Title: Machine translation : can it assist in professional translation of contracts?

Author: Joanna Sycz-Opoń

MACHINE TRANSLATION – CAN IT ASSIST IN PROFESSIONAL TRANSLATION OF CONTRACTS?

Joanna SYCZ-OPOŃ, M.A.
University of Silesia, Faculty of Philology, Institute of English
ul. Grota-Roweckiego 5, 41-200 Sosnowiec, Poland
joanna.sycz@us.edu.pl

Abstract: The aim of this research project is to verify whether machine translation (MT) technology can be utilized in the process of professional translation. The genre to be tested in this study is a legal contract. It is a non-literary text, with a high rate of repeatable phrases, predictable lexis, culture-bound terms and syntactically complex sentences (Šarčević 2000, Berezowski 2008). The subject of this study is MT software available on the market that supports the English-Polish language pair: Google MT and Microsoft MT. During the experiment, the process of post-editing of MT raw output was recorded and then analysed in order to retrieve the following data:
(i)   number of errors in MT raw output,
(ii)   types of errors (syntactic, grammatical, lexical) and their frequency,
(iii)   degree of fidelity to the original text (frequency of meaning omissions and meaning distortions),
(iv)   time devoted to the editing process of the MT raw output.
The research results should help translators make an informed decision whether they would like to invite MT into their work environment.
Key words: machine translation, professional translation, legal translation, legal contract.

TŁUMACZENIE MASZYNOWE – CZY MOŻE WSPOMÓC PROFESJONALNY PRZEKŁAD UMÓW?

Abstrakt: Niniejszy projekt badawczy ma na celu wykazanie czy jakość tłumaczenia maszynowego jest na tyle dobra, by mogło być ono wykorzystywane podczas pracy profesjonalnego tłumacza prawniczego. Podczas badania analizie poddane zostały umowy – teksty użytkowe charakteryzujące się wysoką powtarzalnością wyrażeń, zwrotów i terminów, złożoną składnią oraz nieprzystawalnością terminologiczną (Šarčević 2000, Berezowski 2008). Przyjęta metoda badawcza polegała na nagraniu procesu tłumaczenia przy zastosowaniu narzędzi Google MT oraz Microsoft MT. Badanie umożliwiło wydobycie informacji na temat użyteczności tłumaczenia maszynowego poprzez określenie:
(i)   rodzaju błędów występujących w tekście wygenerowanym przez tłumacza maszynowego,
(ii)   częstotliwości występowania błędów,
(iii)   zgodności merytorycznej z treścią oryginału (liczba pominięć oraz zniekształceń),
(iv)   czasu poświęconego na edycję tekstu wygenerowanego przez tłumacza maszynowego.
Wyniki badania powinny pomóc tłumaczom w podjęciu świadomej decyzji czy chcieliby włączyć tłumaczenie maszynowe do swojego warsztatu pracy.
Słowa kluczowe: tłumaczenie maszynowe, tłumaczenie profesjonalne, tłumaczenie prawnicze, umowa

Introduction

On 26th April 2012 Google researcher Franz Och (2012) announced on the Google official blog that Google MT (machine translation) was at that moment used monthly by 200 million people. He continued quoting even more impressive figures:

> *In a given day we translate roughly as much text as you'd find in 1 million books. To put it another way: what all the professional human translators in the world produce in a year, our system translates in roughly a single day.*

The numbers speak for themselves. Machine translation is gaining popularity at impressive pace, not only among laymen who need it for basic communication, but also in a professional sphere. MT solutions are utilized by large companies (e.g., Xerox, Ford, General Motors) and institutions (e.g., European Commission, Pan American Health Organization), which without MT's assistance would not manage to translate large volumes of text in a short time (Hutchins 2007). These companies have throughout the years understood the limitations of automated translation and no longer expect perfection. They have also learnt how to prepare MT-friendly input texts (characterised by controlled terminology and restricted syntax), which significantly influences quality of MT output (Hutchins 2010). A change of attitude towards MT solutions could be observed also among translators. The studies show that automated translation is slowly, but systematically gaining translators' approval (Fulford 2002, Fulford and Granell-Zafra 2004). It should be also mentioned that machine translation has been recently added to many CAT tools. The number of MT enthusiasts is still small, but it seems that we are now at the breaking point, where automated translation, which has been for decades taken with a pinch of salt, is now beginning to be seriously considered as a helpful tool.

A translation assignment handed over to a client is expected to be faultless. This is, as of today, still unattainable for machines (Graham et al. 2014) . Thus, not soon will MT substitute translators (if ever), but it can provide them with a raw material to work on. However, if MT is to find any application in the translator's work, the process of post-edition needs to be significantly shorter than translation from scratch. This is what this study tested. The aim of this research is to verify usefulness of MT by measuring the effort required to post-edit the MT raw output. The research results should help translators make an informed decision whether they would like to invite MT into their work environment.

Scope of the Study

The focus of this study is utility of automated translation in the work of professional translators. Although there are many studies devoted to MT performance, majority of them were designed with a non-professional user in mind. MT solutions are predominantly used nowadays for *assimilation*, i.e. the purpose of deciphering the meaning of a foreign language text, or for basic *communication* (Hutchins 2003), while translators want MT to facilitate production of a text (*dissemination purpose*). Both groups have different expectations towards the tool and different levels of expertise. Non-professionals want the text to be understandable. Thus, they assess MT utility taking into consideration the errors that

distort the original meaning. What is crucial for translators is the amount of work they need to put into erasing all errors in MT output. Thus, the aspect of the text that the translator is most concerned with is the amount of errors and the time devoted to their correction – the so called *post-editing effort*. This is what this study examined.

It was decided that the subject of this research project had to be narrowed down to translation of (1) agreements (2) from Polish into English (3) with the use of two MT tools: Google MT and Microsoft MT. Such limitation of the scope of the study was necessary due to the fact that quality of MT output is highly dependent on individual features of a text.

(i)   First of all, MT performance differs depending on a genre. Each genre is characterised by many distinct features, e.g., syntactic structure, specific phraseology, lexis, text density, or degree of repeatability. All of these features have impact on MT performance. The genres characterised by simplified syntax, predictable terminology and high rate of repetitions are more MT friendly. On the contrary, genres with long, complex sentences and varied vocabulary are hardly machine-translatable (Kit and Wong 2008, Zervaki 2002 in Seljan, Brkić, and Kučiš 2011). Moreover, since statistical MT works on the basis of bilingual texts stored in its memory, the popularity of a given genre also plays an important role. If a particular genre is well-represented in the realm of the Internet, it is also present in the corpora used by MT software. That increases the chances of it being decently translated by the machine. The genre to be tested in this study is an agreement. It is a non-literary text, with a high rate of repeatable phrases, predictable lexis, culture-bound terms and complex syntax (Šarčević 2000, Berezowski 2008). This is a fundamental document regulating all kinds of business transactions, thus it could be assumed that it is well-represented in MT corpora.

(ii)   The second factor determining MT performance is a language pair, namely, the similarity between the languages and their popularity. The greater the syntactic gap between the languages the worse the MT outcome, especially in case of rule-based MT systems. Polish belongs to a West Slavic language family while English is a West Germanic language. That results in multiple linguistic differences between the two. Polish is an inflected language, equipped with noun cases (singular and plural), verb conjugation, perfective and imperfective aspects, and masculine, feminine and neuter genders. Because of the declension, Polish has relatively free word order in a sentence and subject pronouns are often omitted. It does not make use of articles. English, on the other hand, is generally an uninflected language, yet it is abundant with articles. It has a relatively fixed word order, and generally does not allow omission of personal pronouns.

As for language popularity, English is naturally one of the most widely spoken languages in the world. Polish, on the contrary, is in the minority – used mainly by its native speakers. On the list compiled by Hutchings (2008: unpaginated) it is ranked as follows:

> *The most frequent pairs (for online MT services and apparently for PC systems) are English/Spanish and English/Japanese. These are followed by (in no parti-cular order) English/French, English/German, English/Italian, English/Chinese, English/Korean, and*

French/German. Other European languages such as Czech, Polish, Bulgarian, Romanian,
Latvian, Lithuanian, Estonian, and Finnish are more rarely found on the market.

(iii)   The third factor that needs to be borne in mind is the MT software. Most of MT
tools utilize their own technological solutions, drawing on rule-based (RBMT)
and/or corpus-based approach. In rule-based system the machine generates
translation on the basis of multiple sophisticated linguistic rules and dictionaries
in its memory (Hutchins 2007). In corpus-based system, the machine translates on
the basis of a large corpus consisting of ready-made translations. This is still
a new approach to MT, but many companies have switched to it as more
promising for future development. Within this basic division there exist multiple
subcategories of machine translation. Rule-based technology can be subdivided
into: *direct method (dictionary-based MT), transfer RBMT systems, interlingual*
RBMT systems. The corpus-based MT incorporates, among others, solutions such
as: *statistical machine translation (SMT)* and *example based machine translation*
(Bijimol and Abraham 2014). Each software works on the basis of its own
individual MT engine that uses one of the abovementioned technologies or their
combination. Both Google and Microsoft declare that their MT engines work on
the basis of corpus-based solutions. It is not material to discuss the technological
details here (more information on this topic can be found in Bijimol and Abraham
2014). Yet, it needs to be stressed that the results of the experiment conducted on
particular MT software should not be generalized to other MT tools.

Methodology

Evaluation of MT performance is not an easy task, because it requires assessment of
many parameters, some of them difficult to measure objectively in mathematical counts.
Therefore, various approaches to MT evaluation have been developed so far, ranging
from purely automated (e.g., BLEU, NIST and METEOR evaluation metrics), through
semi-automatic (e.g., HTER) to traditional human assessment. Automatic methods
assess MT quality by comparing MT output with the available translations of the same
text produced by humans (reference texts), using language independent statistical
metrics (Hutchins 2007). The semi-automatic method – HTER – proposes a different
approach. It does not make comparisons between MT output and reference translations
done by humans. Instead, it measures the so called *edit distance* between MT raw output
and its post-edited version performed by a human translator. *Edit distance* is the amount
of editing required to transform MT raw translation into a text of publishable standard.
The evaluation is done by means of automatic count of edits during the post-editing
process. Then, special software automatically compares MT raw output to its post-edited
version. The higher the number of edits, the worse quality is the raw output produced by
the machine (Snover et al. 2006).

Automatic and semi-automatic evaluation methods are fast and low-cost. Yet,
they are burdened with several weaknesses. Every translation is an act of creative
writing; there is not one true version to which MT raw output might be compared.
Therefore, what a machine automatically counts as an error might as well be an
alternative correct translation. Moreover, automatic count of errors or corrections does
not reflect the actual cognitive effort involved in post-editing. This claim was confirmed

in the study by Koponen et al. (2012, 12), which showed that "translator's perception of post-editing effort, as indicated by scores in 1-5, does not always correlate well with edit distance metrics such as HTER. In other words, sentences scored as requiring significant post-editing sometimes involve very few edits, and vice-versa." Finally, the raw data generated in such tests are abstract. To properly understand the results of MT automatic tests, a translator would have to be aware of the amount of corrections or keystrokes made during traditional translation. Taking into account the above arguments, it was decided that automatic evaluation methods do not serve the purposes of this research. Instead, a qualitative approach was applied, namely, the task-based human assessment.

Four participants took part in the experiment. They were graduates from the University of Silesia with one to six years' experience as translators. The participants were asked to translate one of the texts in exactly the same manner they would normally do it, but with the assistance of a selected MT solution (which in practice meant post-editing the output produced by MT tools). Each participant translated two different texts with the use of two different MT solutions. The participants were asked to make the minimum number of changes to MT raw output, according to their own judgment. The experiment was recorded via the screen-capture recording tool *Camtasia Studio*. Then, the recordings were played back and analyzed by human researchers to obtain the required data. The participants were asked to note down the time when they started and finished their translation. These data were used to establish the total time of translation. Moreover, the participants were asked to pause the recording every time they consulted sources, so that the recording did not include the time devoted to consultation of sources. The recording registered only the post-editing process.

In order to create the experiment conditions that resemble natural work environment of a translator, it was necessary to introduce CAT tools into the experiment setting. Wordfast Anywhere was used as a platform to test Microsoft MT, while the performance of Google MT was tested in Google Translator Toolkit – an internet service addressed to translators recently launched by Google.

The most important datum that the study aimed to obtain was the time of post-editing. This is the simplest and most visible indicator of MT quality, since as Koponen et al. (2012) aptly noticed, the shorter the post-editing time, the lower the number of errors and corrections. On top of that, the aim of the study was to reveal common errors appearing in MT output. The errors were counted and classified. The results of the experiment are presented in sections 4 and 5 of this article.

As far as research material is concerned, the main criterion for text selection was its length. For the sake of statistics, it was decided that each text should constitute approximately one translation page (1600-2000 characters) or its multiple (3200-4000 characters), and for the sake of authenticity of the translation task – a text should be an entire document. The texts selected for the experiment are as follows: *umowa kupna-sprzedaży* [*sale agreement*], *umowa o poufności* [*confidentiality agreement*], *umowa najmu* [*lease agreement*], *umowa o dzieło* [*contract for a specific task*]. Due to data protection, no authentic contracts were used. Instead, it was decided to use templates available online, which were filled in with fictional data. In order to ascertain the level of the texts' syntactic complexity, the average sentence length was established (number of words per sentence). Moreover, the readability of the texts was verified with the tool available on the website *logios.pl*. This is a Gunning's FOG index

adjusted to the properties of the Polish language, designed by Polish linguists. The parts of the text that do not constitute grammatical sentences were not taken into account (headings and parties' signatures). The properties of the four texts constituting research material are provided in Table 1 below.

Table 1. Properties of the research material used in the experiment.

Type of agreement	Language	Characters with spaces	No. of sentences	The average sentence length	Readability index (logios.pl)
Umowa kupna-sprzedaży [sale agreement]	Polish	1979 (1 translation page)	17	14.2 words per sentence	FOG index: 13-17 years of education; difficult (higher education level)
Umowa o poufności [confidentiality agreement]	Polish	3600 (2 translation pages)	16	27.6 words per sentence	FOG index: 13-17 years of education; difficult (higher education level)
Umowa najmu [lease agreement]	Polish	3191 (2 translation pages)	19	21.5 words per sentence	FOG index: 13-17 years of education; difficult (higher education level)
Umowa o dzieło [Contract for a specific task]	Polish	3214 (2 translation pages)	22	18.4 words per sentence	FOG index: 13-17 years of education; difficult (higher education level)

General Results

The results of the experiment are shown in Tables 2 and 3. The tables present data pertaining to the general performance of tested MT software, such as: time of post-editing, time devoted to consultation of sources, time of translation, no. of sentences that required editing, no. of faultless sentences and nonsense sentences, as well as no. of sentences translated from scratch by the participants. The results are presented separately for each of the four texts. *Total time of translation* is understood as the time devoted to post-editing plus consultation of sources. *Time devoted to post-editing* excludes the time devoted to consultation of sources. *Nonsense sentences* were classified as such subjectively by the researcher during the analysis of the errors in MT output. There were also several instances during the experiment when the participant decided to delete the whole sentence produced by MT and translate it by himself. Such situations are presented in the category: *sentences translated from scratch.* The lengths of the texts (i.e., no. of characters and sentences) are also presented in the tables for easier comparison of the results.

Table 2. General performance of Google MT.

Researched data	*Umowa kupna-sprzedaży [sale agreement]*	*Umowa o poufności [confidentiality agreement]*	*Umowa najmu [lease agreement]*	*Umowa o dzieło [Contract for a specific task]*
Length of the text	1979 characters (17 sentences)	3600 characters (16 sentences)	3191 characters (19 sentences)	3214 characters (22 sentences)
Total time of translation	36 min. 20 sec.	43 min. 20 sec.	42 min. 10 sec.	55 min. 35 sec.
Time devoted to post-editing	28 min. 20 sec.	39 min. 20 sec.	39 min. 40 sec.	49 min. 50 sec.
Time devoted to sources consultation	8 min.	4 min.	2 min. 30 sec.	5 min. 45 sec.
No. of sentences requiring editing	14 (82.5%)	14 (87.5%)	17 (89.5%)	22 (100%)
No. of sentences not requiring editing (100% accuracy)	3 (17.5%)	2 (12.5%)	2 (10.5%)	0 (0%)
Nonsense sentences	1 (6%)	1 (6%)	0 (0%)	4 (18%)
Sentences translated from scratch	0 (0%)	0 (0%)	2 (10.5%)	6 (27%)

Table 2 above presents results of Google MT performance. *Umowa kupna-sprzedaży* [*sale agreement*] was post-edited in 28 min. 20 sec., *umowa o poufności* [*confidentiality agreement*] – in 39 min. 20 sec., *umowa najmu* [*lease agreement*] – in 39 min. 40 sec. and *umowa o dzieło* [*contract for a specific task*] – in 49 min. 50 sec. The average time of post-editing was 22 min. per one translation page (assuming that one translation page consists of 1700 characters with spaces). It was established by dividing the total time of post-editing of the four texts by the total number of translated pages. The amount of time devoted to consultation of sources varied from 2 min. 30 sec. to 8 min. It is related to individual features of a text (terminological complexity) as well as the knowledge of a participant. That is why this figure was presented in a separate row. The average time of translation, calculated in the same manner as the average post-editing time, was 25 min. per one translation page. Approximately 90% of the sentences in Google MT raw output required some degree of editing. The number of nonsense sentences is surprisingly low – about 8%, and appeared in three out of four texts. This could be an indicator of fast-improving quality of MT tools. Sentences translated from scratch constitute close to 11% of the total.

Table 3. General performance of Microsoft MT.

Researched data	Umowa kupna-sprzedaży [sale agreement]	Umowa o poufności [confidentiality agreement]	Umowa najmu [lease agreement]	Umowa o dzieło [Contract for a specific task]
Length of the text	1979 characters (17 sentences)	3600 characters (16 sentences)	3191 characters (19 sentences)	3214 characters (22 sentences)
Total time of translation	30 min. 30 sec.	46 min. 30 sec.	47 min. 30 sec.	51 min. 10 sec.
Time devoted to post-editing:	27 min.	36 min. 20 sec.	38 min. 30 sec.[56]	45 min. 10 sec.
Time devoted to sources consultation:	3 min. 30 sec.	10 min. 10 sec.	1 min.	6 min.
No. of sentences requiring editing	13 (76.5%)	11 (68.5%)	18 (94.5%)	22 (100%)
No. of sentences not requiring editing (100% accuracy)	4 (23.5%)	5 (31.5%)	1 (5.5%)	0 (0%)
Nonsense sentences	1 (6%)	1 (6%)	0 (0%)	7 (32%)
Sentences translated from scratch	0 (0%)	0 (0%)	2 (10.5%)	5 (23%)

Surprisingly, performance of Microsoft MT and Google MT turned out to be very similar. As Table 3 shows, *umowa kupna-sprzedaży* [*sale agreement*] was post-edited in 27 min., *umowa o poufności* [*confidentiality agreement*] – in 36 min. 20 sec., *umowa najmu* [*lease agreement*] – in 38 min. 30 sec. and *umowa o dzieło* [*contract for a specific task*] – in 45 min. 10 sec. The average post-editing time of one translation page was 21 min. and the average time of translation – 24 min. The lenght of time devoted to consulting sources ranges from 1 min. to 10 min. 10 sec. The sentences not requiring editing constitute 13.5% of the total, and nonsense sentences – 12% of the total. 9.5 % of the sentences were translated from scratch.

Errors in mt output

Let us now look at the types of errors most commonly committed by the tested machine translation software. Table 4 presents the errors encountered in Google MT output, and Table 5

---

[56] The actual time of recording is 46 min. 30 sec. However, it includes 8 min. of a break when the translator stopped translating for unknown reason, but did not pause the recording. The recording does not suggest that the translator consulted sources during that time. Therefore, these 8 minutes were deducted from the total time of recording.

– in Microsoft MT output. Right columns in both tables present the average occurrence of particular types of errors per one translation page (i.e., 1700 characters with spaces). This figure was established by dividing the total amount of errors of particular type (appearing in all four texts) by the total amount of pages (i.e., seven). The average was not calculated for several types of errors, because their occurrence depends on the contents of the original text. Each of the four translated texts contained different number of currencies, numbers, dates, proper names, etc., and generally whenever they appeared they were the source of error. *Total errors* category embraces all errors that appeared in all four texts, including items for which the occurrence frequency was not established. The selected categories of errors are discussed in detail further on in this article.

Table 4. Errors in Google Translate raw output.

Type of error	Umowa kupna-sprzedaży [sale agreement]	Umowa o poufności [confidentiality agreement]	Umowa najmu [lease agreement]	Umowa o dzieło [Contract for a specific task]	Total errors (7 transl. pages)	Average occurrence per 1 transl. page
wrong word/phrase	20	28	34	62	144	approx. 20 errors /page
word/ phrase missing	8	4	20	9	41	approx. 6 errors / page
article	9	2	5	4	20	approx. 2.8 errors / page
surplus word	3	9	2	3	17	approx. 2.4 errors / page
punctuation	5	2	5	4	16	approx. 2.3 errors / page
wrong sentence order	4	1	3	5	13	approx. 1.8 errors / page
wrong word form	3	0	1	3	7	approx. 1 error / page
proper name	3	4	4	2	13	not applicable
word capitalization	3	1	5	3	12	not applicable
surplus clause	0	0	6	0	6	not applicable
currency	0	0	3	0	3	not applicable
number	0	0	2	0	2	not applicable
date	0	0	0	0	0	not applicable
total errors	58	51	90	95	294	approx. 42 errors / page

Table 5. Errors in Microsoft MT raw output.

Type of error	Umowa kupna-sprzedaży [sale agreement] (1 transl. page)	Umowa o poufności [confidentiality agreement] (2 transl. pages)	Umowa najmu [lease agreement] (2 transl. pages)	Umowa o dzieło [contract for a specific task] (2 transl. pages)	Total (7 transl. pages)	Average occurrence per 1 transl. page
wrong word/phrase	18	10	22	59	109	approx. 15.6 errors / page
word/phrase missing	4	7	13	11	35	approx. 5 errors / page
proper name	3	4	10	3	20	approx. 2.8 errors / page
wrong sentence order	7	3	4	4	18	approx. 2.5 errors / page
wrong word form	8	2	2	4	16	approx. 2.3 errors / page
article	1	1	7	5	14	approx. 2 errors / page
surplus word	4	2	2	2	10	approx. 1.4 errors / page
punctuation	0	0	3	3	6	approx. 0.8 error / page
word capitalization	10	30	9	1	50	not applicable
date	2	0	2	0	4	not applicable
currency	1	0	1	1	3	not applicable
number	0	0	0	0	0	not applicable
surplus clause	0	0	6	0	6	not applicable
total errors	58	59	75	93	285	approx. 41 errors / page

Selected categories of errors will now be discussed in detail, accompanied by illustrative examples.

Wrong word or phrase
This is a broad category that encompasses various situations when the participant substituted a lexical item from the MT raw output with, in his opinion, a better lexical solution. The items were replaced due to various reasons. Most typically, the word or phrase was perceived by the participants as stylistically awkward. In Example 1, even though the clause *court of jurisdiction appropriate taking into account the Lessor's seat* is understandable, it is stylistically awkward. It was substituted with the clause: *court having jurisdiction over the Lessor's seat.*

Example 1.

ocjacji rozstrzygane będą przez sąd właściwy dla miejsca siedziby Wynajmującego.

., shall be resolved by the court of jurisdiction appropriate taking into account the Lessor's seat.

There were several cases when the original word was left untranslated in the MT output, usually when there was a spelling error in the original text, as illustrated by Example 2.

Example 2.

{0>
(Słownie: osiemdziesiąc sześć tysięcy zł)..
<}0{>
(In words: osiemdziesiac six thousand zł).

Another common error was the use of imprecise legal terms, as illustrated by Example 3, where *posiadanie* was automatically translated as *ownership* instead of *possession.*

Example 3.

Tłumaczenie maszynowe

OWNERSHIP AND USE OF PREMISES

Znaki: 29

Moreover, MT output exhibited insufficient recognition of context, as illustrated by Example 4. The Polish word *zawierać* has several meanings: to close, to be included in something or to conclude an agreement. Example 4 shows that MT provided the equivalent that did not fit into the context. It needs to be stressed, however, that in multiple other cases registered in this experiment MT solutions proved to be context-sensitive. Yet, they are still not faultless.

Example 4.

Zawarta w dniu 12.10.2012 w Krakowie pomiędzy

Included in the day 12.10.2012 in Kraków between

Example 5 illustrates another common situation, when the translation produced by MT was lexically and grammatically correct, but there existed a well-established equivalent of a term that should rather be used. In Example 5, *contract for the work* is a literal translation of *umowa o dzieło* (type of employment contract in Poland designed for freelancers). The most common renditions of the phrase are: *contract for specific work, contract for a specific task,* or *contract of commission.* One of them was used by the participant to replace the translation done by the machine.

Example 5.

UMOWA O DZIEŁO

CONTRACT FOR THE WORK
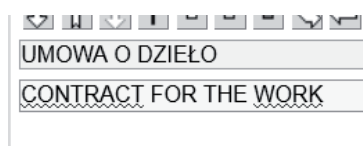
Last but not least, the words or phrases were replaced due to lexical inconsistency in MT output. One basic principle of legal drafting is that for the sake of precision one person or item should in a document be referred to with the same name. In the MT output, however, the term *wynajmujący* was interchangeably translated as *lessor* and *landlord*, and *najemca* as *lessee* and *tenant*. This illustrative example is one of many encountered inconsistencies. As the experiment revealed, this was one of the most common errors committed by MT. Consistency is difficult to achieve for MT solutions, especially corpus-based, because every sentence is translated by the machine independently on the basis of translations found in the corpora. Luckily, the problem can be easily remedied with the use of automatic replacement of terms in a document.

Missing word or phrase
The elements most often omitted in MT raw output were prepositions. Yet, in several cases the sentences also lacked important factual information, as in Example 6, where the final part of the sentence is missing. It has to be stressed that such situations were rare, and general improvement in that respect is noticeable.

Example 6.

Edytowany

Any disputes that may arise in connection with this contract, submit to a court of jurisdiction in which the Employer

Znaki: 118
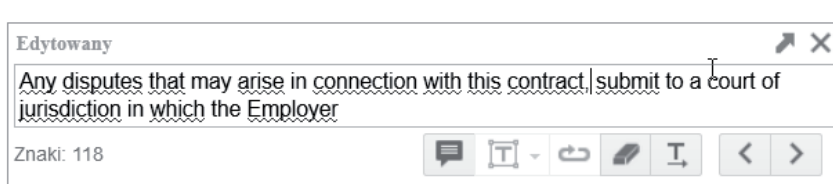
Another source of error was noncompliance with the rules of legal translation. Due to terminological incongruency, it is good practice to provide the original names of system-bound items (such as the name of an institution, a piece of legislation, or a legal term) in square brackets next to their equivalents. This is what MT raw output lacked, as illustrated by Example 7, where the sentence does not include any reference to the country where the legislation is applicable.

Example 7.

FINAL PROVISIONS

Surplus clauses
Surplus clause category pertains to situations when MT raw output included a clause that was not the rendition of an original sentence. This happened during translation of *umowa najmu,* as shown in Example 8. The machine, instead of inserting into the placeable only the number of the paragraph [1.], inserted the number with the accompanying clause which was not related to the original text in any way. There is no logical explanation why the clause appeared in MT output, other than a technical error. The same error reappeared six times, surprisingly, both in Google MT and Microsoft MT. This may suggest that both MT solutions utilize the same corpora of legal texts.
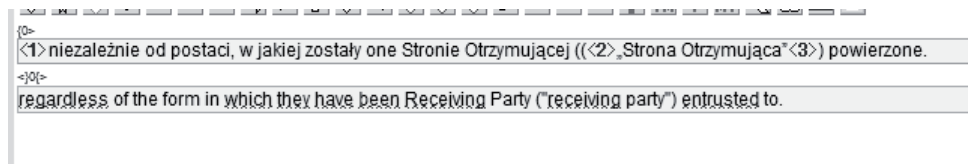
Example 8.
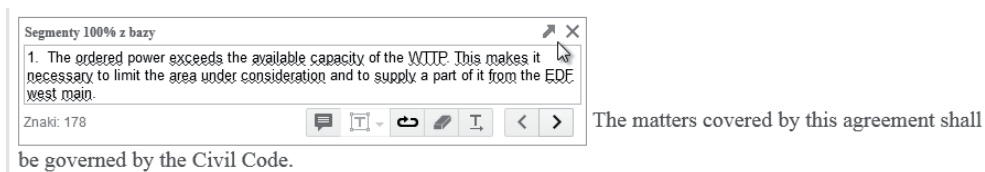
Wrong sentence order
Wrong sentence order appeared fairly rarely – 13 times per 1 translation page in Google MT and 18 times per one translation page in Microsoft MT. This could be attributed to the fact that MT systems tested in this study are not rule-based. Both Google and Microsoft solutions draw on ready-made translations. Thus, in general, syntactic awkwardness, which used to be a serious problem, is now less noticeable. Yet, it is still existent, as illustrated by Example 9, where MT translation is a one-to-one representation of the original sentence order.
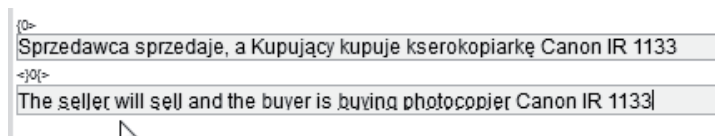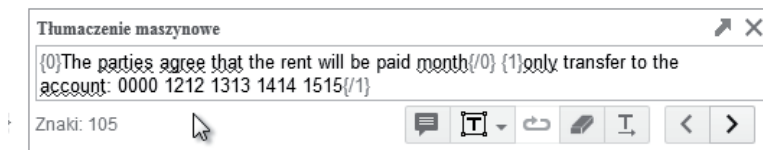
Example 9.

Grammatical errors

Grammatical errors were more common in Microsoft MT output (16 cases) than in Google MT output (7 cases). This category included the following errors: inconsistent use of tenses (Example 10), use of a wrong word category (Example 11) or an incorrect form of a word, as illustrated by Example 12, where the name of the city is not in the nominative case.
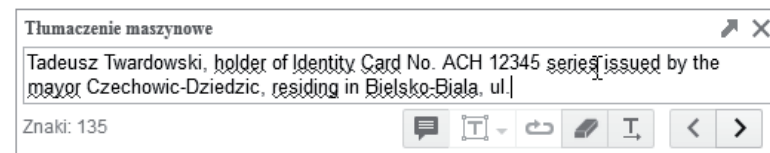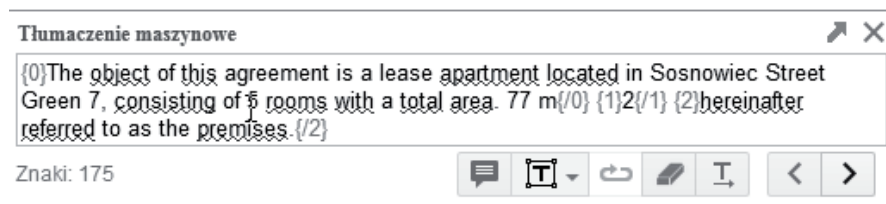
Example 10.

Example 11.

Example 12.

Capitalization

Generally, modern MT technology correctly applies the rules of capitalization. Yet, MT tools are still not aware of the idiosyncrasies of legal writing – namely the rule that the word that has been defined at the beginning of a document is then, in the remaining part of the document, written in capital letters. That was the main source of errors related to capitalization, as illustrated by Example 13, where the words *agreement* and *premises* are not capitalized.
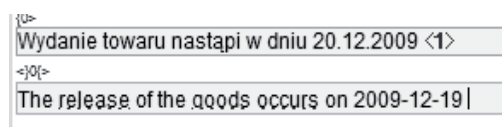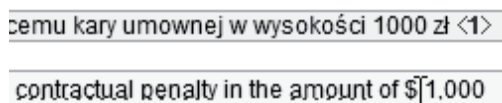
Example 13.

Dates and currencies

MT solutions tested in this study work on the basis of previous translations. That is the reason why there are so many errors related to incorrect dates (Example 14), or currencies (Example 15). This class of errors is especially dangerous, because they pertain to crucial factual information, and can be easily omitted by the human post-editors of MT output, who tend to focus on linguistic aspects of translation.
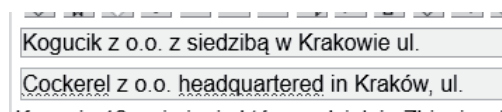
Example 14.

Example 15.

Proper names

MT tools are still unable to recognize proper names, as illustrated by Examples 16, 17 and 18. Example 16 presents awkward rendition of a company's name, Example 17 – wrong rendition of the name of the city, while Example 18 – erroneous rendition of a name and a surname – Jan Kowalski translated as John Smith – plus incorrect punctuation. The experiment revealed that the proper names that resembled standard words were automatically translated by the machine into the target language (even though they were capitalized), whilst the proper names that did not match any dictionary word were left untranslated. The errors of this kind were very common - appeared throughout all eight analyzed texts.

Example 16.

Example 17.

odem osobistym seria nr ACH 12345 wydanym przez burmistrza Czechowic-Dziedzic, zamieszk

, issued by Mayor Czechowic-Heir, residing in Bielsko-Biała, ul.

Example 18.

e przy ul Głowackiego 12, reprezentowanym przez Jana Kowalskiego, zwanym w treści umowy "Sprzedawcą",

, Głowackiego, represented by John Smith, called in the body of the agreement, "Seller", and

Nonsense clauses

MT tools do not produce as many nonsense translations as it is generally believed. There were only 6 registered instances of nonsense clauses in Google MT output and 9 – in Microsoft MT output. However, the experiment revealed that there are still cases when MT output resembles literal translation of the original text, i.e. each particular word is being translated by the machine, even when it should not be, which results in disruption of sentence logic. A few striking examples of nonsense translations are presented below (Examples 19, 20, 21, 22).

Example 19.

Tadeuszem Twardowskim, legitymującym się dowodem osobistym seria nr ACH 12345 wydanym przez burmistrza Czechowic-Dziedzic, Białej, przy ul.

Tadeusz External, giving legitimacy to the ID series no OH 12345 issued by Mayor Czechowic-Heir, residing in Bielsko-Biała, ul.

Example 20.

Kupujący może dochodzić na zasadach ogólnych odszkodowania przewyższającego karę umowną.

The buyer may assert in general damages in excess of the contractual penalty.

Example 21.

Odstąpienie od umowy nie powoduje utraty możliwości dochodzenia przez Zamawiającego odszkodowania i kary umownej.

A waiver of the contract does not result in the loss of the investigation by the customer compensation and contractual penalties.

Example 22.

> Wykonawca zobowiązany jest przedstawić rozliczenie z otrzymanych materiałów i narzędzi, n
>
> The contractor is obliged to present the settlement from materials and tools, not worn out and

Clauses translated from scratch

The decision to delete a whole sentence was absolutely subjective, based on the participant's personal opinion that the sentence is not editable. The participants translated the sentence from scratch when they wanted to apply a completely different grammatical construction than the one proposed by MT, as in Example 23, where the sentence was transformed by the participant into passive. The participants decided to translate from scratch also when logic of the sentence was disturbed and/or the number of errors was too high to make post-editing worthwhile (Examples 24 and 25).

Example 23.

> Wynajmujący do każdej płatności przedstawi rachunek.
>
> The Lessor shall issue a bill in respect to each payment.

Example 24.

> Odstąpienie od umowy nie powoduje utraty możliwości dochodzenia przez Zamawiającego odszkodowania i kary umownej.
>
> A waiver of the contract does not result in the loss of the investigation by the customer compensation and contractual penalties.

Example 25.

> Grochowej 13, zwanym w treści umowy „Wykonawcą" o następującej treści:
>
> 13, makes it possible in the body of the called the agreement "the contractor" reads as follows:

Conclusions

The above presentation of errors should not give the impression that MT performance is of low quality. Quite the contrary. The result of the experiment is very promising. In general, MT tools exhibit good recognition of the genre. The raw output produced by the machine in the experiment looked like a legal text. Moreover, majority of sentences produced by MT were perceived by the participants as logical and understandable. Even though there were instances of imprecision, stylistic awkwardness or even serious factual mistakes, on the whole, the amount of editing was not overwhelming, which was reflected in short post-editing time. It is possible to further improve the time of post-editing with deepened knowledge of what errors to expect and how to use MT tools

more efficiently (e.g., by using Word Processor options to automatically erase all errors of one type in a document).

Is it then recommended to use MT solutions during translation? The experiment revealed that cooperation with MT tools differs significantly from traditional translation. On the one hand, MT assistance releases translators from the excessive use of memory and typewriting, thus it might be welcomed by translators struggling with these aspects of a translation task. On the other hand, cooperation with MT demands critical thinking, perceptiveness and most of all flexibility. Translators who want to use MT in their work need to be willing to accept a different translation than the one that they have in mind. Therefore, the result of this experiment should be matched to individual situation of each translator. Everyone needs to weigh pros and cons of machine translation and decide individually whether it is worth adding to one's workstation. Hopefully, this study, by showing MT's strengths and weaknesses sheds some light on the topic and helps to make an informed decision in this matter.

Bibliograhy

Berezowski, Leszek. 2008. *Jak Czytać i Rozumieć Angielskie Umowy*. Warszawa: C.H. BECK.

Bijimol, T. K., and John T. Abraham. 2014. *A Study of Machine Translation Methods*. http://csidl.org/handle/123456789/675.

Fulford, H. 2002. Freelance Translators and Machine Translation: An Investigation of Perceptions, Uptake, Experience and Training Needs. In *Proc. 6th EAMT Workshop*, 117–22.

Fulford, Heather, and Joaquín Granell-Zafra. 2004. The Freelance Translator's Workstation: An Empirical Investigation. In *Proceedings of 9th European Association for Machine Translation Workshop*, 53-61.

Graham, Yvette, Timothy Baldwin, Alistair Moffat, and Justin Zobel. 2014. Is Machine Translation Getting Better over Time? In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, 443-451.

Hutchings, John. 2008http://www.hutchinsweb.me.uk/OUP-2008.pdf.

Hutchins, John. 2003. Commercial Systems. edited by Harold Somers. *Computers and Translation: A Translator's Guide.* John Benjamins Publishing, Amsterdam, 161–74.

Hutchins, John. 2007. Machine Translation: A Concise History. Edited by C. Wai. *Computer Aided Translation: Theory and Practice,* Hong Kong: Chinese University of Hong Kong.

Hutchins John. 2010. Outline of Machine Translation Developments in Europe and America. In *JAPIO*, 1–8. Tokyo.

Kit, Chunyu, and Tak Ming Wong. 2008. Comparative Evaluation of Online Machine Translation Systems with Legal Texts. In *Law Library Journal* 100 (2): 299–321.

Koponen, M., W. Aziz, L. Ramos, and L. Specia. 2012. Post-Editing Time as a Measure of Cognitive Effort. In *AMTA 2012 Workshop on Post-Editing Technology and Practice (WPTP 2012)*, 11–20.

Och, Franz. 2012. Breaking down the Language Barrier - Six Years in. In *Google Official Blog*. April 26. http://googleblog.blogspot.com/2012/04/breaking-down-language-barriersix-years.html.

Šarčević, Susan. 2000. *New Approach to Legal Translation*. Hague: Kluwer Law International.

Seljan, Sanja, Marija Brkić, and Vlasta Kučiš. 2011. Evaluation of Free Online Machine Translations for Croatian-English and English-Croatian Language Pairs. In *Proceedings of the 3rd International Conference on the Future of Information Sciences: INFuture2011-Information Sciences and E-Society*, 331–45.

Snover, M., B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul. 2006. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA-2006)*, 223–31.

Zervaki, T. 2002. Online Free Translation Services. In *Proceedings of the 24th Int. Conf. on Translating and the Computer*.