in the Use of English Vocabulary Learning Strategies in Chinese Senior High Schools. *Studies in Literature and Language*, vol. 12 (4), pp. 58-62.

Nguyen T.Q., Chiang D., 2017. Improving Lexical Choice in Neural Machine Translation. *Proceedings of NAACL-HLT 2018*. New Orleans,

Louisiana, Association for Computational Linguistics, pp. 334-343.

Novozhilova A.A., 2014. Mashinnye sistemy perevoda: kachestvo i vozmozhnosti ispolzovaniya [Machine Translation Systems: Quality and Possible Ways of Use]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2, Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], no. 3 (22), pp. 67-73. DOI: https://doi.org/10.15688/jvolsu2.2019.3.13.

O'Curran E., 2014. Machine Translation and Post-Editing for User Generated Content: An LSP Perspective. *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas. Vol. 2.* Vancouver, BC, pp. 50-54.

Okamoto S., 2013. Variability in Societal Norms for Japanese Women's Speech: Implications for Linguistic Politeness. *Multilingua*, vol. 32, iss. 2, pp. 203-223.

Oladosu J., Esan A., Adeyanju I., Adegoke B., Olaniyan O., Omodunbi B., 2016. Approaches to Machine Translation: A Review. *FUOYE Journal of Engineering and Technology*, vol. 1 (1), pp. 120-126.

Pasfield-Neofitou S., 2012. 'Digital Natives' and 'Native Speakers': Competence in Computer Mediated Communication. Sharifian F., Jamarani M., eds. *Language and Intercultural Communication in the New Era*. New York, London, Routledge, pp. 138-159.

Razlogova E.E., 2017. Standartnye i nestandartnye varianty perevoda [Standard and Non-Standard Versions of Translation]. *Voprosy yazykoznaniya* [Topics in the Study of Language], no. 4, pp. 52-73.

Riahovskaya A.Yu., 2017. Sravnitelnyy analiz sistem mashinnogo perevoda [Comparative Analysis of Machine Translation Systems]. *Vestnik obrazovatelnogo konsortsiuma Srednerusskiy universitet. Seriya: Informatsionnye tekhnologii*, no. 1 (9), pp. 25-28.

Sanders E.F., 2014. *Lost in Translation: An Illustrated Compendium of Untranslatable Words from Around the World*. Berkeley, California, Ten Speed Press. 112 p.

Sharifian F., 2017. *Cultural Linguistics: Cultural Conceptualisations and Language*. Amsterdam, Philadelphia, John Benjamins, XVII. 171 p.

Steffensen S.V., Fill A., 2014. Ecolinguistics: The State of the Art and Future Horizons. *Language Sciences*, vol. 41, pp. 6-25.

Styblo Jr., M., 2007. *English Loanwords in Modern Russian Language. Master's Dissertation.* Chapel Hill. 72 p.

Sukhoverhov A.V., 2014. Sovremennye tendentsii v razvitii ekolingvistiki [Current Trends and

Developments in Ecolinguistics]. *Yazyk i kultura* [Language and Culture], no. 3 (27), pp. 166-175.

Sukhoverkhov A.V., 2015. Lingvisticheskiy determinizm, kumulyativnaya evolyutsiya i rost nauchnogo znaniya [Linguistic Determenism, Cumulative Evolution and Development of Scientific Knowledge]. *Politematicheskiy setevoy elektronnyy nauchnyy zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta* [Polythematic Online Scientific Journal of Kuban State Agrarian University], no. 105, pp. 1-22.

Sukhoverkhov A.V., Fowler C.A., 2015. Why Language Evolution Needs Memory: Systems and Ecological Approaches. *Biosemiotics*, vol. 8 (1), pp. 47-65.

Toury G., 1995. *Descriptive Translation Studies and Beyond*. Amsterdam, Philadelphia, John Benjamins. 311 p.

Verspoor M., De Bot K., Lowie W., eds., 2011. *A Dynamic Systems Approach to Second Language Development: Methods and Techniques.* Amsterdam, Philadelphia, John Benjamins. 211 p.

Wang X., Lu Z., Tu Z., Li H., Xiong D., Zhang M., 2017. Neural Machine Translation Advised by Statistical Machine Translation. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, California, AAAI, pp. 3330-3336.

Whorf B., 1956. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf.* Cambridge, MIT Press. 290 p.

Wiechetek L., 2008. *Rule-Based MT Approaches Such as Apertium and GramTrans*. URL: https://uit.no/Content/84555/cache=20171811052806/mt.pdf.

Wierzbicka, A., 1992. *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configurations*. New York, Oxford University Press. 487 p.

Wu Y. et. al., 2016. *Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation*. 23 p. arXiv:1609.08144

Yusoff N., Jamaludin Z., Yusoff M.H., 2016. Semantic-Based Malay-English Translation Using N-Gram Model. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 8 (10), pp. 117-123.

Zaremba W., Sutskever I., Vinyals O., 2014. *Recurrent Neural Network Regularization*. 8 p. arXiv:1409.2329

Information about the Authors

Anton V. Sukhoverkhov, Candidate of Sciences (Philosophy), Associate Professor, Department of Philosophy, Kuban State Agrarian University, Kalinina St. 13, 350044 Krasnodar, Russia, sukhoverkhov.ksau@gmail.com, https://orcid.org/0000-0002-0357-4013

Dorothy DeWitt, PhD, Associate Professor, Department of Curriculum and Instructional Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia, dorothy@um.edu.my, https://orcid.org/0000-0003-3123-7150

Ioannis I. Manasidi, Student, Faculty of Applied Informatics, Kuban State Agrarian University, Kalinina St. 13, 350044 Krasnodar, Russia, johnfrom636@gmail.com, https://orcid.org/0000-0002-2090-9970

Keiko Nitta, PhD, Professor, Department of Letters, College of Arts, Rikkyo University, 3-34-1 Nishi Ikebukuro, 171-8501 Toshima City, Tokyo, Japan, knitta@rikkyo.ac.jp, https://orcid.org/0000-0002-6963-711X

Vladimir Krstić, PhD, Honorary Research Associate, Department of Philosophy, University of Auckland, Private Bag 92019, 1142 Auckland, New Zealand, v.krstic@auckland.ac.nz, https://orcid.org/0000-0003-1953-2675

Информация об авторах

Антон Владимирович Суховерхов, кандидат философских наук, доцент кафедры философии, Кубанский государственный аграрный университет, ул. Калинина, 13, 350044 г. Краснодар, Россия, sukhoverkhov.ksau@gmail.com, https://orcid.org/0000-0002-0357-4013

Дороти де Витт, PhD, доцент кафедры учебных программ и технологий обучения, Малайский университет, 50603 г. Куала-Лумпур, Малайзия, dorothy@um.edu.my, https://orcid.org/0000-0003-3123-7150

Иоаннис Игоревич Манасиди, студент факультета прикладной информатики, Кубанский государственный аграрный университет, ул. Калинина 13, 350044 г. Краснодар, Россия, johnfrom636@gmail.com, https://orcid.org/0000-0002-2090-9970

Кейко Нитта, PhD, профессор кафедры литературы, Колледж искусств, Университет Риккё, 3-34-1 Ниши Икебукуро, 171-8501 Тошима, Токио, Япония, knitta@rikkyo.ac.jp, https://orcid.org/0000-0002-6963-711X

Владимир Крстич, PhD, почетный научный сотрудник, кафедра философии, Университет Окленда, Прайвит Бэг 92019, 1142 Окленд, Новая Зеландия, v.krstic@auckland.ac.nz, https://orcid.org/0000-0003-1953-2675