

Improving Machine Translation of English Relative Clauses with Automatic Text Simplification

Sanja Štajner

Data and Web Science Group
University of Mannheim
Germany

stajner.sanja@gmail.com

Maja Popović

ADAPT Centre
Dublin City University
Ireland

maja.popovic@adaptcentre.ie

Abstract

This article explores the use of automatic sentence simplification as a pre-processing step in neural machine translation of English relative clauses into grammatically complex languages. Our experiments on English-to-Serbian and English-to-German translation show that this approach can reduce technical post-editing effort (number of post-edit operations) to obtain correct translation. We find that larger improvements can be achieved for more complex target languages, as well as for MT systems with lower overall performance. The improvements mainly originate from correctly simplified sentences with relatively complex structure, while simpler structures are already translated sufficiently well using the original source sentences.

1 Introduction

Text simplification (TS) was initially proposed in the late nineties as a pre-processing step that would improve machine translation (MT), information extraction (IE), and parsing (Chandrasekar et al., 1996). At that time, text simplification was done manually and focused mainly on syntactic transformations. In the last 20 years, many automatic text simplification (ATS) systems were proposed for various languages. Most of them were done with the goal of making texts more understandable to humans. The most mature systems are those proposed for English language. The initial goal of using automatic syntactic simplification for improving MT systems has been forgotten, with the only exception being the recent work of Štajner and Popović (2016), where two lexico-syntactic ATS systems were used for transform-

ing English sentences before translating them into Serbian. The erroneous automatic simplifications were manually corrected before passing them to the MT system. Both ATS systems performed several types of simplifications, but the effects of any particular simplification type were not investigated.

Apart from being the most studied and the most correctly performed type of automatic syntactic simplification, relative clauses are known to pose difficulties for English-to-Serbian (en-sr) and English-to-German (en-de) machine translation, due to target languages being morphologically rich and with different syntactic structures than English. Two examples of English relative clauses problematic for machine translation are shown in Table 1. In the first sentence, the relative pronoun “which” is problematic. The translation is lexically correct in both target languages. However, due to incorrect gender and/or case, it does not relate to the “plot summary” as in the original sentence, but to “Lorax Film” in the German translation and to “Internet Movie Database Website” in the Serbian translation. The second sentence does not have problems directly with the relative pronoun. However, due to its complex structure, the first part of the sentence is problematic for translation into both target languages. In German, there are several mistranslations (the preposition “zu” two times and the verb “bewegen”), and in Serbian, a substantial part of the sentence is missing (the entire beginning marked bold in the English sentence).

In this work, we investigate the impact of simplification of English relative clauses on the quality of en-de and en-sr neural machine translation in three scenarios: (1) using automatic simplifications without any human intervention; (2) using minimal human intervention to filter out bad simplifications, and in those cases, use the origi-

Original	Cameron's submitted text reads in part like a plot summary of the Lorax film provided on the Internet Movie Database website, which begins: "In the walled city ..."
MT: en→de	Camerons Text liest sich teilweise wie eine Zusammenfassung des Lorax Films auf der Website der Internet Movie Database, der beginnt: "In der ummauerten Stadt..."
MT: en→sr	Cameron-ov podnet tekst delimično pročitao kao rezime snimka Lorax filma koji se nalazi na internet stranici Internet Movie Database-a, koja počinje: "U gradskom zidu..."
Original	Rather than having an executive make the announcement , Rita Masoud, a Google employee who fled Kabul with her family when she was seven years old, wrote about her personal experience.
MT: en→de	Anstatt eine Führungskraft zu dieser Ankündigung zu bewegen , schrieb Rita Masoud, eine Mitarbeiterin von Google, die mit ihrer Familie aus Kabul geflohen war, als sie sieben Jahre alt war, über ihre persönlichen Erfahrungen.
MT: en→sr	<missing clause> Rita Masoud, zaposleni u Google-u koji je napustio Kabul sa svojom porodicom, kada je imala sedam godina, pisao je o svom ličnom iskustvu.

Table 1: Examples of English relative clauses problematic for en-de and en-sr machine translation.

nal source sentences instead; (3) using monolingual manual correction of automatic simplifications where necessary. We also explore in which way simplification of relative clauses can improve the quality of translations, and which types of English relative clauses pose problems to machine translation into Serbian and German.

We focus on English-to-Serbian and English-to-German machine translation, as both target languages are morphologically rich and structurally different from English.

2 Related work

2.1 Automatic Text Simplification

Automatic text simplification systems are usually divided into lexical simplification (LS) systems (e.g. (Baeza-Yates et al., 2015; Glavaš and Štajner, 2015; Paetzold and Specia, 2016)), syntactic simplification (SS) systems (e.g. (Siddharthan, 2011; Aranzabe et al., 2012; Glavaš and Štajner, 2013; Brouwers et al., 2014)), and lexico-syntactic simplification (LSS) systems (e.g. (Siddharthan and

Angrosh, 2014; Saggion et al., 2015; Štajner and Glavaš, 2017)). The first group of systems (LS) is only concerned with vocabulary choices and complexity of short phrases (usually unigrams, and sometimes, shorter n -grams). The second group (SS) only simplifies the syntax by splitting long sentences containing relative clauses, coordinate and subordinate structures, transforming passive to active voice, reordering sentence constituents, etc. The third group (LSS) performs both lexical and syntactic simplification at the same time.

The current state-of-the-art lexical simplification systems are unsupervised (Glavaš and Štajner, 2015; Paetzold and Specia, 2016), and although they have a decent coverage (better than the supervised LS systems) they often lead to ungrammatical output or change of original meaning (Štajner and Glavaš, 2017). The changes in meaning are not subtle, but rather essential, and as such, those systems are suitable as a preprocessing step in machine translation only with a manual correction of their output (Štajner and Popović, 2016).

The state-of-the-art syntactic simplification systems are rule-based (Siddharthan and Angrosh, 2014; Saggion et al., 2015), and as such, provide more grammatical output, at the cost of being too conservative and often not making any changes at all. Out of all syntactic simplification operations, simplification of the relative clauses is the most studied and the most reliable one, especially for English. Therefore, in this study, we focus only on this type of transformations hoping to minimize the necessity for manually correcting simplification output.

2.2 ATS for Improving MT

Many works have so far proposed to rewrite input sentences using paraphrasing or textual entailment to improve machine translation, e.g. (Callison-Burch et al., 2006; Mirkin et al., 2009; Aziz et al., 2010; Tyagi et al., 2015). Mirkin et al. (2013a,b) go one step further, proposing an interactive tool which identifies sentences which are most likely to be translated poorly, offers possible rewritings for the human editor, and then performs translation. Although such approach requires some human post-editing effort, the effort is just monolingual (at the source side only). All these approaches, although being proposed and tested on different language pairs (English-French, English-Spanish, English-Hindu), only focus on out-of-

vocabulary words, or difficult to translate shorter n -grams.

The recent work of Štajner and Popović (2016), investigated the impact of lexico-syntactic automatic text simplification systems on English-to-Serbian machine translation. They used two lexico-simplification systems: the EvLex system (Štajner and Glavaš, 2017) which performs sentence splitting, lexical substitution, and content reduction, and a “classical” lexico-syntactic system (Siddharthan and Angrosh, 2014) which performs sentence splittings and lexical substitutions. Similar to Mirkin et al. (2013a), the ATS outputs were manually inspected before feeding them into the MT system. Unlike in the work of Mirkin et al. (2013a) where human editors could just accept or reject suggested simplifications, in the work of Štajner and Popović (2016), human editors were also able to do minor revisions (correcting the tense, gender, article, etc.). Both ATS systems were found to improve fluency of the translations, and reduce the post-editing effort. The influence of particular simplification types (lexical simplification, or different types of syntactic simplification) was not investigated.

3 Methodology

We perform the following experiments:

1. We select a subset of 1000 sentences of the English test set from the WMT 2016 News translation shared task¹, with English as the original source language, focusing only on the sentences which contain relative clauses.
2. We simplify those relative clauses by the state-of-the-art freely available RegenT simplifier (Siddharthan, 2011) and retain only those that were modified by the system (a total of 106 sentences from the initial 1000).
3. We conduct human evaluation of the quality of automatic simplification, and manual correction of automatic simplification where necessary.
4. We use two English-to-Serbian and one English-to-German state-of-the-art machine translation systems to translate our set of

¹<http://www.statmt.org/wmt16/translation-task.html>

score	definition
5	meaning fully preserved no grammatical errors
4	meaning fully preserved minor grammatical errors
3	meaning partially changed grammar not relevant
2	meaning substantially changed grammar not relevant
1	meaning (almost) completely changed grammar not relevant

Table 2: Guidelines for ATS evaluation

106 sentences, their automatic simplifications made by RegenT, and their manually corrected simplifications (in those cases where human correction was necessary).

5. We manually correct the translation output, and use two automatic scores of post-editing effort as the measures of translation quality.
6. We inspect the type of translation improvements achieved with good simplifications, and the type of relative clauses whose good quality simplifications improve or deteriorate the MT output.

3.1 Simplification of Relative Clauses

For automatic simplification of English relative clauses, we use the state-of-the-art RegenT simplifier (Siddharthan, 2011) which is designed for text regeneration tasks such as text simplification, style modification or paraphrasing. The system applies transformation rules (specified in XML files) to a typed dependency representation obtained from the Stanford Parser (De Marneffe et al., 2006). The transformation rules were manually created, and are grouped according to the simplification operation they model: simplifying coordination, subordination, apposition and relative clauses, as well as conversion of passive to active voice. The rule files can be used in combinations or independently; for our experiments, we used only the rules for relative clauses.² The system keeps the entire information in the simplified sentence, it does not tend to remove any information from the original sentence, and as such it is well-suited as a pre-processing step for MT.

The quality assessment was done by three annotators, all three native English speakers. The

²<http://homepages.abdn.ac.uk/cgi-bin/cgiwrap/csc323/RegenT/demo.cgi>, simplified in November 2017.

(1) good “5”	<i>meaning preserved, no grammar errors</i>
original	Both taught in the Division of Social Sciences and History, which lists 17 faculty members, and many students took courses from both.
simplified	Both taught in the Division of Social Sciences and History and many students took courses from both. The Division lists 17 faculty members.
(2) good “4”	<i>meaning preserved, two additions (comma and determiner “this”)</i>
original	Unlike light, which has to be sent down an optic fibre to the desired location inside the brain, low frequency ultrasound waves can pass through tissue unhindered.
simplified	Light, has to be sent down an optic fiber to the desired location inside the brain. Unlike this light, low frequency ultrasound waves can pass through tissue unhindered.
(3) bad “3”	<i>meaning partly changed, some grammatical errors</i>
original	Human breast milk is composed of a variety of proteins, fats, vitamins, and carbohydrates, which give babies all the nutrients they need.
simplified	Human breast milk is composed of a variety and fats and vitamins of proteins, carbohydrates. This variety give babies all the nutrients they need.
(4) bad “2”	<i>meaning changed to a large extent due to lack of negation, no grammar errors</i>
original	There’s no consensus about what the Fed will do, which in itself is causing financial market jitters .
simplified	There’s no consensus about what the Fed will do. This consensus in itself is causing financial market jitters.
(5) bad “1”	<i>meaning changed, low grammaticality</i>
original	A student who praised Lamb , Brandon Beavers, said he also seemed agitated and jittery, ” like there was something wrong with him. ”
simplified	A student praised Lamb , Brandon Beavers , said he also seemed agitated and jittery, ‘like there. This student was something wrong with him.’
(6) bad (“1”)	<i>meaning changed (wrong co-reference), no grammar errors</i>
original	The bubbles, he found, amplify the ultrasonic waves which then pass inside the worms.
simplified	The bubbles, he found, amplify the ultrasonic waves. The bubbles then pass inside the worms.
(7) bad (“1”)	<i>meaning changed (all companies instead of some), no grammar errors</i>
original	Broadly speaking, companies that do the majority of their business in the U.S. will win...
simplified	Companies do the majority of their business in the U.S. Broadly speaking, these companies will win...

Table 3: Examples of good and bad simplifications and their ATS-RC scores. Related elements in a sentence are presented in bold, and erroneous parts in red.

final score was calculated as the arithmetic mean of the three scores, rounded at the closest integer. The inter-annotator agreement, calculated as the weighted Cohen’s kappa, was 0.65, 0.72, and 0.62, respectively.

Seven example sentences with their scores presented in Table 3 illustrate the simplification scores and the mechanism of assigning them.

3.2 Manual Correction of Simplifications

The sentences which were assigned “bad” scores in the previous step, were manually corrected, using the minimal effort for corrections. Similar as in (Štajner and Popović, 2016; Štajner and Glavaš, 2017), the editor (native English speaker) was in-

structed not to introduce any additional simplifications, but rather minimally correct the output so that the original meaning and grammaticality of the sentences are preserved. The second editor (native English speaker) controlled the quality of the corrections.

3.3 Machine Translation

All original, automatically simplified, and corrected English sentences were translated into Serbian and German by the Google translate system³ in February 2018. For the analysis of intrinsic limits of using simplification of English relative clauses as a pre-processing step for NMT, avail-

³<https://translate.google.com/>

ATS-RC score	sentences		group	sentences	
	#	%		#	%
1	37	34.9	bad	56	52.8
2	5	4.7			
3	14	13.2			
4	17	16.0	good	50	47.2
5	33	31.2			

Table 4: Distribution of simplification quality scores (with meaning preservation as the primary criterion, and grammaticality as the secondary).

ability of two distinct target languages is a big advantage, since possible influences of language-related characteristics are reduced. To avoid possible dependencies on the MT system, translations produced by another publicly available NMT system for English-to-Serbian, Asistent⁴, were included in the in-depth analyses (Section 5). In this way, two target languages of the same MT system, as well as two different systems for the same target language were taken into account.

3.4 Evaluation

Although German reference translations were available (Serbian were not, as Serbian is not among the languages investigated at the WMT shared task), using reference translations is not convenient for this type of evaluation since it would penalize too harsh the translations of simplified sentences (especially in the case of syntactic simplification involving sentence splitting and reordering of clauses). The translation outputs were post-edited minimally and the edited translations were used as reference translations to calculate two MT evaluation scores: the character n -gram F-score, chrF (Popović, 2015), and edit distance. The chrF score operates on sub-word level by matching character sequences, and it correlates very well with human direct assessment scores which are, as mentioned in Section 3.1, based mainly on adequacy and partly on fluency (Bojar et al., 2017). Edit distance represents the amount of words which have to be changed in order to transform the translation output into the reference.

4 Results and Discussion

The number and percentage of automatically simplified English clauses with each of the five possible quality scores is presented in Table 4. The sentences were further grouped into two broader

⁴<http://server1.nlp.insight-centre.org/asistent/>

	chrF / edit rate	
	original	simplified
en-sr	85.2 / 15.9	83.0 / 20.8
en-de	93.3 / 7.01	89.7 / 12.8

Table 5: chrF / edit rate for Serbian and German translations of all original English sentences and all their automatic simplifications (higher chrF scores and lower edit rates indicate better translations).

		chrF / edit rate	
		original	simplified
en-sr	good	84.6 / 16.0	86.7 / 14.6
	bad	85.8 / 15.8	79.5 / 26.4
en-de	good	92.6 / 8.05	92.9 / 7.81
	bad	94.0 / 6.04	86.5 / 17.4

Table 6: chrF score / edit rate for translations of good, and bad simplifications of English sentences into Serbian and into German. For each group, better scores are presented in bold.

categories, “good” and “bad”: scores 4 and 5 are considered as good, the rest as bad.

4.1 Impact of Automatic Simplifications

The two MT scores, chrF and edit rate, are presented in Table 5 for the translations of all original and all automatically simplified English sentences (without any quality control or manual corrections). Passing the automatically simplified sentences to MT system, without any quality analysis or manual correction, seems to deteriorate the quality of translations. This can be intuitively expected, since a number of simplifications contains major errors, as shown in Table 4.

The scores for the German translations are better than for Serbian translations, probably due to Serbian being morpho-syntactically more complex language with fewer resources than German.

4.2 Impact of Simplification Quality

To explore the influence of simplification quality on translation quality, MT scores were calculated separately for the translations of good simplifications, and the translations of bad simplifications (Table 6). As expected, the simplification quality of a source sentence has a strong influence on the machine translation output: good simplifications improve the MT scores, whereas bad simplifications clearly deteriorate them. These results indicate that automatic simplification can improve machine translation of English relative clauses into Serbian and German, if we introduce a quick quality check of automatic simplifications, either human (could also be just binary assess-

	chrF / edit rate		
	original	automatic	corrected
en-sr	85.2 / 15.9	83.0 / 20.8	86.4 / 9.4
en-de	93.3 / 7.01	89.7 / 12.8	93.3 / 4.5

Table 7: chrF / edit rate for translations of all original English sentences, all automatic simplifications, and automatic simplifications with manual corrections into Serbian and into German (the higher chrF scores and the lower edit rates, better the translations).

ment as “good”/“bad”) or automatic (automatically checking meaning preservation and grammaticality). Even the first option, the human assessment, improves MT as it requires faster and less demanding (monolingual only) human intervention than post-editing of the MT output.

4.3 Impact of Automatic Simplifications with Manual Corrections

When the bad simplifications are corrected,⁵ the MT scores for Serbian translation output improve, whereas for German they reach the original values by chrF scores, and improve on edit rate scores (Table 7). Taking into account the overall better performance of the English-to-German MT system, the results indicate that ATS is more helpful for translating into more complex and less supported languages (like Serbian).

We further calculated the percentages of improved, deteriorated and unchanged machine translated sentences in terms of both MT evaluation scores (Table 8). The results confirm that the influence of simplification quality is substantial. In English-to-Serbian translation, 84%–88% of bad simplifications deteriorate the translations. At the same time, only 30-50% of correctly simplified source sentences (either directly by the ATS system or by manual correction afterwards), improve the translations. The percentage of improved translations is higher for translations into Serbian, and the percentage of deteriorated translations is slightly higher for translations into German. These results are also consistent with our previous findings (Štajner and Popović, 2016), that only a subset of (correctly) simplified sentences improves the MT output. These results indicate that there are certain limits of current ATS systems when used for MT as the target application. These limitations seem not to be related to the quality of

⁵Erroneous simplifications in our set required technical post-editing effort (edit rate) of 14.2%, of which 9.2% were lexical edits and 5.0% reordering edits.

(a) en→sr

chrF	better	worse	same
good	26 / 52.0%	19 / 38.0%	5 / 10.0%
bad	9 / 16.1%	47 / 83.9%	0 / 0%
corrected	29 / 51.8%	21 / 37.5%	6 / 10.7%

edit rate	better	worse	same
good	24 / 48.0%	21 / 42.0%	5 / 10.0%
bad	7 / 12.5%	49 / 87.5%	0 / 0%
corrected	28 / 50.0%	19 / 33.9%	9 / 16.1%

(c) en→de

chrF	better	worse	same
good	19 / 38.0%	23 / 46.0%	8 / 16.0%
bad	7 / 12.5%	46 / 46.0%	3 / 5.4%
corrected	20 / 35.7%	19 / 33.9%	17 / 30.4%

edit rate	better	worse	same
good	16 / 32.0%	24 / 48.0%	10 / 20.0%
bad	4 / 7.1%	48 / 85.7%	4 / 7.1%
corrected	19 / 33.9%	20 / 35.7%	17 / 30.4%

Table 8: Number / percentage of improved, deteriorated and unchanged machine translated sentences in terms of the chrF score (above) and edit rate (below). Results for translations of correct (good and corrected) simplifications are presented in bold.

produced simplifications, because in all scenarios only a subset of correctly simplified sentences improves the MT output.

5 In-Depth Analysis

In order to explore the limits of simplification of English relative clauses for improving MT systems, we manually analyzed translations of all good and corrected simplifications. In this set of experiments, we used an additional English-to-Serbian MT system, as explained in Section 3.

Table 9 shows the amount of improved, deteriorated and unchanged translations when translating only the correctly simplified source sentences (either being correctly automatically simplified, or being manually corrected). For both English-to-Serbian MT systems, about a half of the simplified sentences improves the MT scores, whereas for English-to-German, improvement is achieved for only about one third of sentences. These results indicate that it is difficult to improve a very strong MT system by simplifying relative clauses. Surprisingly, even for the system with the lowest overall performance (Asistent), half of the correctly simplified sentences exhibit worse or unchanged MT scores.

In order to get more details about the two groups

chrF	better	worse	same
sr (Google)	55 / 51.9%	40 / 37.7%	11 / 10.4%
sr (Asistent)	51 / 48.1%	54 / 50.9%	1 / 1.0%
de (Google)	39 / 36.8%	42 / 39.6%	25 / 23.6%

edit rate	better	worse	same
sr (Google)	52 / 49.0%	40 / 37.7%	14 / 13.2%
sr (Asistent)	53 / 50.0%	51 / 48.1%	2 / 1.9%
de (Google)	35 / 33.0%	44 / 41.5%	27 / 25.5%

Table 9: Number / percentage of improved, deteriorated and unchanged translations in terms of the chrF score (above) and edit rate (below) for translation of good and corrected simplifications.

of translation outputs, improved and worsened, we performed error classification using Hjerison (Popović, 2011). Hjerison classifies the errors into five categories: inflection, order, omission, addition and mistranslation, but with a high level of confusions between omissions, additions and mistranslations. Therefore we applied the same tactic as Toral and Sánchez-Cartagena (2017), merging additions, omissions and mistranslations into one “lexical” category. The three classes of edit rates are presented in Table 10.

All three error categories are improved in “better” translations and deteriorated in “worse” translations. For the system with high overall MT score (Google), the largest changes are in the number of lexical errors. For the system with lower overall MT score (Asistent), the changes in reordering (syntactic) errors are larger and the changes in lexical errors smaller than for the better performing system (Google). Grammatical errors in the Asistent translations are much more frequent than in the Google translations, and these errors can be reduced by syntactic simplification of relative clauses. The amount of errors in translations of original versions of “better” sentences is higher than for “worse” sentences. This suggests that the MT systems can already handle the “worse” sentences sufficiently well, so that the simplification only introduces confusion which results in increased number of lexical errors.

These error rates shed some light on differences between improved and worsened translation outputs, but they did not provide any information about the corresponding source sentences.

We investigated what the source sentences (correct simplifications), both those that improve and those that deteriorate MT output, have in common regardless of the MT system and the target language. The number of such overlapping source

three types of edit rates (%)		better		worse	
		orig.	simp.	orig.	simp.
sr (Google)	inflection	5.8	3.5	2.5	3.7
	order	2.0	1.2	0.9	1.7
	lexical	12.3	8.3	9.8	13.2
sr (Asistent)	inflection	8.0	7.8	7.7	7.9
	order	8.0	6.6	6.7	7.2
	lexical	32.9	30.9	30.2	32.4
de (Google)	inflection	1.2	0.5	0.4	1.2
	order	1.2	0.4	0.7	1.5
	lexical	9.6	3.8	4.1	9.2

Table 10: Three classes of edit rates (inflectional, ordering and lexical) for improved and deteriorated translations when translating good and corrected simplifications. For each group, better scores are presented in bold.

	better	worse	same
sr (Google) \cap sr (Asistent)	27	20	1
sr (Google) \cap de (Google)	21	23	6
sr (Asistent) \cap de (Google)	18	19	1

Table 11: Number of source sentences whose simplification improves/deteriorates/does not change the MT scores for different MT systems. The numbers in parentheses denote the number of corresponding sentences in each of the two involved translation outputs.

sentences between each pair of translation outputs is presented in Table 11. The smallest overlap can be noted between German Google translations and Serbian Asistent translations, which can be expected since in this case both the target language and the MT system differ.

Several examples of improved and deteriorated sentences are presented in Table 12. Relatively simple structures where the relative pronoun, or determiner, almost immediately follows its corresponding noun are already well handled by MT systems. Simplifying such structures only introduces disturbances, which are mostly manifested in the form of increased number of lexical errors (see Table 10). More complex structures with distant relative pronouns and/or more than one possible co-reference are more difficult to translate correctly and these are the structures where simplification of relative clauses generally helps, independently of the language pair and the MT system.

Table 13 represents the most frequent POS 4-grams for the source sentences which lead to “better” and “worse” translations. Both tables clearly indicate that the structure of the sentences in the two groups differs.

(a) English sentences for which TS improves the MT scores

orig.	A student who praised Lamb, Brandon Beavers , said he also seemed agitated and jittery, "like there was something wrong with him."
simp.	A student Brandon Beavers who praised Lamb, said he also seemed agitated and jittery," like there was something wrong with him."
orig.	Cameron's submitted text reads in part like a plot summary of the Lorax film provided on the Internet Movie Database website, which begins: "In the walled city of Thneed-Ville, where everything is artificial and even the air is a commodity, a boy named Ted hopes to win the heart of his dream girl, Audrey."
simp.	Cameron's submitted text reads in part like a plot summary of the Lorax film provided on the Internet Movie Database website. The summary begins: 'In the walled city of Thneed-Ville, where everything is artificial and even the air is a commodity, a boy named Ted hopes to win the heart of his dream girl, Audrey.'
orig.	Rather than having an executive make the announcement, Rita Masoud, a Google employee who fled Kabul with her family when she was seven years old, wrote about her personal experience.
simp.	A Google employee fled Kabul with her family when she was seven years old. Rather than having an executive make the announcement, Rita Masoud, this employee , wrote about her personal experience.

(b) English sentences for which TS deteriorates the MT scores

orig.	Experts believe shoppers could be holding off making purchases ahead of the event, which takes place on the last Friday in November.
simp.	Experts believe shoppers could be holding off making purchases ahead of the event . The event takes place on the last Friday in November.
orig.	The tiny nematodes change direction the moment they are blasted with sonic pulses that are too high-pitched for humans to hear.
simp.	The tiny nematodes change direction the moment they are blasted with sonic pulses . These pulses are too high-pitched for humans to hear.
orig.	Human breast milk is composed of a variety of proteins, fats, vitamins, and carbohydrates, which give babies all the nutrients they need.
simp.	Human breast milk is composed of a variety of proteins, fats, vitamins, and carbohydrates. This variety gives babies all the nutrients they need.

Table 12: Examples of English source sentences whose simplification (a) improves MT scores for distinct MT systems and (b) deteriorates MT scores for distinct MT systems.

effect on MT scores:	
better	worse
. N N , N PREP N , PREP DET ADJ N DET ADJ N PREP	PREP DET N PREP N , WH-DET V-PRES N PREP DET N DET N PREP N

Table 13: Most frequent POS 4-grams in the two groups of overlapping original English source sentences.

6 Summary and outlook

In this work, we showed (on a small data set) that the automatic simplification of English relative clauses can improve English-to-Serbian and English-to-German machine translation (MT) if used as a pre-processing step before translating the sentences with a neural machine translation (NMT) system, only if used with the quality control of the simplifications, or some minimal manual correction of the simplifications. We found that such simplifications improve the output of Google's English-to-Serbian and English-to-German MT mostly by decreasing the number

of lexical errors, while the output of the lower performing English-to-Serbian NMT system (Asistent) mostly benefit from a decreased number of reordering errors. We also found that both target languages and both MT systems share the patterns of relative clauses whose simplification improves the translations. The described limitations of using simplification of English relative clauses for improving MT output are not surprising: the state-of-the-art ATS systems were tailored for improving comprehension of texts by different target users. Those transformations do not necessarily coincide with improving machine translation. An important direction for future work is to develop ATS systems which are tailored for structures problematic for MT.

Acknowledgements: This research was supported by the ADAPT Centre for Digital Content Technology at Dublin City University, funded under the Science Foundation Ireland Research Centres Programme (Grant 13/RC/2106) and co-

funded under the European Regional Development Fund.

References

- María Jesús Aranzabe, Arantza Díaz De Ilarraza, and Itziar González. 2012. First Approach to Automatic Text Simplification in Basque. In *Proceedings of the first Natural Language Processing for Improving Textual Accessibility Workshop (NLP4ITA 2012)*, Istanbul, Turkey.
- Wilker Aziz, Marc Dymetman, Shachar Mirkin, Lucia Specia, Nicola Cancedda, and Ido Dagan. 2010. Learning an expert from human annotations in statistical machine translation: the case of out-of-vocabulary words. In *Proceedings of EAMT*.
- Ricardo Baeza-Yates, Luz Rello, and Julia Dembowski. 2015. Cassa: A context-aware synonym simplification algorithm. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistic and Human Language Technologies (NAACL-HLT 2015)*, pages 1380–1385, Denver, Colorado.
- Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Aurelie Neveol, Mariana Neves, Martin Popel, Matt Post, Raphael Rubino, Carolina Scarton, Lucia Specia, Marco Turchi, Karin Verspoor, and Marcos Zampieri. 2016. Findings of the 2016 Conference on Machine Translation. In *Proceedings of the First Conference on Machine Translation (WMT 2016)*, pages 131–198, Berlin, Germany.
- Ondřej Bojar, Yvette Graham, and Amir Kamran. 2017. Results of the WMT17 Metrics Shared Task. In *Proceedings of the Second Conference on Machine Translation (WMT 2017)*, pages 489–513, Copenhagen, Denmark.
- Laetitia Brouwers, Delphine Bernhard, Anne-Laure Ligozat, and Thomas François. 2014. Syntactic sentence simplification for french. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 47–56, Gothenburg, Sweden.
- Chris Callison-Burch, Philipp Koehn, and Miles Osborne. 2006. Improved statistical machine translation using paraphrases. In *Proceedings of the 2016 Human Language Technology Conference of the North American Chapter of the ACL (HLT-NAACL)*, pages 17–24.
- Raman Chandrasekar, Christine Doran, and B. Srinivas. 1996. Motivations and Methods for Text Simplification. In *Proceedings of COLING 1996*, pages 1041–1044, Copenhagen, Denmark.
- Marie-Catherine De Marneffe, Bill McCartney, and Christopher Manning. 2006. Generating Typed Dependency Parses from Phrase Structure Parses. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, pages 449–454, Genoa, Italy.
- Goran Glavaš and Sanja Štajner. 2013. Event-Centered Simplification of News Stories. In *Proceedings of the Student Workshop at RANLP 2013*, pages 71–78, Hissar, Bulgaria.
- Goran Glavaš and Sanja Štajner. 2015. Simplifying Lexical Simplification: Do We Need Simplified Corpora? In *Proceedings of the ACL&IJCNLP 2015 (Volume 2: Short Papers)*, pages 63–68, Beijing, China.
- Shachar Mirkin, Lucia Specia, Nicola Cancedda, Ido Dagan, Marc Dymetman, and Idan Szpektor. 2009. Source-language entailment modeling for translating unknown terms. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2, ACL '09*, pages 791–799, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Shachar Mirkin, Sriram Venkatapathy, and Marc Dymetman. 2013a. Confidence-driven Rewriting for Improved Translation. In *Proceedings of the XIV MT Summit, Nice, France*, pages 257–264.
- Shachar Mirkin, Sriram Venkatapathy, Marc Dymetman, and Ioan Calapodescu. 2013b. SORT: An Interactive Source-Rewriting Tool for Improved Translation. In *Proceedings of ACL, Sofia, Bulgaria*, pages 85–90.
- Gustavo Henrique Paetzold and Lucia Specia. 2016. Unsupervised lexical simplification for non-native speakers. In *Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016)*, Phoenix, Arizona.
- Maja Popović. 2011. Hjerson: An Open Source Tool for Automatic Error Classification of Machine Translation Output. *The Prague Bulletin of Mathematical Linguistics*, (96):59–68.
- Maja Popović. 2015. chrF: Character n-gram F-score for Automatic MT Evaluation. In *Proceedings of the 10th Workshop on Statistical Machine Translation (WMT 2015)*, pages 392–395, Lisbon, Portugal.
- Horacio Saggion, Sanja Štajner, Stefan Bott, Simon Mille, Luz Rello, and Biljana Drndarevic. 2015. Making It Simplex: Implementation and Evaluation of a Text Simplification System for Spanish. *ACM Transactions on Accessible Computing*, 6(4):14:1–14:36.

- Advait 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 2011)*, page 2–11, Nancy, France.
- Advait Siddharthan and M. A. Angrosh. 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 2014)*, pages 722–731, Gothenburg, Sweden.
- Antonio Toral and Víctor Manuel Sánchez-Cartagena. 2017. A Multifaceted Evaluation of Neural versus Statistical Machine Translation for 9 Language Directions. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017)*, Valencia, Spain.
- Shruti Tyagi, Deepti Chopra, and Iti Mathur. 2015. Classifier based text simplification for improved machine translation. In *Proceedings of International Conference on Advances in Computer Engineering and Applications (ICACEA)*, Ghaziabad, India, pages 46–50.
- Sanja Štajner and Goran Glavaš. 2017. Leveraging event-based semantics for automated text simplification. *Expert Systems with Applications*, 82:383–395.
- Sanja Štajner and Maja Popović. 2016. Can text simplification help machine translation? In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation (EAMT 2016)*, pages 230–242, Riga, Latvia.