Text Globalization and Machine Translation (Case of English-Russian Translation of U.S. Travel Site)

Irina D. Volkova[*]

Translation Studies Department, Volgograd State University, 400062 Volgograd, Russia

Abstract. The paper deals with the pre-editing process before it goes through machine translation. The main argument is to improve the source in order to improve the raw output quality. The mentioned process is investigated through the prism of the concept of a globalized text, thus making contribution to the research into "translation and globalization paradigm" being of interest over the past years. The content analysis of original and Google-translated texts of the U.S. tourism portal allows one to point out some principles of composing a global English-language text that would make the process of English-Russian translation more simple and productive. We conclude that the content writer should take into account certain peculiarities of the translation process, and they have been pointed out as well.

1 Introduction

Technological advances in translation studies provide wider opportunities for dealing with various translation tasks. Machine translation turns out to be one of such advances, which has constantly been improved and upgraded through the enhanced programs of statistics analysis and data processing. Such popular services as Babelfish and Google Translate are often used to meet the communication needs of the globalized world community, in particular on the Internet platform. Content of international websites as a translation object provides information to customers from all over the world, thus making the translator face the challenge of content adaptation. This challenge lies at the core of such widely known phenomena as localization and globalization – the processes aimed at ensuring availability and readability of information for large audiences from all over the world. While a large number of international websites have been successfully localized over the last decade, the concept of a global or globalized text requires much attention at the current stage of multicultural communications taking place in different contexts and formats. The key thesis of the present research: an English-language text, which fulfils the functions of the global text within the mono- or multilingual website space, should be composed in a manner allowing for its further correct machine translation. In this connection, the purpose of the research consists in pointing out the principles of composing a global English-language text which would serve its pragmatics functions even through the use of online translation tools by foreign readers. The relevance is determined by the need to make up such tourism texts and post them on international travel websites under the mark 'Global Site'.

2 Translation in the tourism sector

2.1. Website Global Version

The tourism sector is characterized by great importance from the viewpoint of economic development, and the information logistics of this sector needs linguistic interpretation and guidelines. Tourism websites, travel portals in particular, represent a good example of the global resource for travellers from all over the world. Along with several localized versions, websites may have a special language option marked 'Global Site' or 'International Site' aimed at audiences not included into a localization range due to some reasons, namely, high costs and the audience range being too diverse to adapt for each one. A large number of official online portals (for example, travel portals of Poland, Great Britain, Norway, New Zealand, Spain, Ireland etc.) have a Global English version, and even the larger number needs one. However, the knowledge on creating a global resource is scattered and has not yet been actually applied. The task of making up tourism texts, which may be addressed to large audience, a "global addressee", seems to be rather difficult since up-to-date the status of a global addressee is still under question – we do not know exactly what their level of proficiency in English, what countries they come from (countries of inner, outer or extending circles) [1]. Modeling a global text contributes to the development of a communication & pragmatics paradigm in the field of translation studies. The inclusion of the globalization process into the practical tasks of a translator seems to be one of the urgent tasks in the conditions of expanding global

---

[*] Corresponding author: idvolkova@volsu.ru

village. Moreover, revealing certain principles of text globalization provides tools for editing and translation activity with regard to a number of online and print discourses.

2.2. Machine translation of website content

Enthony Pym substantiates the idea that the global text should be suitable for machine translation since this is a way to ensure its readability [2: 180]. The author argues that when a document has a limited number of syntactic structures and a completely controlled multilingual terminology, machine translation plus reviewing can make the translation process almost automatical. We are aware that even the most accurate machine translation technology can't automatically account for branded content, terminology preferences, formatting considerations, subject-matter, cultural issues, and other important factors [3]. The typical mistakes of machine translation done by Google Translate and Prompt were described by I.A. Borisova [4: 55-56]. They include improper use of articles, wrong translation of words and word combinations, use of original punctuation marks, grammatical mistakes, word for word translation, use of personal pronouns in impersonal constructions, etc. [4]. None of machine translators is able to translate the whole presented text correctly, the greatest challenge being the Russian case pattern [5]. Perhaps the most promoted in Russia nowadays Google Translate is still far from being perfect giving rise to complaints of professional translators but helps unsophisticated users to understand text fragments in an unknown language at large [5].

3. Literature review

The concepts of Global English and World Englishes got much attention in Western research on incorporating varieties of English into education patterns and ELT in particular [6]. It is a global language with a global ownership [6], and this is the very meaning of the word 'global' which is often implied in diverse studies and contexts. According to B. Seidlhofer, the concept of International English rather means the use of English on the international scale, than a certain language variety [7].

There is also a specific research area dealing with the concept of 'global text' originating from developments in the field of English language simplification. The simplified versions of English include Ogdens' Basic English, Jean-Paul Nerrière's Globish, as well as Plain English and Minimal English. Texts written in either of these versions are meant to simplify the process of information decoding by non-native speakers of English. Basic English is a simplified subset of regular English created by linguist and philosopher Charles Kay Ogden as an international auxiliary language, and as an aid for teaching English as a second language [8]. Jean-Paul Nerrière's Globish is a trademarked name for a subset of standard English grammar, and a list of 1500 English words. Nerrière claims it is "not a language", but rather it is the common ground that non-native English speakers

adopt in the context of international business [9]. The Plain English Campaign (PEC), established in 1979 and based on the principle of everyone having access to clear and concise information, helped many government departments and other official organisations with their documents, reports and publications [10]. Today the campaign got wide popularity within the special web-platform offering a large number of commercial services on editing English texts. The concept of Minimal English was introduced by Anna Wierzbicka in 2014 to denote a radically reduced 'Mini-English' – an inter-language, a global means for discussions and information exchange [11: 194]. Cliff Goddard points to cross-translatability as a key distinction between 250 words of Minimal English and all the rest simple English models [12].

It should be noted that these conceptions consider the global language and the global text within the scale of one language – English. They got further interpretation and development both in foreign and Russian studies. John Kohl gives guidelines how to write documentation that is optimized for non-native speakers of English, translators, and even machine-translation software [13]. Russian researchers analyze Internet communication [14-16] proceeding from the need to make an international website play the role of a multicultural platform by means of global text options.

4. Materials and methods

The focus on the two-facet nature of the globalization process covering both editing an English text and translating from English is of special importance, particularly if we take into consideration the problem of working out a multilingual tourism website. The efficiency of machine translation technologies depends on obligatory presettings and initial text quality [5], and for many language pairs, we are now at the stage where it is quicker to "post-edit" machine-translation output than to start translating from scratch [17; 18].

As shown in [19], global versions of tourism websites rarely conform with the existing techniques of English text adaptation, which, by the way, proved to be effective in the process of translating texts of online tourism discourse widely represented on numerous web portals to different cities and countries.

The initial stages of ongoing research on the attitude of Russians to English-language online content let reveal the trend of using Google Translate services instead of further search for the necessary information in Russian. In the present research we analyze the U.S. tourism portal VisitTheUSA.com (https://www.visittheusa.com/) – the official travel site of the USA which has several language options – Global English, English (U.K.), English (Australia), English (Canada), English (India), French (Canada), Portuguese (Brazil), Sweden, French (France), German (Germany), Korean (Korea), Spanish (Chile), Spanish (Colombia), Spanish (Mexico), Japanese (Japan), Simplified Chinese (China). The content analysis shows that the Global English version is the primary one in the site structure and is supposed to

be 'global' that is simple, clear and easy for translation by non-native speakers. We will focus rather on "pre-edit" machine translation and analyze the translation mistakes that can be prevented by means of creating texts in Global English specially designed for further translation into the Russian language via Google Translate.

5. Results

Translation of texts from section Destinations. Alyaska (https://www.visittheusa.com/state/alaska) let reveal some common Google mistakes. Table 1 represents the fragments of original, adapted and translated versions of the text *The Alaskan wilderness: America's final frontier*.

Table 1. The fragments of original adapted and translated versions of the text *The Alaskan wilderness: America's final frontier*.

	Text fragment	Google Translate (Russian transliteration)
original	*With awe-inspiring scenic beauty, enormous ice-age glaciers and abundant wildlife,* ***Alaska is like nowhere else on earth***.	*Obladaya vpechatlyayushchey zhivopisnoy krasotoy, ogromnymi lednikami lednikovogo perioda i obilnoy dikoy prirodoy,* ***Alyaska nigde ne vstrechaetsya na zemle***.
adapted	*Alaska is a* ***unique / exceptional / one-of-a-kind*** *place on earth.*	*Alyaska yavlyaetsya* ***unikalnym / isklyuchitelnym / edinstvennym*** *v svoem rode mestom na zemle*

According to Collins' dictionary [20], the pronoun 'nowhere' has 9 shades of meaning in different contexts, 'in, at, or to no place; not anywhere' being the primary one. This means that we should edit the phrase by replacing the negative grammatical form 'like nowhere else' with a positive one. The presented adjectives (unique / exceptional / one-of-a-kind) preserve the original proposition of exceptional nature. As we can see from Table 1, testing the adapted options in Google Translate proves the efficient decoding of sense. The illustrated negative-for-positive substitution coincides with one of the Plain English principles – 'Write in the positive' implying that positive sentences are shorter and easier to understand than their negative counterparts (A Plain English Handbook).

Table 2 represents the fragments of original, adapted and translated versions of the text *Watch the Wildlife*. Conversion as the derivation of a new word without any overt marking, in particular noun to verb and verb to noun, is one of the many ways to create new English words on the basis of already existing ones [21: 134]. The above-mentioned fragment demonstrates the mistake of speech part identification in the process of machine translation: the original 'spawn' is used as a

verb, while the translated lexeme functions as a noun. This implies that the global text should be structured so as to exclude such variance. For example, pronoun 'which' may be used before an ambiguous word (see adapted fragment 1). We may also point to deviance in grammar structure caused by punctuation mistake – absence of comma after 'traffic jams' in the original phrase. Inserting comma between two clauses recovers the actual sense (see adapted fragment 2).

Table 2. The fragments of original, adapted and translated versions of the text *Watch the Wildlife*.

	Text fragment	Google Translate (Russian transliteration)
original	*Bald eagles gather by the hundreds, moose cause traffic jams and* ***millions of salmon spawn in its rivers***.	*Lysye orly sobirayutsya sotnyami, losi vyzyvayut probki i* ***milliony lososevykh nerestov v ego rekakh***.
adapted (1)	*...and* ***millions of salmon which spawn in its rivers***>	*...**milliony lososey, kotorye poyavlyayutsya v ego rekakh***.
adapted (2)	*...moose cause traffic jams, and* ***millions of salmon spawn in its rivers***.	*...losi vyzyvayut probki, a v rekakh - milliony lososey*.

Table 3. The fragments of original, adapted and translated versions of the text *Take a Breathtaking Drive*.

	Text fragment	Google Translate (Russian transliteration)
original	***Take a Breathtaking Drive*** *...**Fly** to Alaska and **rent** a car or **RV**. You'll find 3 routes that have been designated **National "Scenic Byways"***	***Vozmite zakhvatyvayushchiy drayv*** *...**Vylet** na Alyasku i **arenda** avtomobilya ili **RV**. Vy naydete 3 marshruta, kotorye byli oboznacheny **National «Scenic Byways»***
adapted	***Go on a breathtaking trip*** ***You can fly to Alaska*** *and **rent** a car or **research vessel**. You'll find 3 routes that have been designated **national scenic byways**.*	***Otpravtes v zakhvatyvayushchuyu poezdku*** ***Vy mozhete letet*** *na Alyasku i arendovat avtomobil ili **issledovatelskoe sudno**. Vy naydete 3 marshruta, kotorye byli oboznacheny **natsionalnymi zhivopisnymi marshrutami**.*

Table 3 represents the fragments of original, adapted and translated versions of the text *Take a Breathtaking Drive*.

The analysis of the given fragment allows one to point out the following words and word combinations, which should be corrected. According to Wayne Magnuson Dictionary of English Idioms, Sayings and Slang [22], 'take a drive' is an idiom with a meaning 'go for a drive, go on a trip', and the best option, in our view, is to use the heading 'Go on a breathtaking trip', which is translated by Google with a tempting offer – 'Otpravtes v zakhvatyvayushchuyu poezdku'. Abbreviation RV requires decoding in the English text – 'research vessel', since the translation of such lexemes represents one of the most urgent issues of machine translation, particularly when we talk about specialized abbreviations which are not characterized by frequent use. The translation of the verbs 'fly' and 'rent' with nouns 'vylet' and 'arenda' illustrates the highly mentioned problem of determining the part of speech by translation program. In this case this problem may be solved by using a pronoun and a modal verb (You can fly and rent).

As we can see from Table 3, National "Scenic Byways" is not translated by Google due to the use of quotation marks and capital letters. Elimination of these markers allows for the correct translation (see adapted fragment).

Table 4 represents the fragments of original, adapted and translated versions of the text *Visit the Country's Largest National Park*.

Table 4. The fragments of original, adapted and translated versions of the text *Visit the Country's Largest National Park.*

	Text fragment	Google Translate (Russian transliteration)
original	*Alaska is home to 17 national park areas.*	*Na Alyaske prozhivayut 17 natsionalnykh parkov*
adapted (1)	*Alaska has 17 national park areas*	*Na Alyaske naschityvaetsya 17 natsionalnykh parkov*
adapted (2)	*17 national park areas are located in Alaska*	*17 natsionalnykh parkov raspolozheny na Alyaske*
adapted (3)	*Alaska includes 17 national park areas*	*Alyaska vklyuchaet v sebya 17 natsionalnykh parkov*

Using the word 'home' in the given context certainly intensifies the journalism style of the original text, but this intensifier turns out to be undeciphered by Google Translate. In our opinion, a globalized text should mostly consist of lexical units excluding the problem of sense ambiguity in machine translation to a maximum possible extent. Talking about the location of parks in Alaska, we may consider such verbs as *locate, have, include*, etc. (see adapted fragments 1, 2, 3).

Table 5 represents the fragments of original, adapted and translated versions of the text *Cruise to a Glacier*.

Table 5. The fragments of original, adapted and translated versions of the text *Visit the Country's Largest National Park*.

	Text fragment	Google Translate (Russian transliteration)
original	*You can get up close to glaciers on cruise tours and day cruises in many areas of Alaska including **Glacier Bay, Kenai Fjords** and **Prince William Sound**.*	*Vy mozhete podnyat'sya blizko k lednikam v kruiznykh turakh i dnevnykh kruizakh vo mnogikh rayonakh Alyaski, vklyuchaya **lednik, Kenay Fyords i Prints Uilyam Zvuk**.*
adapted	*You can get up close to glaciers on cruise tours and day cruises in many areas of Alaska including **"Glacier Bay", "Kenai Fjords"** and **"Prince William Sound"**.*	*Vy mozhete podnyatsya blizko k lednikam v kruiznykh turakh i dnevnykh kruizakh vo mnogikh rayonakh Alyaski, vklyuchaya **«Glacier Bay», «Kenai Fjords» i «Prince William Sound»**.*

A special role in the process of translating tourism texts is played by linguistic and cultural adaptation of microtoponyms – proper names of human-created environment objects, such as squares, parks, museums, theaters and other cultural facilities, which create the attractive image of a city or a certain tourist destination [23]. As we can see from Table 4, Google Translate misrepresented both the meaning and the original form of the three given microtoponyms. We have already revealed that Google does not translate the words in quotation marks (see Table 3), therefore the best option in this case would be to place the microtoponyms in quotation marks, thus preserving their original English spelling (see adapted text fragment in Table 5). These words will surely not be understood by Russian readers, but will ensure easy search for more detailed information about the destinations. Moreover, the fragments of local languages bear the concept of personal enrichment [24], and a function as original markers of 'otherness' [25: 25].

Table 6. The fragments of original, adapted and translated versions of the text *Visit the Country's Largest National Park*.

	Text fragment	Google Translate (Russian transliteration)
original	***Stay** at a **backcountry lodge** for **fly-fishing** and **bear watching**.*	***Prebyvanie v bekkantri lodzh** dlya **lovit rybu i smotret medved**.*
adapted	*You **can stay** at a **remote house**, go **fishing** and **watch bears**.*	*Vy **mozhete ostatsya v otdalennom dome**, **otpravitsya na rybalku i nablyudat za medvedyami**.*

Table 6 represents the fragments of original, adapted and translated versions of the text "Explore the Wilderness".

The given fragment illustrates the need to use more common words and word combinations (backcountry – remote; lodge – house, fly-fishing –fishing) and to test gerund forms since Google Translate is not able to preserve their grammatical meaning in some cases.

6. Conclusion

In general, we may note a relatively high level of Google translatability of the analyzed texts: only 20% of the whole text corpus needs post-editing or, in our case, pre-edit translation. The content analysis of original and Google-translated texts of the U.S. tourism portal let point out some principles of composing a global English-language text that would make the process of English-Russian translation more simple and productive. In case an English-language website has no localized Russian version, but it is implied that Russian readers may be interested in the content, then the content writer should take into account the following peculiarities of the translation process:

– negative-for-positive substitution makes sentences clearer and easier to Google translate;
– the conversion as a word-formation model should be taken into account to avoid the variance in part-of-speech decoding by machine translator (for example, in order to highlight the verb grammatical meaning, pronoun *which* or modal verbs may be used: *spawn – which spawn; fly – can fly*);
– it is necessary to check the punctuation of the original text to avoid mistakes;
– Google Translate preserves the original spelling of words in quotation marks and capitalized words. Therefore, when we need the equivalent translation, we should exclude these markers;
– in some cases, preserving the original spelling of words by the abovementioned means may be productive, microtoponyms being a good example since they are hard for machine-decoding;
– abbreviations should be explained;
– words in transferred use create ambiguity of meaning in the text of translation.

Actually, the analyzed process of editing English texts aimed at making them appropriate for machine translation is characterized by creative character and may imply a number of possible variants each ensuring the correct further translation. The proposed technique of adjusting original texts to Google Translate and other translation services may also be used with regard to other languages.

References

1. B. Kachru, *Institutionalized second-language varieties* (The English language today) ed. S. Greenbaum (Oxford: Pergamon, 1985)
2. A. Pym, *Exploring Translation Theories* (New York: Routledge, 2014)
3. *Best Practices on Using Machine Translation.* URL: http://www.transperfect.com/blog/best-practices-using-machine-translation.
4. I.A. Borisova, Vestnik Moskovskogo gosudarstvennogo lingvisticheskogo universiteta **13**, 53-59 (2014)
5. D.A. Morel, E.I. Gimazitdinov, *Science, Engineering, Innovations, 2014: Collected Articles International Scientific and Technological Conference* (Bryansk, March 25-27, 2014)
6. N. Galloway, H. Rose, ELT Journal, 1-14 (2017)
7. B. Seidlhofer, *A concept of international English and related issues: from real English' to 'realistic English'?* (Strasbourg: Council of Europe, 2003)
8. Ch.K. Ogdens, *Basic English: a general introduction with rules and grammar* (Trench, Trubner & Co., Ltd., 1930)
9. J.-P. Nerrière, *Parlez Globish! Don't speak English. Broché* (Paris, Groupe Eyrolles, 2004)
10. *Plain English Campaign.* URL: http://www.plainenglish.co.uk/.
11. A. Wierzbicka, Talking about the Universe in Minimal English: Teaching Science through Words that Children Can Understand. *Minimal English for a Global World: Improved Communication Using Fewer Words* (Palgrave Macmillan, 2018)
12. C. Goddard, *Minimal English for a Global World: Improved Communication Using Fewer Words.* (Palgrave Macmillan, 2018)
13. J.R. Kohl, *The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market* (SAS Institute, 2008)
14. I.G. Fomicheva, Foreign languages in higher education **4**, 54-59 (2009)
15. E.A. Grigoryeva, *Student Scientific Forum: materials of the VII International Student Electronic Scientific Conference* (2015)
16. A.A. Ionina, "Globalization" of the English text in international translation studies. *Language and culture in the era of globalization: materials of the first international scientific conference.* (St. Petersburg, 2013)
17. A. Pym, *On the ethics of translators' interventions Tarragona* (Spain, 2009)
18. I. Garcia, Machine Translation **25** 217-237, (2011)
19. I.D. Volkova, *Globalization and Localization as the Vectors of Travel E-Discourse Development. PhD Thesis* (Volgograd, 2017)

20. *Collins' dictionary*. URL: https://www.collinsdictionary.com/dictionary/english/nowhere.

21. I. Plag, *Word-formation in English*. (Universität Siegen, 2002)

22. *Wayne Magnuson Dictionary of English Idioms, Sayings and Slang*. URL: http://esl-bits.net/idioms/id585.htm.

23. I.D. Romadina (Volkova), V.A. Mityagina, X Linguae Journal **10**, 112-124 (2017)

24. P. Bourdieu, *Language and Symbolic Power*. (Cambridge: Polity Press, 1991)

25. E.M. Maslennikova, Bulletin of the Chelyabinsk State University **6,** 22-25 (2014)