

MACHINE TRANSLATION IN THE COURSE “COMPUTER TECHNOLOGIES IN LINGUISTICS” AT THE PHILOLOGICAL DEPARTMENT OF THE SAINT-PETERSBURG UNIVERSITY

Andrei Masevich, Victor Zakharov

Abstract: *Machine translation now left laboratories and became one of the practices of information service. A large number of free or partly free systems of machine translation became available in the net. Consequently, the task of their comparison and evaluation criteria arises. The paper describes the procedure of the estimation of the machine translation of text executed by different machine translation systems. The systems did translation into the Russian language of text (fragment “Communist manifesto” by Marx and Engels) from German, English and French. Students of the philological department of Saint-Petersburg State University systematized errors of translation, tried to determine the sources of errors, and considered the possibility (or impossibility) of their elimination.*

Keywords: *machine translation, evaluation, teaching*

ACM Classification Keywords: *1.2.7 Natural Language Processing*

Introduction

Machine translation (MT) now became one of the widely used services of the global network. The task of comparison and evaluation of numerous systems that appeared recently arises. The growing popularity of systems of machine translation systems, the constant improvement of them are factors which cause the need of including certain aspects of these technologies into the different training courses and first of all for the students of linguistics. In this paper we describe the approach to the teaching machine translation in at? the Philological Department of St. Petersburg University (the Chair of Mathematical Linguistics) within training course “Computer technologies in linguistics”.

To evaluate the machine translation one can use different criteria. From the? point of view of user, the output text should first semantically correspond to the source text and it should be coherent and understandable to target language audience [Helmreich, Farwell 1996: 49. The development of MT evaluation principles is both an important research task and the need of user facing the problem of choice of the system.

The Consortium Translate4eu developed a protocol of the translation evaluation from the user's point of view. The protocol is based on the set of tests (working package 4). The users are asked to evaluate the adequacy of the translation and some other features of systems. The users compare translations of sentences selected at random. Users have a chance to compare their own evaluation and complex summary of evaluations of a system. Result of the summary will be available to any user.

The quality of MT systems is in many respects determined by solution of linguistic problems. That is why one of the approaches to the evaluation of a systems - this is the examination of their linguistic special features.

We describe the general methodology of the evaluation of the productions of various systems of machine translation, which includes comparison of translated texts and statistical analysis of errors of translation. The students of the department act as experts who execute the evaluation. The paper also contains description of the practical training tasks.

The essence of the approach

We assume that when starting the course students have already certain knowledge of the linguistic disciplines and the theory and current practice of machine translation.

The practical task consists of identifying, classifying and statistical analysis of errors of several systems of machine translation and development of suggestions for improvement of the system's performance.

For the training tasks we use the following Internet resources: the translating systems Google, Prompt, AltaVista and the service <http://itranslate4.eu/>, which provide simultaneous translations by seven systems Bing, Google, Lingenio, Prompt, SkyCode, Systran, Trident, also the National Corpus of Russian Language, and Electronic dictionary of German (Digitales Wörterbuch der deutschen Sprache (DWDS))

Using the systems, the students translate separate sentences and small fragments of text from English, German, and French into Russian. For the comparison one uses the human translations, executed by professional interpreters and sometimes students make translations themselves. Students further classify the errors of machine translators try to understand their reasons and consider possibilities of the corrections.

Classification of errors

For the classification of errors a rough scheme was developed. We divide errors of translation into three large groups. One can further divide each group into subgroups.

1. Lexical and semantic errors

Wrong translation of words and word-groups, for example:

- Bill has been out of work for several months (Билл был без работы в течение нескольких месяцев) Законопроект был без работы в течение нескольких месяцев (Homonymy of words Bill – a draft of law and Bill – personal name)
- She was writing letters all afternoon (После полудня она все время писала письма) Она писала помечает буквами все время после обеда (Homonymy letters written messages and letters characters)
- It's so nice sitting here with you (Так хорошо сидеть здесь с вами) Это такое хорошее заседание здесь с вами. (The gerund "sitting" should be translated with infinitive form)

2. Grammar Errors

I am going to Moscow next month (Я собираюсь в Москву следующем месяце)

Я иду к Москва следующий месяц – coherence error

To know her is to like her Знать ее полюбить она – coherence error

When he lived in London he went to the theatre once a week (Когда он жил в Лондоне, он ходил в театр каждую неделю) Когда он жил в Лондоне, он пошел в театр один раз в неделю – Predicate, expressing repeatable action, is translated as predicate of once occurring action

The earth moves round the sun (Земля движется вокруг Солнца) Земля двигает вокруг солнца. The relation agent –subject is rendered wrongly.

The sun shines during the day (Солнце светит днем) Блески солнца во время дня. The predicate is translated as subject

3. The other errors (generating of pseudo words, lack of translation of words and collocations etc.)

It was a rule for men and women to sit apart Было правилом для людей и женщин, котор нужно сидеть врозь

Studying all night made him ill Изучать всю ночу сделал его больно

When he lived in London he went to the theatre once a week (Когда он жил в Лондоне, он ходил в театр каждую неделю) Когда он жил в лондоне он пошел к театру once a week

Each type of error has a corresponding code 1, 2 or 3. The absence of error is designated by the 0

Students assign one or more codes to each translated sentence. The codes should be

inserted in brackets to the boxes of the table as shown bellow (see tabl.1).

Translation of sentences

To illustrate this task we selected 9 sentences from the book [Hornby, 1975] (see Table 1).

Table 1. Results of translation of sentences executed by 4 systems

English sentence	Prompt	Trident	Systran	WordLINGVO
Studying all night made him ill	Изучение всю ночь сделало его плохо (2)	Изучение всей ночи сделало его больным (2)	Изучать всю ночь сделал его больно (2,3)	Изучать всю ночь сделал его больно (2,3)
It was a rule for men and women to sit apart	Это было правило для мужчин и женщин сидеть обособленно(2)	Это было правило, чтобы мужчины и женщины сидели отдельно (0)	Было правилом для людей и женщин, котор нужно сидеть врозь (1, 2, 3)	Было правилом для людей и женщин, котор нужно сидеть врозь (1, 2, 3)
When he lived in London he went to the theatre once a week	Когда он жил в Лондоне, он пошел в театр один раз в неделю (2)	Когда он жил в Лондоне, он ходил в театр раз в неделю (0)	Когда он жил в Лондоне он шел к театру раз неделя (2)	Когда он жил в лондоне он пошел к театру once a week (2,3)
She was writing letters all afternoon	Весь день она писала письма (1)	Она писала помечает буквами все время после обеда (1,2)	Она писала письмам все после полудня (2)	Она была письмами сочинительства все после полудня (2,3)
To know her is to like her	Знать ее означает понравиться она (2)	Чтобы знать ее - понравиться ее (2)	Знать ее полюбить она 2)	Знать ее к как ей (1,2)
Я иду в Москву в следующем месяце	Я иду в Москву в следующем месяце (2)	Я собираюсь в Москву в следующем месяце (0)	Я иду к Москве следующий месяц (2)	Я иду к Moscow следующий месяц (2,3)
The sun shines during the day	В течение дня светит солнце (0)	Солнце сияет в течение дня (1)	Блески солнца во время дня (2)	Солнце светит во время дня (0)

Table 1-continued. Results of translation of sentences executed by 4 systems

English sentence	Prompt	Trident	Systran	WordLINGVO
Bill has been out of work for several months	Билл был безработным в течение нескольких месяцев (0)	Билл был безработным несколькими месяцами (2)	Билл из работы на несколько месяцев (2)	Билл из работы на несколько месяцев (2)
It's so nice sitting here with you	Это - так хорошее заседание здесь с Вами (1,2)	Это - такое хорошее заседание здесь с вами (1,2)	Оно настолько славный сидеть здесь с вами (1,2)	Будет настолько славным усаживанием здесь с вами(1, 2)

Table 2. Evaluation of translation according to types and number of errors

Type of error	Code of error	Prompt, number of errors	Trident, number of errors	Systran, number of errors	World - LINGO, number of errors
Lexical and semantic errors	1	2	3	2	3
Grammar errors	2	6	5	9	8
Other	3	0	0	1	5
General number of errors		8	8	12	16
Number of sentences translated without errors	0	2	3	2	1

For the examples, we selected 9 English sentences. But even a small amount of material allows to come to certain conclusions, like following:

Predominated type of errors are grammar errors.

Since the Trident system produce the more errorless translations than any other one can suggest that the translation of this system are of the best quality.

The amount of material could be enlarged and for its processing one can apply statistical methods.

Further, the students summarize the results obtaining in several experiments.

In addition to evaluation of machine translators, the students are asked to identify the grammar and syntax constructions especially difficult for machine translation from English into Russian. To perform this it is necessary to analyze sentences with most serious errors of translation.

Thus, judging by our example, one can suggest that one of the most complex for the machine translation from English into Russian is gerund constructions. The students are asked to find corresponding examples to confirm or to refute the suggestion as follows (Table 3).

Table 3. Translations of sentences with gerund

Sentences	Prompt	Trident	Systran	WordLINGVO
She enjoys going to concerts	Она любит идти в концерты	Она наслаждаются собрался в концерты	Она наслаждается идя концертами	Она наслаждается пойти к согласиям
She can't bear seeing animals treated cruelly	Она не может родить видящих животных, которых рассматривают безжалостно	Она не может носить виденье, что животные лечили жестоко	Она не может принести увидеть животных обработанных жестокосердно	Она не может принести увидеть животных обработанных жестокосердно
Don't start borrowing money	Не начинайте занимать деньги	Не запускайте одолжение денег	Не начните одолжить деньгам	Не начните одолжить деньг

The analysis of the examples seems to confirm the suggestion. It also can help the regularities of the some errors appearance.

The translation of the fragment of the text

As another training task we take a translation by translating systems of the fragment from the Manifesto of Communist Party by K. Marx and F. Engels from German, English and French. The choice is because the versions of the text are available in Internet in all three languages. There is also Russian translation (we use the Russian version published by the Marx-Engels-Lenin Institute in 1955).

In our paper we limit ourselves with analysis of some errors in translating from German into Russian

Source text (German)

Ein Gespenst geht um in Europa - das Gespenst des Kommunismus. Alle Mächte des alten Europa haben sich zu einer heiligen Hetzjagd gegen dies Gespenst verbündet, der Papst und der Zar, Metternich und Guizot, französische Radikale und deutsche Polizisten.

Wo ist die Oppositionspartei, die nicht von ihren regierenden Gegnern als kommunistisch verschrien worden wäre, wo die Oppositionspartei, die den fortgeschritteneren Oppositionsleuten sowohl wie ihren reaktionären Gegnern den brandmarkenden Vorwurf des Kommunismus nicht zurückgeschleudert hätte?

Translation into Russian by the Marx-Engels-Lenin institute, 1955 г.):

Призрак бродит по Европе - призрак коммунизма. Все силы старой Европы объединились для священной травли этого призрака: папа и царь, Меттерних и Гизо, французские радикалы и немецкие полицейские.

Где та оппозиционная партия, которую ее противники, стоящие у власти, не ославили бы коммунистической? Где та оппозиционная партия, которая в свою очередь не бросала бы клеймящего обвинения в коммунизме как более передовым представителям оппозиции, так и своим реакционным противникам?

Translation into English

A specter is haunting Europe—the specter of Communism. All the powers of old Europe have entered into a holy alliance to exorcise this specter; Pope and Czar, Metternich and Guizot, French radicals and German police spies.

Where is the party in opposition that has not been decried as Communistic by its opponents in power? Where the opposition that has not hurled back the branding reproach of Communism, against the more advanced opposition parties, as well as against its reactionary adversaries?

Table 4. Examples of translating the fragment

Systems	Translation into Russian
Prompt German - Russian	<p>Привидение (1) идет вокруг в Европе - привидение коммунизма. Все власти (1) старой Европы объединились к святой (1) травле против этого привидения, папа и царь, Metternich и Guizot (3), французские радикалы и немецкие полицейские.</p> <p>Где оппозиционная партия, которая не была бы поносившей (2) ее правящими противниками как коммунистическая, где оппозиционная партия, у которой не было бы клеймщего упрека коммунизма zurückgeschleudert (3) более прогрессивным оппозиционным людям как ее реакционным противникам</p>
Trident German - Russian	<p>Привидение идет, чтобы в Европе - привидение коммунизма. Все власти старой Европе союзные себя к святому Hetzjagd против этого привидения, Papst и Zar, Metternich и Guizot, французский радикальный и немецкий полицейский.</p> <p>Где оппозиционная партия есть, которая не стала бы обесславлена их владетельными противниками как коммунистический, где оппозиционная партия, которая не назад бросила бы fortgeschritteneren оппозиционным людям как и их реакционным противникам клеймющий упрек коммунизма</p>
Prompt German - English, Systran English- Russian	<p>Привидение идет вокруг в Европе - привидение коммунизма. Вся власть старой Европы формировали союзничество к святейшей охоте против этого привидения, Папы и царя, Metternich и Guizot, французских радикалов и немецких полицейские.</p> <p>Где оппозиционная партия которая не была бы обозначена ее господствующими противниками как коммунист, где оппозиционная партия которая не имела бы упрек zurückgeschleudert коммунизма к людям противовключения fortgeschritteneren как к ее охотнорядческим противникам</p>
Lingatec German - English, Trident English- Russian	<p>Призрак циркулирует в Европе - призрак коммунизма. Союзнический, римский папа и Царь, Metternich и Guizot, имеют себя для священной охоты с гончими против этой ghost Европы все полномочия века, Французских радикалов и Немецких полисменов.</p> <p>Где оппозиционная партия, оппоненты, не управляющие над этим hers, получены были печально известными как коммунист, где оппозиционная партия, которая не сможет иметь обеих обратных сторон, бросила клеймющий упрек коммунизма для более передовых оппозиционных людей подобно ее реакционным оппонентам</p>

Table 4 continued. Examples of translating the fragment

Systems	Translation into Russian
Lingemo German - English, Promt English- Russian	<p>Призрак циркулирует в Европе, призраке коммунизма. Все полномочия старого человека Европа соединились с собой для священной стремительной охоты против этого призрака, Папы Римского и Царя, Metternich и Guizot, французских радикалов и немецких полицейских.</p> <p>Где оппозиционная партия, неуправление противников по этому ее было столь коммунистическим печально известно оппозиционная партия, которые оба не бросились бы назад, упрек объявления коммунизма для более продвинутых людей оппозиции как ее реакционные противники будет иметь, где</p>
Lingemo German - English, Trident English - Russian	<p>Призрак циркулирует в Европе, призрак коммунизма. Все полномочия старой Европы особы имеют вступил себя в союз для священного разбивания, охотничьего против этого призрака, римского папы и Царя, Metternich и Guizot, Французские радикалы и Немецкие полисмены.</p> <p>Где оппозиционная партия, оппоненты, неуправляющие над этим hers, были как коммунист общеизвестно оппозиционная партия, которая не сможет бросить оба назад клеймящий упрек коммунизма для более передовых оппозиционных людей подобно ее реакционным оппонентам имел бы, где</p>
Google German - English, Trident English - Russian (2010)	<p>А спектр является навязчивый Европе - призрак коммунизма. Все державы старой Европы взяли на себя обязательства по Святым Духом охоту против этого альянса, Папы Римского, и царь, Меттерних и Гизо, французские радикалы и немецкие полицейские.</p> <p>Где оппозиционной партии, а не его оппонентов, как правящая Коммунистическая verschrien был обнаружен, когда оппозиционная партия, оппозиция более продвинутых людей, как своих реакционных противников, таких, как марка упрек коммунизма zurückgeschleudert бы и нет</p>
Google German - Russian (2012)	<p>Призрак бродит по Европе - призрак коммунизма. Все силы старой Европы объединились для священной травли этого призрака: папа и царь, Меттерних и Гизо, французские радикалы и немецкие полицейские.</p> <p>Где та оппозиционная партия, партия, которая не была осуждена своих противников как коммунистической правящей, где оппозиционная партия, которая не бросала бы более продвинутой оппозиции, а также своим реакционным противникам брендинга обвинения в коммунизме?</p>

Semantic problems. We shall show one of semantic errors. Let us consider semantics of the Russian word “Призрак” which translated in versions 1 and 2 as “привидение”

Students are asked to look at the very detailed description of the lexical unit “Gespenst” in the electronic dictionary of German language (Digitales Wörterbuch der deutschen

Sprache (DWDS) <http://www.dwds.de/?qu=Gespent> (Fig.1) and to make a search for the examples of using words «призрак» and «привидение» in National Corpus of Russian Language.

The screenshot shows the DWDS website interface. At the top, there's a search bar containing 'Gespent' and a 'Suche' button. Below the search bar, there are several panels:

- DWDS-Wörterbuch:** Shows the word 'Gespent' with its grammatical information: 'neutr., -s/-es; -er', 'Ausssprache: ►', and 'Spukgestalt, Geist'. It also lists 'drohende Gefahr übertragen, gehoben' and provides a link to 'dwsdw-0.9 Eintrag | Zusammensetzungen | Belege | Beispiele'.
- Etymologisches Wörterbuch des Deutschen (nach Pfeffer):** Provides a detailed etymology: 'Gespent, ferner: gespenstern, gespenstig, gespenstisch. Gespent n. 'Spukgestalt, Geist', ahd. *gispēstn.* 'Anlockung, Verführung, Eingebung' (9. Jh.), mhd. *gespēst, gespenstf., gespenste n., asächs. gispēstn., mnd. gespēstf.* stehen wie ahd. *spanstf.* (9. Jh.), mhd. *spēstf. n.* 'Verlockung, Verführung' als Verbalabstrakta zu dem unter *spannen* (s. d.) angeführten, im Nhd. ausgestorbenen starken Verb ahd. *spanan*, mhd. *spanen* 'locken, reizen', asächs. aengl. *spanan* (s. auch ab-, widerspenstig). *Gespent* ist daher eigentl. das, 'was jmdn. wegzieht, weglockt', ein ('beufliches) Trugbild'. – *gespenstern* Vb. 'wie ein Gespent umhergehen, spuken' (19. Jh.). *gespenstig* Adj. 'gespenstähnlich, unheimlich' (19. Jh.), ahd. *spēstg* 'lockend, verführerisch' (um 1000), mhd. *gespenstec* 'verführerisch, zauberisch', auch *gespenstisch* Adj. (18. Jh.). etymw-1.0.6
- OpenThesaurus:** Shows synonym groups for 'Gespent', including 'Geist, Gespenst, Spuk' and a link to 'OpenThesaurus 2011-09-09 Openthesaurus Webseite'.
- DWDS-Wortprofil 2010:** Displays a word cloud with terms like 'Arbeitslosigkeit', 'Geister', 'Kommunismus', 'Vergangenheit', etc.
- DWDS-Kernkorpus (eingeschränkte Version):** Shows 63 hits with example sentences, such as '...Eine Abrechnung. **Gespenter** am Kamin. Ein Kammerp...

Fig. 1. Description of the word «Gespent» in DWDS

The words «призрак» and «привидение» in Russian are synonyms. However, the word «призрак» is polysemic, it can mean both hope, threat, apprehension etc. and the host, mystical creature. The same is true for the German word “Gespent”.

We selected several example of sentences with word «призрак» with different meanings from the National Corpus of Russian Language.

*На горизонте появился **призрак** войны с Турцией, вырисовывались контуры большой морской войны. [Е. В. Тарле. Экспедиция адмирала Сенявина в Средиземное море (1805-1807) (1954)]*

*Она дала мне хоть иллюзию, хоть **призрак** любви, и это истинно царский, неоплатный подарок... [А. И. Куприн. Телеграфист (1911)]*

Второе: суровая жизнь закалила Диму и Геню до состояния стали, а Егорушка рос

мягкий, ласковый, как бы безвольный, и уже виделся за горами **призрак** его возможного будущего, будущего всех мужчин Диминой семьи, и история зарождения Егорушки сулила также и со стороны матери легкомыслие и случайные связи. [Людмила Петрушевская. Два бога (1998-1999)]

То был **призрак** короля Арнульфа Второго, изменнически, из засады убитого в 1527 году, на пятнадцатом году жизни. [Ф. К. Сологуб. Королева Ортруда (1909)]

And also three phrases from the Corpus with the word “Привидение”

Охваченный состраданием, я хочу подойти к ней, но нимфа быстро поднимается: она только взглянула на меня, — и убегает, а на лице её ужас, точно она увидела **привидение**. [И. Ф. Анненский. Вторая книга отражений (1909)]

И **привидение**, пройдя в отверстие трельяжа, беспрепятственно вступило на веранду. [М. А. Булгаков. Мастер и Маргарита, часть 1 (1929-1940)]

Тут все увидели, что это— никакое не **привидение**, а Иван Николаевич Бездомный — известнейший поэт. [М. А. Булгаков. Мастер и Маргарита, часть 1 (1929-1940)]

Polysemy of the word “призрак” is evident that the examples above (ex.1-4). The word “привидение” is monosemic (ex.5-7) and its single meaning coincide with meaning of the word “призрак”, in which it is used in the ex.4

The word “Gespenst” in the source text joins metaphorically both meanings. On one side it is threat, apprehension. On the other side it is a subject of “heilige Hetzjagd”, i.e. of chase, persecution. In the literary translation of the text this metaphor is fully kept. But how should one set a task of this kind to the system of machine translation? Possibly, though we are not sure, when the system have to make a choice between two synonyms it should prefer the more general one.

Table 5. The example of Grammar error (from the version 1)

Wo ist die Oppositionspartei, die nicht von ihren regierenden Gegnern als kommunistisch verschrien worden wäre...	Где оппозиционная партия, которая не была бы <u>поносившей</u> ее правящими противниками как коммунистическая...
---	--

Should the participle «поносившей» (abusing, denouncing) be in the passive voice like this «Где оппозиционная партия, которая не была бы поносимой ее правящими противниками, как коммунистическая». However, it does not sound euphonic, so it is might be better to replace it with synonym «ославленной».

Errors of the third group are obvious. The personal names and the complex German words as zurückgeschleudert or fortgeschritteneren are frequently remaining not

translated.

Other particular qualities. In the version 7 the word “Gespenst” is translated as “спектр”. Neither in Russian nor in German (Spektrum) the word is used in the meaning of «призрак». The cause is that the system used an intermediate language. The text was first translated from German into English and then from English into Russian. The system translated English word “specter” as “спектр”, which has completely different meaning. Note, that the Google translator executed the version 7 in 2010.

The same system executed The version 8 was done in 2012. The first paragraph of this version copies literally Russian literary translation. The second paragraph is the usual machine translation with typical errors

Conclusion

In our paper, we could not give detailed analysis of all errors. We just had a task describe method of the evaluation of the system using the rough scheme of the classifying errors.

In conclusion, we would like to note that one can use our method could for other goals. If we refine the classification of errors than with the aid of students-linguists, we could possibly create a database of errors for improving systems of machine translation.

Bibliography

[Hornby, 1975] A.S.Hornby. Guide to patterns and usage in English London: Oxford University Press, 1975.

[Helmreich, Farwell, 1996] Helmreich St., Farwell D. Translation Differences and Pragmatic-Based MT //Expanding MT Horizons. Proc. of the Sec. Conf. Of the Assoc. for MT in the Americas. Montreal, Canada, 1996. Pp. 66-75.

Authors' Information



Andrei Masevich – Assistant professor. Saint-Petersburg State University of Culture and Arts; e-mail: andmasev@mail.ru

Major Fields of Scientific Research: Computational Linguistics, Library information systems



Victor Zakharov – Associate Professor, Saint-Petersburg State University, Universitetskaya emb., 11, Saint-Petersburg 199034 Russia ; e-mail: vz1311@yandex.ru

Major Fields of Scientific Research: Natural Language Processing