The Improvement of Machine Translation Quality with Help of Structural Analysis and Formal Methods-Based Text Processing

Anna Mylnikova and Aigul Akhmetgaraeva
Perm National Research Polytechnic University, Komsomolsky Ave. 29, Perm, Russia
novikova@yandex.ru, aigul15.ahmetgaraeva@yandex.ru

Keywords: Machine Translation, Machine-Aided Translation, Language Pair, Classification, BLEU Scores, the Frequency of Vocabulary Use, Algorithm, the Evaluation of Translation Quality.

Abstract: This article considers the issues of enhancing the quality of machine translation from one language into another one by structuring linguistic patterns and using identification methods for the situations that cannot be processed by the suggested approach and are subject to individual processing. According to the BLEU score metrics, the described approach allows to increase the quality of machine translation on average by 0.1 and reduce postprocessing time due to the identification of idioms and words with context-dependent meanings by translation. The experiment data base of the study was built upon online available pairs of texts that cover the events of FIFA World Cup 2018 and well-known idioms.

1 INTRODUCTION

Beginning with the industrialization age, there is an ongoing growth in labor efficiency. The issues of sharing knowledge, activity outcomes and technologies have become global and, hence, communication in foreign languages has gradually penetrated from elite environment into routine activity. The emergence of new products and projects leads to a big amount of in-line documentation and correspondence issues. This process entails works on preparing various documents in different languages and their translation.

First, we encountered machine and machine-aided translation systems in science fiction books and movies, but in the middle of the XX century the organizations Warren Weaver of the Rockefeller Foundation and RAND announced the possibility of making translations from one foreign language into another one through a mediator – computer [1] and started to implement that idea as a part of projects. The conceptual guidelines of machine translation system operation are laid down in the works of A. Vakher, W. Weaver, H.P. Edmundson, P.G. Hays, G. Artsrouni and P.P. Smirnov-Troyanskii [2]. The research outcomes of this scientific school distinguished machine translation into a science intensive direction, that got exponential development by introducing such methods as 1) structural grammatical methods (GAT, COMIT, METAL, ESPERANTO), 2) syntactic methods (P. Garvin, E. Brown, A. Lukjanow, etc.) 3) semantic approaches (ETAP-1,2,3, DLT, Rosetta, KANT).

The current studies are focused on the issues of enhancing the quality of machine translation and, as a rule, take advantage of hybrid models that combine the methods of corpus linguistics, statistical analysis and cognitive analysis on the basis of the methods that are developed in the theory of intelligence systems [3], [4], [5].

2 THE CURRENT STATE OF MACHINE-AIDED TRANSLATION METHODS AND SYSTEMS

Meaningful translation from one foreign language into another one underlies the identification of the syntactic structure of source-language and the model that actualizes the in-depth and external semantics of this phrase and, hence, the identification of a single value matching on the syntactic and semantic levels of target-language. This task is challenging due to some reasons. First of all, the difference in syntactic structures of natural languages leads to an effect of rigid and “not rigid” localization effects [6], when, in particular, one syntactic structure of English or German languages can be assigned to up to 4 variants of syntactic structures in Russian language due to its not rigid theme - rheme based order; however, the semantic content in the latter 4 Russian variants maintains generally equal.

It is assumed that the formalization of linguistic structures for source-language and target-language as well as the development of their match pattern base can help achieve meaningful machine translation. In the 98s, K.A. Papenini suggested that this problem can be tackled by using direct maximum entropy translation models [7]. The drawback of such models is a strict limitation of parallel data. The German scientists Franz J. Och and Hermann Ney developed this idea for statistical machine translation by introducing conventional dynamic programming search algorithms. With help of Bayes' decision rule, they included a dependence parameter on the hidden variable of the translation model [8]. However, this model works only by true probability distributions, which is not always the case due to differences in language systems and the nature of thought unfolding in different languages [9], [10].

Another popular approach today is an approach which is based on the methods of machine learning. For instance, in 2010 the corporation Google developed and embedded the method of cross-language near duplicate detection by using parallel document mining for statistical machine translation system learning [11]. In this approach they extend the local distribution distance of a word or phrase to be translated and apply deep learning methods to teach neural networks. Currently, the system of machine translation Google identifies the local distribution distance within 8 words [12], [13] and does not cover the lexical and grammatical context of the whole phrase. As a result, it entails a number of translation mistakes.

3 THE COMPARISON OF MACHINE TRANSLATIONS AND THE ANALYSIS OF MISTAKES

Based on the analysis of text translations of various thematic scope websites, news blocks devoted to the coverage of FIFA World Cup 2018 events (official texts translated in many languages were taken as most accurate translations since they were translated by professional translators which ensures the accuracy of professional terms, well-known expressions and idioms used in translation), performed by the machine translation systems Google, PROMT, SYSTRAN, Babylon, Microsoft translator, Yandex translator we can observe only a low quality of machine translations. See the results of the BLEU score metrics used for the evaluation of machine translation quality [14], [13] in the Table 1.

Table 1: The evaluation of machine translation quality made by the BLEU score metrics.

Machine translation system	BLEU score metrics (Russian-English)	BLEU score metrics (English-Russian)
Google	0.298	0.5
PROMT	0.232	0.413
SYSTRAN	0.155	0.175
Babylon	0.26	0.45
Microsoft translator	0.307	0.51
Yandex translator	0.304	0.58

Taking into account the fact that according to the BLEU score metrics the highest result corresponds to the value «1», we can conclude that nowadays the problem of producing accurate meaningful translation from the source-language to the target language is not solved yet. Therefore, it is important to understand the reasons of such low quality. For this case, let us analyze the most frequently observed mistakes, see the Figure 1.

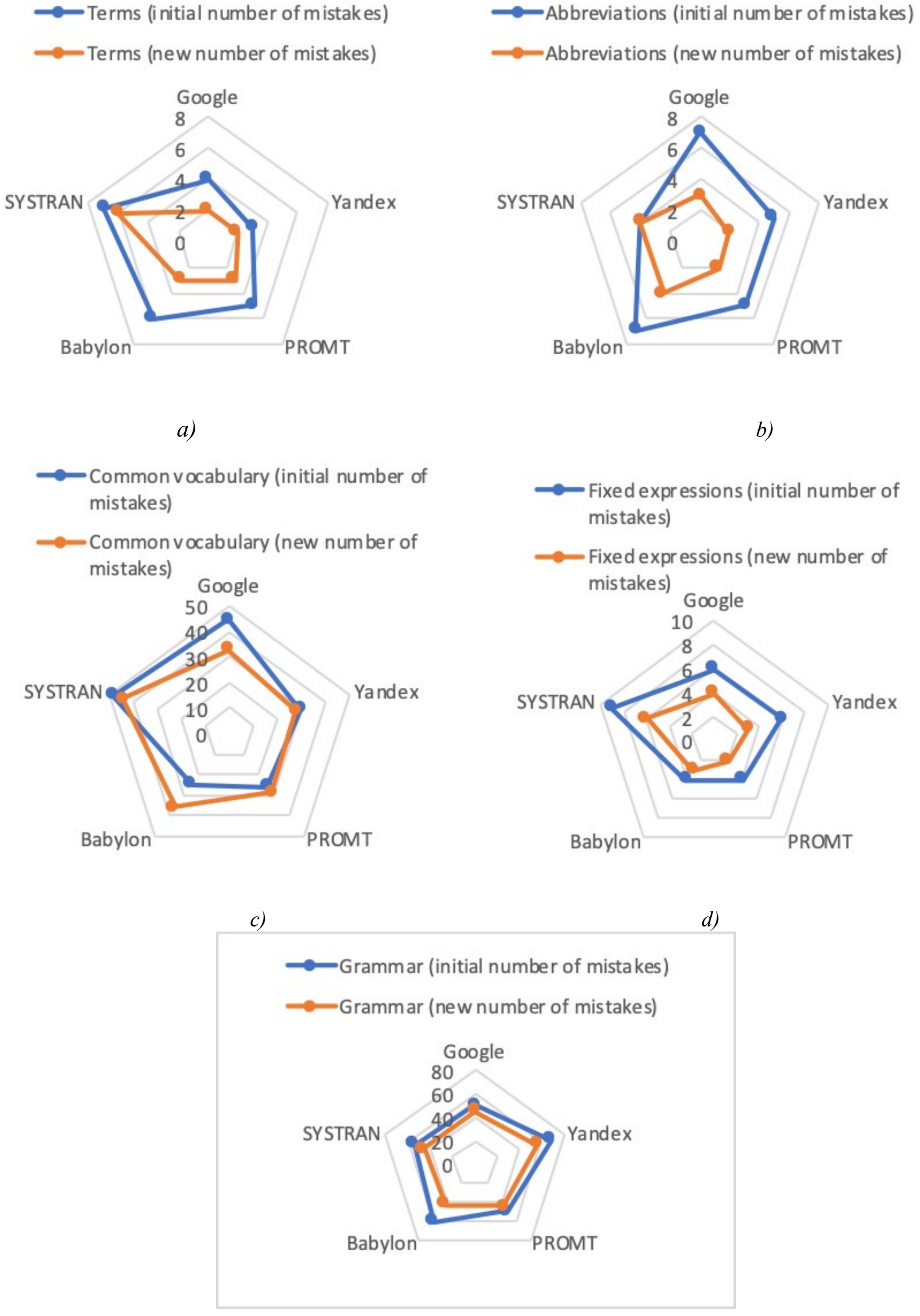

Figure1: Numerical reduction of mistakes by using initial source-language structures and decoded source-language structures: *a)* terms, *b)* abbreviations, *c)* fixed expressions, *d)* common vocabulary, *e)* grammar.

After seeing the given statistics, it is obvious that systems make various errors. In particular, grammar and in some cases semantic errors are dependent on the structure used in the source-language. At the same time, the distributive location of head lexical transducers is significant since it affects the lexical-grammatical phrase realization in English language [9]. We could not agree more with the statement of E. Sumita and H. Iida that «example-retrieval cost is high when the input sentence is syntactically ambiguous» [15].

The decoding of source-language syntactic structures in line with the structures of target-language grammar system allows to reduce not only the number of grammar mistakes, but also semantic mismatches [16], [17].

However, in machine translation this approach does not help avoid all the sense distorting semantic mistakes. The use of idioms, fixed and professional terms leads to a word-by-word translation, that distorts the meaning of the phrase. For instance, the phrase in Russian language «Я не уверен, смогу ли выслать доклад сегодня вечером, он еще совсем сырой» was translated by Google into English language as follows: «I'm not sure if I could send a report tonight, he's still very raw». The sense of the phrase is not given correctly though, as the accurate translation of the Russian phrase into English language corresponds to the following phrase: «Most likely I won’t be able to send the report tonight, it’s far from done».

It is getting worse when it comes to idioms as whole phrases (see the Table 2).

It is logically to assume that such situations cannot be processed correctly with help of the existing concepts and methods. The identification of such “special” linguistic expressions with further individual processing could be a way to deal with this challenging situation. Hence, we need to determine the attributes which are required for identifying such expressions.

Table 2: Examples of idiom machine translation.

Source-language phrase	**Accurate translation into target-language**	**Google**	**PROMT**	**Yandex**	**Babylon**	**SYSTRAN**
A) English-Russian						
1. Born with a silver spoon in his mouth	**Родившийся под счастливой звездой**	Родился с серебряной ложкой во рту	Терпевший серебряная ложка в его рту	Родился с серебряной ложкой во рту	Родился с серебряной ложкой во рту	Принесенный с серебряной ложкой в его рте
2. An old head on young shoulders	**Мудр не по годам**	Старая голова на молодых плечах	Старая голова на молодых плечах	Старая голова на молодых плечах	Старая голова на молодых плечах	Старая голова на молодых плечах
3. To have one’s head in the clouds	**Витать в облаках**	Иметь голову в облаках	Витать в облаках	Чтобы иметь голову в облаках	На голова в облаках	Иметь one голову в облаках
4. To take it on the chin	**Не падать духом**	Взять его на подбородок	Взять его на подбородке	Чтобы взять его на подбородок	Принять его на подбородке	Принять его на подбородке
B) Russian-English						
1. Уйти по-английски	**To take French leave**	Take French leave	To take French leave	Leave in English	Take French leave	To leave in English
2. Подложить свинью	**To play a dirty trick**	Put a pig	Play a dirty trick	A pig in a poke	Send to a pig	To place the pig
3. Ударить в грязь лицом	**To have egg on one’s face**	Smash face	To lose face	To strike in a dirt the person	Hit the dirt in the face	To strike into mud by face
4. У чёрта на куличиках	**In the middle of nowhere**	At the damn thing	At the world's end	In the middle of nowhere	The feature on the куличках	In feature on kulichkakh

4 MACHINE TRANSLATION IMPROVEMENT

4.1 The Use of Distributive Localizations to Enhance the Quality of Machine Translation

Distributive localizations are used on the bases of the structuring methods described in [17]. In contrast to the suggested approach, we will use a group of overlapping dependencies that cancel the action of other dependencies by the emergence of certain distributive localizations. This expands the baseline translation system by adding new functional dependencies and, hence, allows to achieve meaningful alignment of the source-language and the target-language without parallel data limitation. For instance, for the case 1A,2A,3A,4B,5A,6A, where 1A – the declarative sentence, 2A – the indicative mood, 3A – the active voice; 4B – the present simple tense, 5A – the affirmation, 6A – the simple predicate (actualized by a notional verb or a copulative verb):

[Parenth][NP3]<NP1>[AdvP1.2][AdvPM][Adv/measure]<VP>[NP2][NP2ext][AdvP3][AdvP2][AdvP1.1]<".">,

where Adv/measure is not applied in one distributional context together with AdvP1.2, AdvP1.3., AdvPM; AdvP2 is not applied in one distributional context together with AdvP1.1, AdvPM, Adv/measure; AdvP1.1 is not applied in one distributional context together with c AdvP2.

Such action will allow to improve machine translation. The Table 3 shows the translation examples from Russian language into English language performed by Google based on the source-language phrase structures and those decoded in accordance with the structure of the target-language. The formalization language suggested in [17] is taken to describe linguistic structures.

4.2 Machine Translation of Idioms and Terms

When working with idiomatic expressions, phrases and terms, which were mentioned in the previous section, the suggested approach helps achieve grammatically correct, but not semantically sound translations. Hence, the outcome does not make sense to the native speaker.

It is assumed that one of the ways to handle this issue is the identification of such phrases, expressions and words and their special processing (post-editing or manual translation).

In this study, idioms are assigned to phrases and expressions, that have similar meanings yet different lexical-grammatical actualization in the source-language and target-language (see examples in the Table 2), terms are assigned to certain words that generate context-dependent meanings (for instance, professional terms).

Table 3: Machine translation examples made by Google without the decoding of the source-language structures and with the decoding of source-language structures.

	Phrase structure	Machine translation	Number of mistakes
Example 1 – «Они в компании всегда быстро проводят обновление программного обеспечения»			
Source-language structure	<NP1>[NP3][AdvP1.2][AdvP2] <VPvf1>[NP2][NP2ext] <".">	They always update the software in the company.	3
Decoded source-language structure	[NP3]<NP1>[AdvP1.2] <VPvf1>[NP2][NP2ext][AdvP2] <".">	In the company, they always carry out software updates quickly.	0
Example 2 – «Раньше ваша компания когда-либо обновляла программное обеспечение для переводчика?»			
Source-language structure	[AdvP1.3/2*]<NP1>[AdvP1.3/4] <VP3> [NP2][NP2ext]<’’?’’>	Did your company ever update the software for an interpreter?	3
Decoded source-language structure	<NP1>[AdvP1.3/4]<VP3>[NP2] [NP2ext][AdvP1.3/2*]<’’?’’>	Has your company ever updated the software for an interpreter before?	0

Table 4: Pairwise comparison of the BLEU score metrics for the idiom «To make a push in the development» (English) - «Сделать толчок в развитии» (Russian).

	Google	**PROMT**	**Yandex**	**Babylon**	**SYSTRAN**
Google	1	0.4	0.14	0.28	0.4
PROMT	0.4	1	0.57	0.57	0.4
Yandex	0.14	0.57	1	0.5	0.16
Babylon	0.28	0.57	0.5	1	0.4
SYSTRAN	1	0.4	0.16	0.4	1

The experiments of translating texts of various stylistic codes with help of the above-mentioned machine translation systems show that when such special situations emerge – i.e. the presence of idioms and terms in the source-language context – we observe differences in translations. If we take formalization language to represent the source-language phrase, it will be possible to identify such special cases.

Case 1. If the difference is observed on the segment <NP1><VP> [NP2], the special situation is assigned to the whole phrase (or sentence).

In order for the differences to be evaluated, let us perform a pairwise comparison of translations' quality with help of the BLEU score metrics (see an example in the Table 4). To specify the situations, we can use the frequency of vocabulary use in the observed text. Idiomatic expressions are generally based on common vocabulary, which means that the frequency of a notional verb use should be above the average value.

Case 2. If the difference is observed on the segment [NP2][NP2ext][AdvP3][AdvP2][AdvP1.1], the special situation is assigned to a word-combination.

Case 3. If the difference is observed on the segments <NP1> or [NP2], the cause of inaccurate translation is a certain word or a stem-compound.

In the latter two cases the frequency of vocabulary use should be not higher than the average value (see the Figure 2).

Therefore, if we introduce 2 classes and check their relations we can identify the situations that require additional processing with help of a translator.

In general, the machine-aided translation algorithm can be demonstrated by the algorithm given in the Figure 3.

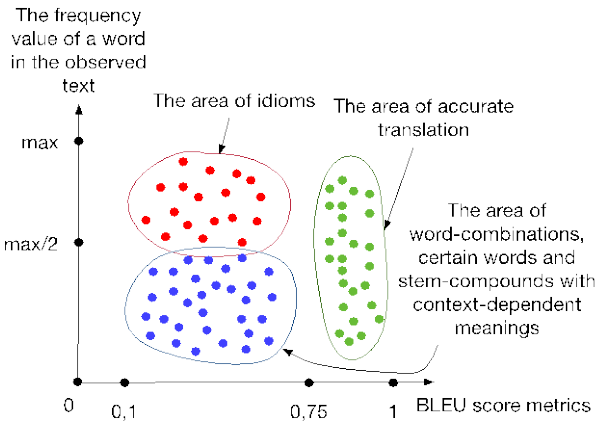

Figure 2: The examples of value distribution built on the analyzed set of values for idioms, word-combinations and certain words and stem-compounds.

Figure 3: The machine-aided translation algorithm.

1. The syntactic parsing of a source-language sentence.
2. The coding of a source-language syntactic structure in the formalization language [17].
3. The decoding of a source-language syntactic structure in accordance with the model of a target-language syntactic structure.
4. The rearrangement of words in a source-language phrase in accordance with the new syntactic structure.
5. The translation of a rearranged phrase with help of the existing machine translation systems.
6. The calculation of the BLEU score metrics under the cases 1-3 specified in the section 4.2. If the evaluation metrics can be assigned to one of the classes in the Figure 2, the corresponding phrase segment is marked with a special label.
7. The selection of a baseline translation (for instance, according to the experiment results of the present study Google delivers best translation outcome).

5 CONCLUSIONS

By manual processing, the suggested algorithm delivers the increase in translation quality by 0,1 of the BLEU score metrics. This evidence is a significant step forward as the existing machine translation systems are competing for basis points. Only in certain cases the difference in the translation quality comes up to tenths among the existing machine translation systems. Besides, the identification of situations that need close attention will considerably save translator's time. Today, the translation algorithm for big texts consists in the use of a machine translation system with further professional proofreading and post-editing.

The highlighted advantages make it clear that the suggested approach will work only with big texts, that will provide a sufficient amount of data for frequency calculation. More than that, the configuration of class memberships will be dependent on the knowledge domain (medicine, law, information technologies, programming, technics, etc.) of an analyzed text and on the language pair. These issues need background investigation. Efficient algorithm operation might need an introduction of a special non-traditional text classification [18], [19]. Besides, not all the phrases can be identified this way (for instance, the idiomatic expression «An old head on young shoulders» from the Table 2 was not identified as a special situation). This requires an additional analysis of the obtained translation result (on how it makes sense to a native speaker) and a possibility of introducing additional classification attributes.

REFERENCES

[1] B. Henisz-Dostert, R. R. Macdonald, and M. Zarechnak, Machine translation. The Hague ; New York: Mouton, 1979.

[2] P. Sojka, Ed., Text, speech and dialogue: 13th international conference, TSD 2010, Brno, Czech Republic, September 6-10, 2010: proceedings. Berlin ; New York: Springer, 2010.

[3] M. R. Costa-Jussà and M. Farrús, "Statistical machine translation enhancements through linguistic levels: A survey," ACM Comput. Surv., vol. 46, no. 3, pp. 1-28, Jan. 2014.

[4] M. R. Costa-jussà, A. Allauzen, L. Barrault, K. Cho, and H. Schwenk, "Introduction to the special issue on deep learning approaches for machine translation," Comput. Speech Lang., vol. 46, pp. 367-373, Nov. 2017.

[5] E. Hasler, A. de Gispert, F. Stahlberg, A. Waite, and B. Byrne, "Source sentence simplification for statistical machine translation," Comput. Speech Lang., vol. 45, pp. 221–235, Sep. 2017.

[6] F. J. Och and H. Ney, "A Systematic Comparison of Various Statistical Alignment Models," Comput. Linguist., vol. 29, no. 1, pp. 19–51, Mar. 2003.

[7] K. A. Papineni, S. Roukos, and R. T. Ward, "Maximum likelihood and discriminative training of direct translation models," in Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181), Seattle, WA, USA, 1998, vol. 1, pp. 189-192.

[8] F. J. Och, C. Tillmann, and H. Ney, "Improved Alignment Models for Statistical Machine Translation," vol. 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora.

[9] H. Alshawi, S. Bangalore, and S. Douglas, "Automatic acquisition of hierarchical transduction models for machine translation," in Proceedings of the 36th annual meeting on Association for Computational Linguistics -, Montreal, Quebec, Canada, 1998, vol. 1, p. 41.

[10] D. Wu, "Stochastic inversion transduction grammars and bilingual parsing of parallel corpora," Comput. Linguist., vol. 23, no. 3, pp. 377-403, Sep. 1997.

[11] J. Uszkoreit, J. M. Ponte, A. C. Popat, and M. Dubiner, "Large scale parallel document mining for machine translation," COLING 10 Proc. 23rd Int. Conf. Comput. Linguist., pp. 1101–1109, Aug. 2010.

[12] M. Aiken, K. Ghosh, J. Wee, and M. Vanjani, "An Evaluation of the Accuracy of Online Translation Systems," Commun. IIMA, vol. 09, no. 04, 2009.

[13] P. N. Astya et al., Proceeding, International Conference on Computing, Communication and Automation (ICCCA 2016): 29-30 April, 2016. 2016.

[14] Association for Computational Linguistics, P. Isabelle, and Association for Computational Linguistics, Eds., Proceedings of the conference, 40th annual meeting of the Association for Computational Linguistics: Philadelphia, [6 - 13] July 2002, University of Pennsylvania, Philadelphia, Pennsylvania. Hauptbd. ... San Francisco: Morgan Kaufmann, 2002.

[15] E. Sumita and H. Iida, "Heterogenous Computing for Example-Based Translation of Spoken Language," Proc. Sixth Int. Conf. Theor. Methodol. Issues Mach. Transl., pp. 273-286, 1995.

[16] A. V. Novikova and L. A. Mylnikov, "Problems of machine translation of business texts from Russian into English," Autom. Doc. Math. Linguist., vol. 51, no. 3, pp. 159-169, Jun. 2017.

[17] A. Novikova, "Direct Machine Translation and Formalization Issues of Language Structures and Their Matches by Automated Machine Translation for the Russian-English Language Pair," Proc. Int. Conf. Appl. Innov. IT, vol. 6, no. 1, p. 85-92., 2018.

[18] E. Stamatatos, N. Fakotakis, and G. Kokkinakis, "Automatic Text Categorization in Terms of Genre and Author," Comput. Linguist., vol. 26, no. 4, pp. 471-495, Dec. 2000.

[19] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," Comput. Linguist., vol. 37, no. 2, pp. 267-307, Jun. 2011.