Baltic J. Modern Computing, Vol. 4 (2016), No. 2, pp. 230–242

# Can Text Simplification Help Machine Translation?

Sanja ŠTAJNER[1] and Maja POPOVIĆ[2]

[1] Data and Web Science Group, University of Mannheim, Germany
[2] Humboldt University of Berlin, Germany

sanja@informatik.uni-mannheim.de, maja.popovic@hu-berlin.de

**Abstract.** This article explores the use of text simplification as a pre-processing step for statistical machine translation of grammatically complex under-resourced languages. Our experiments on English-to-Serbian translation show that this approach can improve grammaticality (fluency) of the translation output and reduce technical post-editing effort (number of post-edit operations). Furthermore, the use of more aggressive text simplification methods (which do not only simplify the given sentence but also discard irrelevant information thus producing syntactically very simple sentences) also improves meaning preservation (adequacy) of the translation output.

## 1 Introduction

Machine translation for under-resourced languages is facing a number of problems. First, there is not enough parallel data to build robust statistical machine translation (SMT) systems. Second, most of these languages (including Serbian) have a very rich morphology and suffer from data sparsity when it comes to less frequently used cases, tenses, etc. Third, there is a number of syntactic differences which are difficult to capture. For English-to-Serbian SMT, a number of language related problems has been identified so far (Popović and Arčan, 2015). Most of them are related to syntactic differences, e.g. missing verb parts due to distinct structure of certain verb tenses, incorrect prepositions, or incorrect translations of English sequences of nouns.

In this paper, we explore whether it is possible to improve the performance of the machine translation for under-resourced languages by introducing a pre-processing step in which source sentences are first simplified by an automatic text simplification (ATS) system. We focus on English-to-Serbian MT and apply two state-of-the-art ATS systems as a pre-processing step for simplifying the original English sentence before feeding it into a phrase-based SMT system.

We exploit two different types of ATS systems, a more conservative one (which, while simplifying the input sentence lexically and syntactically, retain all the information contained in the original sentence), and the more aggressive one (which, while sim-

plifying the input sentence lexically and syntactically, also tries to reduce the amount of information by discarding irrelevant information and high-level details). In this way, we address two different usage scenarios in MT: (1) when it is important to maintain all the information contained in the source text (e.g. translations of whole texts or documents); and (2) when it is enough to get a gist of the source text (e.g. skimming through news articles and looking for the most important news).

The results of the human evaluation of the news articles translated using the two above-mentioned approaches, in terms of grammaticality (fluency) and meaning preservation (accuracy) of the output, and the analysis of the post-editing effort (number of post-edit operations) shows that both approaches improve the MT output.

The remainder of the article is structured as follows. Section 2 briefly reports on the existing approaches to automatic text simplification and motivates our choice of ATS systems. Section 3 describes the chosen ATS systems in more details, presents the datasets and the SMT system used in experiments and describes the evaluation procedure. Section 4 presents and discusses the results of our experiments, while Section 5 summarises the main findings and presents ideas for future research.

## 2 Related Work

Automatic text simplification (ATS) systems aim to transform original texts into their lexically and syntactically simpler variants. In theory, they could also simplify texts on the discourse level, but most of the systems still operate only on the sentence level.

The motivation for building the first ATS systems was to improve the performance of machine translation systems and other text processing tasks, e.g. parsing, information retrieval, and summarisation (Chandrasekar et al., 1996). It was argued that simplified sentences (which have simpler sentential structures and reduced ambiguity) could lead to improvements in the quality of machine translation (Chandrasekar, 1994).

Since then, a great number of ATS systems has been proposed not only for English, but also for other languages, e.g. Basque (Aranzabe et al., 2013), Portuguese (Specia, 2010), Spanish (Saggion et al., 2015), French (Brauwers et al., 2014), and Italian (Barlacchi and Tonelli, 2013).

For English, the state-of-the-art ATS systems range from those performing only lexical (Glavaš and Štajner, 2015) or only syntactic (Siddharthan, 2011) simplification, to those combining lexical and syntactic simplification (Angrosh and Siddharthan, 2014). Recently, several ATS systems have been proposed which do not only simplify given text/sentences but also reduce the amount of information contained by removing high-level details, such as appositions, adverbial phrases, or purely descriptive sentences ((Glavaš and Štajner, 2013), (Siddharthan et al., 2014), (Narayan and Gardent, 2014)).

However, in these twenty years, the motivation for building ATS systems has shifted from improving text processing systems to making texts more accessible to wider audiences (e.g. children, non-native speakers, people with low literacy levels, and people with various language or learning disabilities). Therefore, ATS systems have only been evaluated for the quality of the generated output, its readability levels, and usefulness in making texts more accessible to target populations (reducing reading speed and improving comprehension). To the best of our knowledge, there has been no evaluation of

the state-of-the-art ATS systems in terms of how much (if at all) they can help improve MT systems (which was, as previously mentioned, their first intended goal and main motivation).

## 3 Experiments

In this study, we use two state-of-the-art ATS systems:

1. **TS-A**: A combination of lexical TS system (Glavaš and Štajner, 2015) with the EventSimplify (Glavaš and Štajner, 2013) which performs syntactic simplification with a significant content reduction. This is the most "aggressive" system of all above-mentioned systems which perform content reduction (Section 2), i.e. it is the system which performs the highest level of content reduction and achieves the most readable (simplest) output (due to a high number of sentence splitting operations).
2. **TS-C**: The lexico-syntactic TS system proposed by Angrosh and Siddhathan (2014) which belong to the "conservative" ATS systems which do not perform any content reduction and thus, completely preserve the original meaning of the sentence;

We used 100 news articles from the EMM NewsBrief[3] for which the output of the EventSimplify (Glavaš and Štajner, 2013) ATS system was freely available.[4] We further focused on the output of the event-wise simplification scheme (which achieved the highest readability of all four provided schemes) and applied the lexical simplification system (Glavaš and Štajner, 2015) on top of it in order to obtained a full simplification system which encompasses lexical simplification, syntactic simplification and content reduction (TS-A). Next, we applied the TS-C system on all those 100 original articles.

### 3.1 Text Simplification Systems

The examples presented in Table 1 illustrate the potential of the two ATS systems used (*TS-C* and *TS-A*) and differences among them. In general, the TS-A performs more sentence splitting than the TS-C (see examples 2, 3, and 4 in Table 1, with the extreme case of producing four simplified sentences instead of one original sentence in the fourth example). The TS-A system also removes some details (e.g. *"several minutes later"* in the third example, or *"in Port St. John"* in the second example), or entire subordinate clauses (e.g. *"a steep fall from.."* in the first example).

The main focus of both ATS systems is on structural simplification, although there are occasional cases of lexical simplification as well (e.g. *"arrived"* → *"came"* in the second example, or *"recieved"* → *"got"* and *"refuge"* → *"shelter"* in the fourth example).

It is interesting to note that both systems (though TS-A more frequently) also simplify the tense of the verbs, as in the following examples: *"before* **turning** *the gun* → *"After that, ...* **turned** *the gun* (ex. 2), *"Deputies arrived...* **to hear** *..."* → *"Deputies came ... Deputies* **heard***..."* (ex. 3), and "**had received**" → "**got**" (ex. 4). Furthermore, the AT-C system consistently changes constructions of the type **"<clause>, X said."** into **"X said that <clause>"** (as illustrated in the second example in Table 1).

[3] http://emm.newsbrief.eu/NewsBrief/clusteredition/en/latest.html
[4] http://takelab.fer.hr/data/evsimplify/

**Table 1.** Examples of sentence simplification performed by the two ATS systems (TS-C and TS-A). Differences to the original sentences are shown in bold.

| Ex. | Version | Sentence |
|---|---|---|
| 1 | Original | Vladimir Putin's United Russia party won less than 50% of Sunday's vote, a steep fall from its earlier two-thirds majority, according to preliminary results. |
| | TS-C | Vladimir Putin's United Russia party won less than 50% Sunday's vote**, according to preliminary results. This is a steep fall from its earlier two-thirds majority.** |
| | TS-A | Putin's United Russia party won less than 50%. |
| 2 | Original | A Florida mother shot her four children early Tuesday morning before turning the gun on herself at her home in Port St. John, police said. |
| | TS-C | **Police said that a** Florida mother shot her four children early Tuesday morning before turning the gun on herself at her home in Port St. John. |
| | TS-A | A Florida mother shot her four children early Tuesday morning**. After that, a Florida mother turned** the gun on herself at her home. |
| 3 | Original | Deputies arrived at the house several minutes later to hear more shots fired. |
| | TS-C | Deputies **came** at the house several minutes later to hear more shots fired. |
| | TS-A | Deputies **came** to the house**. Deputies heard** more shots. |
| 4 | Original | The Chinese Embassy said it had received a report that a dozen Chinese fishing boats had taken refuge in a lagoon of Huangyan Island to escape foul weather when the Philippine gunboat blocked the lagoon entrance and sent 12 Philippine soldiers to harass the Chinese fishermen. |
| | TS-C | The Chinese Embassy said it **also got** a report that a dozen Chinese fishing boats had taken refuge in a lagoon of Huangyan Island to escape foul weather. Then the Philippine gunboat **sent 12 Philippine soldiers to harass the Chinese fishermen. At that time, the gunboat blocked the lagoon entrance.** |
| | TS-A | The Chinese Embassy had received a report**. A** dozen Chinese fishing **ships** had taken **shelter** in a lagoon of Huangyan Island**. T**he Philippine gunboat blocked the lagoon entrance**. The Philippine gunboat** sent 12 Philippine soldiers to harass the Chinese fishermen. |

As ATS systems do not always produce perfectly grammatical output and lexical simplification sometimes lead to changed meaning (Angrosh and Siddharthan, 2014; Glavaš and Štajner, 2015), we manually inspected a randomly selected subset of 65 original sentences and their automatically simplified sentences produced by both systems (TS-A and TS-C).[5] In those cases where the meaning or grammaticality was incorrect, we performed a minimal post-editing (PE) necessary to restore the original meaning and grammaticality of the sentence. As the goal of this PE is not to make any further simplifications and the mistakes were easy to notice, this type of PE was very fast (11.3 seconds per sentence for TS-A and 15.2 seconds per sentence for TS-C) and did not even require a native speaker or trained annotator, but only someone with the proficiency level of English. For illustration, several sentences are given in Table 2.

**Table 2.** Examples of post-editing performed on the automatically simplified sentences generated by the TS-C and TS-A systems. Differences between the automatically simplified sentences and their PE versions are shown in bold.

| Ex. | Version | Sentence |
|---|---|---|
| 1 | Original | Ex-Soviet leader Mikhail Gorbachev says Russian authorities must annul the parliamentary vote results and hold a new election. |
| | TS-C (no PE) | Ex-Soviet leader Mikhail Gorbachev says**.** Russian authorities must annul the parliamentary vote results. These authorities hold a new election. |
| | TS-C (PE) | Ex-Soviet leader Mikhail Gorbachev says **that** Russian authorities must annul the parliamentary vote results. These authorities **must** hold a new election. |
| 2 | Original | A 21-year-old man was arrested on April 30, on suspicion of murder and was released on bail until May 29 pending further enquiries. |
| | TS-C (no PE) | A 21-year-old man was arrested on April 30, on suspicion of murder. This man was **followed** until May 29 pending further enquiries. |
| | TS-C (PE) | A 21-year-old man was arrested on April 30, on suspicion of murder. This man was **released** until May 29 pending further enquiries. |
| | TS-A (no PE) | A 21-year-old man was arrested on April 30 on suspicion. A 21-year-old man was released on **jail** until May 29. |
| | TS-A (PE) | A 21-year-old man was arrested on April 30 on suspicion of murder. A 21-year-old man was released on **bail** until May 29. |

[5] This subset of sentences was later used for MT experiments and human evaluation and post-editing.

### 3.2 Statistical Machine Translation System

For the machine translation from English to Serbian, we used the ASISTENT system.[6] It is a freely available SMT system, based on the widely used phrase-based SMT framework (Koehn et al., 2003) and it supports translations from English to Slovene, Croatian and Serbian and vice versa. Additionally, translations between those three Slavic languages are also possible.

The system was trained using the Moses toolkit (Koehn et al., 2007). The word alignments were built with GIZA++ (Och and Ney, 2003), and the 5-gram language model was built using the SRILM toolkit (Stolcke, 2002) The training dataset originates from the OPUS website[7] (Tiedemann, 2012) where three domains were available for the Serbian-English language pair: the enhanced version of the SEtimes corpus[8] (Tyers and Alperen, 2010) containing "news and views from South-East Europe", OpenSubtitles[9], and the KDE localisation documents and manuals, i.e. technical domain. Approximately 20.7M sentences, in total, were used for training (20.5M subtitles, 200,000 news, 30,000 technical), and 2,000 sentences were used for tuning (retaining the same proportions of the sentences from the three corpora as in the training dataset).

The English-to-Serbian part of the ASISTENT system (Arčan et al., 2016) was tested on 2,000 sentences from the three corpora used for training and tuning (the 2,000 sentences which were not used for training and tuning) and achieved a 38.88 BLEU score (Papineni et al., 2002), a 31.18 METEOR score (Denkowski and Lavie, 2014), and a 61.62 chrF3 score (Popović, 2015).

### 3.3 Evaluation Procedure

From the initial set of 100 news articles, we randomly selected 65 original sentences and evaluated all translation outputs (from original sentences, and TS-A and TS-C systems, which led to a total of 195 target sentences) with respect to the following aspects:

1. adequacy, i.e. meaning preservation
2. fluency, i.e. grammaticality
3. technical post-editing effort, i.e. amount of necessary edit operations

Each of the tasks has been carried out separately, i.e. the evaluation of adequacy and fluency were carried out in two separate passes, and post-editing was carried out in the third pass.

For adequacy, a quality score from 1 to 5 was assigned to each segment according to the following guidelines:

- 1 = very bad (regardless of a potentially good grammaticality)
- 2 = difficult to understand and different from the source meaning
- 3 = the main idea is preserved but some parts are unclear/different from the source
- 4 = understandable with minor ambiguities/differences

[6] http://server1.nlp.insight-centre.org/asistent/
[7] http://opus.lingfil.uu.se/
[8] http://nlp.ffzg.hr/resources/corpora/setimes/
[9] http://www.opensubtitles.org/

– 5 = perfectly understandable (regardless of a potentially poor grammar)

For fluency scores, the following guidelines were used:

– 1 = very bad (regardless of a potentially good meaning preservation)
– 2 = many grammatical errors
– 3 = a number of grammatical errors but mostly minor ones
– 4 = almost correct (a small number of minor errors)
– 5 = perfectly grammatical (regardless of possible loss/change of meaning)

The post-editing effort was analysed in the following way:

– Each translated segment was post-edited by looking into the corresponding source segment, i.e. using English originals for translations of originals, using the corresponding simplified English sentences for translations of simplified segments.
– The raw edit counts and edit rates (raw counts normalised with the segment length) were calculated using Hjerson (Popović, 2011) for:
  • five classes of edits/errors
  • all edit operations

Reference translations were not available.

## 4 Results and Discussion

The average adequacy and fluency scores, and the percentages of sentences with each of the scores are presented in Table 3. It can be noted that the use of TS-C does not improve the overall adequacy, but it might improve fluency, whereas the use of TS-A improves MT in both aspects.

**Table 3.** Average scores for adequacy and fluency (first row) and percentage of sentences for each of the five scores (1–5).

| Score | Adequacy | | | Fluency | | |
|---|---|---|---|---|---|---|
| | Orig | TS-C | TS-A | Original | TS-C | TS-A |
| Average | 3.17 | 3.02 | 3.63 | 2.91 | 3.13 | 3.45 |
| 1 | 15.2 | 13.0 | 6.5 | 4.3 | 2.3 | 2.3 |
| 2 | 10.9 | 17.4 | 8.7 | 23.9 | 17.4 | 11.4 |
| 3 | 32.6 | 32.6 | 30.4 | 47.8 | 45.6 | 31.8 |
| 4 | 23.9 | 28.3 | 23.9 | 23.9 | 34.8 | 47.7 |
| 5 | 17.4 | 8.7 | 30.4 | 0 | 0 | 6.8 |

A closer look into the distribution of the sentence scores indicates that the use of the TS-C system in MT decreases the number of sentences with very bad accuracy score, but it also decreases the number of sentences with perfect adequacy scores. The TS-A

system, however, significantly increases the number of sentences with perfect adequacy scores, at the same time decreasing the number of sentences with low adequacy scores.

As for the fluency, both TS systems significantly increase the number of sentences with high fluency scores (score 4, and in the case of TS-A, score 5 as well) while at the same time they decrease the number of sentences with low fluency scores. It should be noted that the fluency is generally problematic for the SMT system – none of the original English sentences has been translated into a perfectly grammatical sentence, and the use of TS-C does not succeed in improving this either. However, the use of the TS-A system leads to a 6.8% of sentences being translated into perfectly grammatical sentences.

**Table 4.** Percentage of changes in adequacy and fluency scores.

(a) Adequacy

| Original | TS-C | | | | | TS-A | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 1 | 10.9 | **2.2** | **2.2** | **0** | **0** | 4.3 | **2.2** | **2.2** | **4.3** | **2.2** |
| 2 | 0 | 4.3 | **4.3** | **2.2** | **0** | 0 | 2.2 | **6.5** | **0** | **2.2** |
| 3 | 2.2 | 8.7 | 15.2 | **6.5** | **0** | 2.2 | 2.2 | 15.2 | **2.2** | **10.9** |
| 4 | 0 | 0 | 4.3 | 17.4 | **2.2** | 0 | 0 | 4.3 | 13.0 | **6.5** |
| 5 | 0 | 2.2 | 6.5 | 2.2 | 6.5 | 0 | 2.2 | 2.2 | 4.3 | 8.7 |

(b) Fluency

| Original | TS-C | | | | | TS-A | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 1 | 2.2 | **2.2** | **0** | **0** | **0** | 0 | **4.3** | **0** | **0** | **0** |
| 2 | 0 | 4.3 | **17.4** | **2.2** | **0** | 0 | 2.2 | **8.7** | **13.0** | **0** |
| 3 | 0 | 8.7 | 21.8 | **17.4** | **0** | 2.2 | 2.2 | 17.4 | **19.6** | **6.5** |
| 4 | 0 | 2.2 | 6.5 | 15.2 | **0** | 0 | 2.2 | 4.3 | 13.0 | **4.4** |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 4 presents the results of further analysis, showing the percentage of each particular change in adequacy and fluency scores for each of the TS systems. The desired changes (from lower to higher score) are presented in bold.

For the TS-C system, it is confirmed that a number of sentences with a bad adequacy score is improved, and on the other hand, a number of sentences with a good adequacy score is deteriorated. The majority of sentences does not change. As for the fluency, the main improvement comes from improving poor sentences into medium ones and medium sentences into almost good ones. The majority of sentences does not change.

For the TS-A system, the main changes in adequacy originate from improving sentences with very bad adequacy scores even up to perfect, and from the improvement

of sentences with a medium adequacy score into perfect. The main contribution for fluency, using the TS-A system, comes from improving medium sentences into almost perfect, and from improving poor ones into medium and almost good.

For illustration, Table 5 contains several examples of original sentences, their automatically simplified sentences by both systems and the fluency and adequacy scores for the produced translations into Serbian. The first example shows how a strong reordering of clauses within a sentence (without any sentence splitting) can improve both fluency and adequacy of the translation output. The second example demonstrates how even one lexical change (replacement of a phrasal verb with a more frequently used non-phrasal verb) can also improve the fluency and adequacy of the translation. The third example shows how much sentence splitting and its combination with lexical simplification can improve the translation in the case of a long source sentence. In the penultimate example, we again see how much sentence splitting in a combination with tense simplification and discarding details can improve translation, leading to a perfect fluency and adequacy. The last example demonstrates how retaining only the most important information can improve the fluency of the translation output.

### 4.1 Post-Editing Effort

Results for the post-editing effort are shown in Table 6. The overall raw count of edit operations decreases for both TS systems albeit significantly more for the TS-A, which is expected since the sentences are shorter. Edit rates also decrease for both TS systems, but more for the TS-C due to the reduced sentence lengths of the TS-A system.

Furthermore, the TS-A reduces raw counts for each of the five error classes, whereas the improvement with the TS-C comes mainly from the reduction of reordering errors. This is still an important improvement since it has been shown that the reordering edit operations strongly correlate with the cognitive post-editing effort (Popović et al, 2014).

Table 7 shows the percentage of improved, deteriorated and unchanged sentences for both TS systems with regard to all evaluation aspects, i.e. adequacy, fluency, edit rate, and raw count of edit operations.

For about one half of the sentences (54.3% for the TS-C and 43.5% for the TS-A) the adequacy scores do not change. Among those sentences which do change the adequacy score, in the case of the TS-C, more sentences deteriorate their score than improve it (26.1% as opposed to 19.6%), while in the case of the TS-A, in contrast, more sentences improve their adequacy instead of deteriorating it (39.1% as opposed to 17.4%).

The number of sentences that improve their fluency is higher than the number of sentences that deteriorate it for both TS systems, and it is particularly pronounced for TS-A.

Edit rates are improved significantly with using the TS-C (47.8%) and for the majority of sentences (54.3%) using the TS-A. Raw counts of edit operations are improved for more than one half of the sentences by the TS-C (60.9%) and for more than 82% of the sentences by the TS-A.

**Table 5.** Examples of the adequacy and fluency scores received for the translation of original sentence and two automatically simplified sentences (using the TS-C and TS-A systems), for the cases where TS led to improvements in the output. Differences between the original sentences and their automatically simplified versions are shown in bold.

| Ex. | Version | A | F | Sentence |
|---|---|---|---|---|
| 1 | Original | 2 | 3 | ”As we emerge from a decade of conflict abroad and economic crisis at home, it’s time to renew America,” Obama said, speaking against a backdrop of armored vehicles and a U.S. flag. |
| | TS-C | 4 | 4 | **Speaking against a backdrop of armored vehicles and a U.S. flag, Obama said it’s time to renew America as we emerge from a decade of conflict abroad and economic crisis at home.** |
| 2 | Original | 3 | 2 | Several Israeli security delegations have visited Egypt during the past two months to **decide on** a new embassy location. |
| | TS-C | 4 | 4 | Several Israeli security delegations have visited Egypt during the past two months to **choose** a new embassy location. |
| 3 | Original | 1 | 2 | The Chinese Embassy said it had received a report that a dozen Chinese fishing boats had taken refuge in a lagoon of Huangyan Island to escape foul weather when the Philippine gunboat blocked the lagoon entrance and sent 12 Philippine soldiers to harass the Chinese fishermen. |
| | TS-C | 2 | 3 | The Chinese Embassy said it **also got** a report that a dozen Chinese fishing boats had taken refuge in a lagoon of Huangyan Island to escape foul weather. Then the Philippine gunboat **sent 12 Philippine soldiers to harass the Chinese fishermen. At that time, the gunboat blocked the lagoon entrance.** |
| | TS-A | 3 | 3 | The Chinese Embassy had received a report**. A** dozen Chinese fishing **ships** had taken **shelter** in a lagoon of Huangyan Island**. T**he Philippine gunboat blocked the lagoon entrance**. The Philippine gunboat** sent 12 Philippine soldiers to harass the Chinese fishermen. |
| 4 | Original | 4 | 3 | A Florida mother shot her four children early Tuesday morning before turning the gun on herself at her home in Port St. John, police said. |
| | TS-A | 5 | 5 | A Florida mother shot her four children early Tuesday morning**. After that, a Florida mother turned** the gun on herself at her home. |
| 5 | Original | 4 | 3 | Vladimir Putin’s United Russia party won less than 50% of Sunday’s vote, a steep fall from its earlier two-thirds majority, according to preliminary results. |
| | TS-A | 4 | 4 | Putin’s United Russia party won less than 50%. |

**Table 6.** Raw counts and edit rates (%) normalised with the segment length.

| Edit | Raw counts | | | Edit rates (%) | | |
|---|---|---|---|---|---|---|
| Operations | Orig. | TS-C | TS-A | Orig. | TS-C | TS-A |
| Σ errors | 565 | 542 | 321 | 46.2 | 43.0 | 45.0 |
| Morphology | 209 | 210 | 132 | 17.2 | 16.9 | 18.6 |
| Order | 100 | 66 | 43 | 8.2 | 5.3 | 6.1 |
| Omission | 76 | 80 | 38 | 5.8 | 5.9 | 5.0 |
| Addition | 21 | 26 | 10 | 1.7 | 2.1 | 1.4 |
| Mistranslation | 159 | 160 | 98 | 13.1 | 12.8 | 13.8 |

**Table 7.** Percentage of sentences with better/worse/same sentences with respect to adequacy (A), fluency (F), edit rate (ER) and raw edit counts (REC).

| % | TS-C | | | | TS-A | | | |
|---|---|---|---|---|---|---|---|---|
| | A | F | ER | REC | M(A) | G(F) | ER | REC |
| better | 19.6 | 39.1 | 60.9 | 47.8 | 39.1 | 52.1 | 54.3 | 82.6 |
| worse | 26.1 | 17.4 | 34.8 | 39.1 | 17.4 | 15.2 | 45.6 | 8.7 |
| same | 54.3 | 43.5 | 4.3 | 13.0 | 43.5 | 32.6 | 0 | 8.7 |

## 5 Summary and Outlook

In this article, we investigated whether the state-of-the-art automatic text simplification systems (ATS) can improve English-to-Serbian machine translation (MT) if used as a pre-processing step to simplify source sentences before translating them with the SMT system. We tested this hypothesis by using two ATS systems, a more "conservative" one (TS-C) which only performs lexical and syntactic simplifications, and a more "aggressive" one (TS-A) which performs more lexical and syntactic changes but also performs a significant content reduction thus leading to a loss of some information details.

All the presented results indicate that the use of the TS-C can improve the fluency of the MT output and reduce technical and cognitive post-editing effort through reduction of reordering errors. The use of the TS-A introduces even more improvements for adequacy, fluency and all types of edit operations, but at the cost of losing some details in the information. This approach, however, could be very useful for tasks where the main meaning of the text is crucial and the loss of some details is affordable.

In addition, our results show that the use of a TS system as a pre-processing step in a MT pipeline is only useful for a subset of sentences, whereas the rest of the sentences either deteriorates or remains unchanged. Therefore a method for filtering sentences into two or three classes (TS improves/TS worsens or TS improves/TS does not influence/TS worsens) would be very useful and should be investigated in the future work.

In future research, we will also include more language pairs and domains.

## Acknowledgements

We would like to thank Mihael Arčan for the help with the English-to-Serbian SMT system, and to Goran Glavaš, Advaith Siddharthan and Mandya Angrosh for the help with the automatic text simplification systems.

## References

Mandya Angrosh and Advaith Siddharthan. 2014. Hybrid text simplification using synchronous dependency grammars with hand-written and automatically harvested rules. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Gothenburg, Sweden, pages 722–731.

María Jesús Aranzabe, Arantza Díaz de Ilarraza, and Itziar Gonzalez-Dios. 2013. Transforming Complex Sentences using Dependency Trees for Automatic Text Simplification in Basque. *Procesamiento del Lenguaje Natural*, Volume 50, pages 61–68.

Mihael Arčan, Maja Popović and Paul Buitelaar. Asistent – a machine translation system for Slovene, Serbian and Croatian. In *Proceedings of the 10th Conference on Language Technologies and Digital Humanities*, Ljubljana, Slovenia.

Gianni Barlacchi and Sara Tonelli. 2013. ERNESTA: A Sentence Simplification Tool for Children's Stories in Italian. In *Computational Linguistics and Intelligent Text Processing*, pages 476–489.

Laetitia Brouwers, Delphine Bernhard, Anne-Laure Ligozat and Thomas François. 2014. Syntactic Sentence Simplification for French. In *Proceedings of the EACL Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, Gothenburg, Sweden, pp. 47–56.

Raman Chandrasekar. 1994. Hybrid Approach to Machine Translation using Man Machine Communication. *PhD Thesis*. Tata Institute of Fundamental Research, University of Bombay, Bombay, India.

Raman Chandrasekar, Christine Doran and Bangalore Srinivas. 1996. Motivations and Methods for Text Simplification. In *Proceedings of the Sixteenth International Conference on Computational Linguistics (COLING)*, pages 1041–1044.

Michael Denkowski and Alon Lavie. 2014. Meteor Universal: Language Specific Translation Evaluation for Any Target Language. In *Proceedings of the EACL 2014 Workshop on Statistical Machine Translation*, pages 376–380.

Goran Glavaš and Sanja Štajner. 2013. Event-Centered Simplication of News Stories. In *Proceedings of the Student Workshop held in conjunction with RANLP Conference*, Hissar, Bulgaria, pages 71–78.

Goran Glavaš and Sanja Štajner. 2015. Simplifying Lexical Simplification: Do We Need Simplified Corpora? In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 63–68.

Philipp Koehn and Franz Josef Och and Daniel Marcu 2003. Statistical phrase-based translation. in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, pages 48–54.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, Stroudsburg, PA, USA.

Shashi Narayan and Claire Gardent. 2014. Hybrid Simplification using Deep Semantics and Machine Translation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 435–445.

Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.

Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311–318.

Maja Popović. 2011. Hjerson: An Open Source Tool for Automatic Error Classificatio n of Machine Translation Output. *The Prague Bulletin of Mathematical Linguistics*, pages 59–68, Prague, Czech Republic, October.

Maja Popović, Arle Lommel, Aljoscha Burchardt, Eleftherios Avramidis, Hans Uszkoreit. 2014. Relations between different types of post-editing operations, cognitive effort and temporal effort. In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT 14)*, pages 191–198, Dubrovnik, Croatia.

Maja Popović. 2015. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the 10th Workshop on Statistical Machine Translation*, pages 392–395.

Maja Popović, Mihael Arčan. 2015. Identifying main obstacles for statistical machine translation of morphologically rich South Slavic languages. In *Proceedings of the 18th Annual Conference of the European Association for Machine Translation (EAMT-15)*, pages 97–104, Antalya, Turkey.

Horacio Saggion, Sanja Štajner, Stefan Bott, Luz Rello, Simon Mille and Biljana Drndarević. 2015. Making It Simplext: Implementation and Evaluation of a Text Simplification System for Spanish. *ACM Transactions on Accessible Computing*, Volume 6, Chapter 14.

Advaith Siddharthan. 2011. Text Simplification using Typed Dependencies: A Comparison of the Robustness of Different Generation Strategies. In *Proceedings of the 13th European Workshop on Natural Language Generation (ENLG)*, pages 2–11.

Angrosh Mandya, Tadashi Nomoto and Advaith Siddharthan. 2014. Lexico-syntactic text simplification and compression with typed dependencies. In*Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, Dublin, Ireland, pages 1996–2006.

Lucia Specia. 2010. Translating from complex to simplified sentences. In*Proceedings of the 9th international conference on Computational Processing of the Portuguese Language*, pages 30–39.

Andreas Stolcke. 2002. SRILM – an extensible language modeling toolkit. volume 2, pages 901–904, Denver, CO, September.

Jörg Tiedemann. 2012. Parallel data, tools and interfaces in OPUS. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, pages 2214–2218.

Francis M. Tyers and Murat Alperen. 2010. South-East European Times: A parallel corpus of the Balkan languages. In *Proceedings of the LREC Workshop on Exploitation of Multilingual Resources and Tools for Central and (South-) Eastern European Languages*, pages 49–53, Valetta, Malta.

Received May 2, 2016 , accepted May 12, 2016