A Comparison of Statistical and Neural Machine Translation for Slovene, Serbian and Croatian

Mihael Arčan

Data Science Institute
National University of Ireland Galway
IDA Business Park, Lower Dangan, Galway
mihael.arcan@insight-centre.org

Abstract

In this paper we present a comparison of translation quality using of Statistical Machine Translation (SMT) and Neural Machine Translation (NMT), considering translation directions between English, Slovene, Serbian and Croatian. Our experiments show that on a reduced training dataset with around two million sentences, SMT outperforms the NMT neural models. Furthermore, we present experiments with enlarged neural architectures, using 1,000 nodes and 4 hidden layers, which shows improved translation quality in terms of the BLEU metric.

1. Introduction

Although automatically generated translations using machine translation approaches are far from perfect, studies have shown significant productivity gains when human translators are supported by machine translation output rather than starting a translation task from scratch (Federico et al., 2012; Läubli et al., 2013; Green et al., 2013).

Due to the large success of NMT in recent years (Kalchbrenner and Blunsom, 2013; Bahdanau et al., 2014; Sutskever et al., 2014), we evaluate its translation performance against the usage of SMT, focusing on translation direction between English, Slovene, Serbian and Croatian. Figure 1 illustrates how a sequence-to-sequence neural network used in our experiments can be trained on parallel data. First, a sequence-to-sequence framework reads a source sentence using an encoder to build a dense vector, a sequence of non-zero values that represents the meaning of the source sentence. A decoder processes this vector to predict a translation of the input sentence. In this manner, these encoder-decoder models can capture long-range dependencies in languages, e.g., gender agreements or syntax structures. The challenges involved with the less supported Slavic languages (Krek, 2012) lie in the morphological complexity for all word classes. Furthermore, these languages have rather a free word order and are highly inflected. There are six distinct cases affecting not only common nouns but also proper nouns as well as pronouns, adjectives and some numbers. Some nouns and adjectives have two distinct plural forms depending on the number. There are also three genders for the nouns, pronouns, adjectives and some numbers leading to differences between the cases and also between the verb participles for past tense and passive voice.

Since training a neural translation model is computational expensive, we first limit the data used to train the NMT models to two million parallel sentences. In the next experiment, we then extend the parallel corpus for the English-Slovene language pair to evaluate how the neural architecture, number of nodes and hidden layers, affect the translation quality.

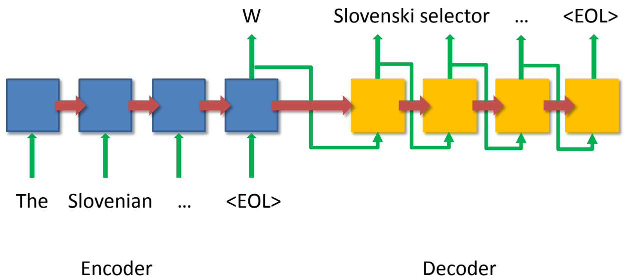

Figure 1: Neural network with the encoder-decoder architecture.

Finally, the neural models trained during this work are publicly accessible through the *Asistent* system[1] (Arčan et al., 2016), an SMT system, which enables automatic translations between English, Slovene, Croatian and Serbian language.

The remainder of the paper is organised as follows: Section 2. gives an overview of the related work on machine translation for the targeted south Slavic languages. Section 3. describes the methods of reducing the number of parallel sentences and subword unit transformation. After this, we give insights on the used parallel resources, translation frameworks and evaluation methods in Section 4. In Section 5. the results of our experiments described in the previous section are illustrated. Finally, we conclude our findings and give an outlook for our further research.

2. Related Work

One of the first results with automatic translations for Slovene was shown in the *Presis* System (Romih and Holozan, 2002). The rule-based translation system annotates each source sentence with grammatical features and uses built-in rules for converting annotated source sentences into the target language.

First publications dealing with SMT systems for

[1] `http://server1.nlp.insight-centre.org/asistent/`

Serbian-English (Popović et al., 2005) and Slovene-English (Maučec et al., 2006) are reporting results using small bilingual corpora. Using morpho-syntactic knowledge for the Slovene-English language pair was shown to be useful for both translation directions in Žganec Gros and Gruden (2007). However, no analysis of results has been carried out in terms of what actual problems were caused by the rich morphology and which of those were solved by the morphological preprocessing. Recent work in SMT also deals with the Croatian language, which is very closely related to Serbian. First results for Croatian-English are reported in Ljubešić et al. (2010) on a small weather forecast corpus, and an SMT system for the tourist domain is presented in Toral et al. (2014). Furthermore, SMT systems for both Serbian and Croatian are described in Popović and Ljubešić (2014) and more recently in Toral et al. (2016) and Sánchez-Cartagena et al. (2016). Work on rule based machine translation between Croatian and Serbian was shown in Klubička et al. (2016).

Different SMT systems for subtitles were developed in the framework of the SUMAT project, including Serbian and Slovene (Etchegoyhen et al., 2014). First effort in the direction of collecting a larger amount of existing parallel datasets for Serbian and Slovene was carried out in Popović and Arcan (2015). The authors built several SMT systems in order to identify the most important language related issues which may help to build better translation systems. However, all the translation systems described were built and used only locally, mainly only on one particular genre and/or domain. In this proposed work, we are building a publicly available mixed-domain SMT system built on existing parallel corpora, which we believe will be useful for the given under-resourced language pairs.

Popovic et al. (2016a) perform a systematic evaluation of MT results between Croatian, Serbian and Slovenian on the differences between the structural properties represent the most prominent issue for all translation directions between the Slavic languages. For translations between Croatian and Serbian, the constructions involving the verb *trebati* (en. should/need) definitely represent the larger obstacle for both translation directions and for both MT approaches, statistical as well as rule-based.

Maučec and Brest (2017) present an overview of numerous relevant works and the main issues on SMT of highly inflectional Slavic languages. The authors give insights on the most difficulties related to inflectional richness and relaxed word order. Furthermore they stress big differences between translation from a highly inflectional language and translation to a highly inflectional language. The research has shown that simple reduction of rich morphology of Slavic language does not improve translation to English, because some important information is also lost. Translation to a highly inflectional language poses a question about morphological features of target words as they are not evident from morphologically less rich source language. In this sense taking source context in account and additional tagging of source text, based on target language, shows promising results. Manojlović et al. (2017) investigate the treatment of idioms in state-of-the art SMT systems involving English and Croatian. The authors construct three short stories abundant with idioms per each language, and translate them into the other language by two state-of-the-art SMT systems. They manually inspect the outputs and present results and devise an error taxonomy for handling idioms. Popović et al. (2016b) demonstrate that a small amount of in-domain training data is very important for the translation quality for the English-to-Croatian SMT of the specific genre of Massive Open Online Courses (MooC), especially for capturing appropriate morpho-syntactic structures. Adding in-domain parallel data containing the closely related Serbian language improves the performance, especially when the Serbian part is translated into Croatian thus producing an artificial English-Croatian in-domain corpus. The improvements consist mainly from reducing the number of lexical errors. Further improvements have been achieved by adding a relatively large out-of-domain news corpus reaching performance comparable with systems trained on much larger (out-of-domain) parallel texts. The authors show that adding this corpus reduces the number of additions and lexical errors, nevertheless it introduces more morphological and ordering errors due to the different nature and structure of the segments.

3. Methodology

In this section, we describe the data selection approach of finding relevant sentences within parallel data. Due to the large vocabulary of morphological rich languages, we use subword unit NMT models instead of word-based models and give therefore insights into *Byte Pair Encoding* to minimise the out-of-vocabulary (OOV) issue.

3.1. Relevant Sentence Selection

Due to the computational complexity of training NMT sequence-to-sequence models, we experiment on minimising the set of relevant parallel sentences. Therefore, we start selecting parallel sentences for each language pair from the set of all sentences and select those containing words, which do not appear in the set of previously selected sentences. We repeat this step till we obtain a corpus with the targeted size. With this approach, we plan to exclude duplicate sentences and minimise the set of very similar sentences into the training dataset. For this initial setting, we perform the selection approach for all language pairs limiting the training dataset to two million sentences. Furthermore, focusing only on the English-Slovene language pair, we also generate a parallel corpus with five million sentences.

3.2. Byte Pair Encoding

A common problem in machine translation, in general, are rare and unknown words, e.g. terminological expressions, which the system has rarely or never seen. Therefore, if the training method does not see a specific word or phrase multiple times during training, it will not learn the correct translation. This challenge is even more evident in NMT due to the complexity associated with neural networks. Therefore the vocabulary is often limited only to 50,000 or 100,000 words (in comparison to 200,000 or more words in a two million corpus). To overcome this

Sentence	Tokenised Sentence	BPE Segmented Sentence
procedures for proper handling and disposal of antineoplastic medicinal products should be used.	procedures for proper handling and disposal of antineoplastic medicinal products should be used .	procedures for proper handling and disposal of ant■ ine■ op■ lastic medicinal products should be used ■.
many thanks to mrs van lancker, mr berman, mr lambsdorff, mr hutchinson, mrs scheele, mrs doyle, mrs weber, mr varvitsiotis, mrs hassi and mrs gomes.	many thanks to mrs van lancker , mr berman , mr lambsdorff , mr hutchinson , mrs scheele , mrs doyle , mrs weber , mr varvitsiotis , mrs hassi and mrs gomes .	many thanks to mrs van lanc ■ ker ■, mr berman ■, mr lamb■ s■ dor■ ff ■, mr hut■ chin■ son ■, mrs sche■ ele ■, mrs doyle ■, mrs weber ■, mr var■ vit■ si■ otis ■, mrs hass■ i and mrs g■ omes ■.
kdo vam daje pravico, da invalidom odrekate neomejen dostop do izobraževanja, ali da starejšim ljudem odrekate enako obravnavo pri zavarovanjih in finančnih storitvah?	kdo vam daje pravico , da invalidom odrekate neomejen dostop do izobraževanja , ali da starejšim ljudem odrekate enako obravnavo pri zavarovanjih in finančnih storitvah ?	kdo vam daje pravico ■, da invali■ dom odre■ kate ne■ omejen dostop do izobraževanja ■, ali da stare■ jšim ljudem odre■ kate enako obravnavo pri zavarov■ anjih in finančnih storitvah ■?

Table 1: Examples of tokenised sentences and subword unit (BPE) segmentation.

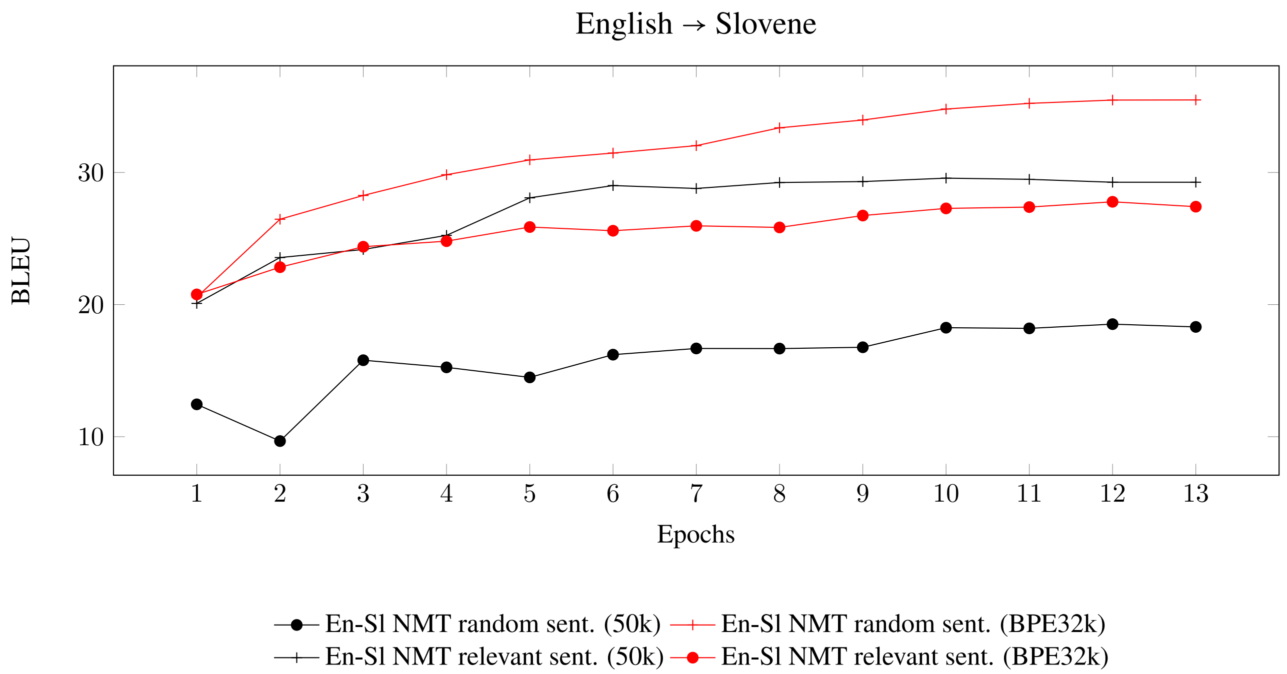

Figure 2: Comparison of the evaluation dataset with NMT systems using word level and subword units.

limitation, different methods were suggested, i.e. character based NMT (Costa-Jussà and Fonollosa, 2016; Ling et al., 2015) or using subword units, e.g. Byte Pair Encoding (BPE). The latter one was successfully adapted for word segmentation specifically for the NMT scenario Sennrich et al. (2015). BPE (Gage, 1994) is a form of data compression that iteratively replaces the most frequent pair of bytes in a sequence with a single, unused byte. Instead of merging frequent pairs of bytes as shown in the original algorithm, characters or character sequences are merged for the purposes of NMT. To achieve this, the symbol vocabulary is initialised with the character vocabulary, and each word is represented as a sequence of characters, plus a special end-of-word symbol (■), which allows restoring the original tokenisation after the translation step. This process is repeated as many times as new symbols are created. Table 1 shows the differences between word-based tokenisation and subword unit (BPE) segmentation, while Figure 2 shows the translation quality improvement of subword unit models and the word-based models in terms of the BLEU metric within the span on 13 epochs using a parallel corpus with two million sentences.

4. Experimental Setting

In this Section, we give an overview on the datasets and the translation toolkits used in our experiment. Furthermore, we give insights into the evaluation techniques, considering the translation directions between English and

	L1 Lang.	-	L2 Lang.	Sentences	L1 Language Tokens	L1 Language Types	L2 Language Tokens	L2 Language Types
Training Dataset	English	-	Slovene	2,299,805	37,849,280	631,114	33,379,920	587,018
	English	-	Croatian	2,464,895	34,880,415	626,353	29,744,620	593,672
	English	-	Serbian	2,152,740	27,465,108	658,660	23,536,540	573,241
	Croatian	-	Serbian	2,177,242	22,717,151	1,200,577	22,790,166	1,170,801
	Slovene	-	Croatian	2,004,229	17,759,047	464,392	18,150,733	545,023
	Slovene	-	Serbian	2,131,301	20,515,466	769,019	21,257,242	883,175
Development Dataset	English	-	Slovene	2,017	38,280	32,918	10,092	13,650
	English	-	Croatian	2,114	45,605	40,536	8,264	12,638
	English	-	Serbian	2,092	37,757	35,419	10,373	14,017
	Croatian	-	Serbian	2,000	14,716	14,774	4,601	4,620
	Slovene	-	Croatian	2,000	14,339	14,447	3,924	4,215
	Slovene	-	Serbian	2,000	12,985	13,575	3,890	4,060
Evaluation Dataset	English	-	Slovene	2,015	44,559	39,561	7,414	10,972
	English	-	Croatian	2,113	45,768	40,462	8,218	12,727
	English	-	Serbian	2,036	40,349	37,346	6,833	10,866
	Croatian	-	Serbian	2,000	12,805	13,043	3,909	3,984
	Slovene	-	Croatian	2,000	13,799	14,187	3,794	4,109
	Slovene	-	Serbian	2,000	13,090	13,606	3,900	4,138

Table 2: Statistics on parallel data used for the training, development and evaluation set (tokens = running words; types = unique words).

the targeted Slavic languages.

4.1. Training Datasets

The parallel data used to train the translation systems were mostly obtained from the OPUS web site (Tiedemann, 2012), which contains various corpora, i.e. DGT, ECB, EMEA, Europarl, KDE among others, of different sizes and domains. For the Serbian-English language pair, a small language course corpus of about 3,000 sentence pairs was added as well. Furthermore, a small phrase book with about 1,000 entries was added to the Slovene-Serbian training set. From the set of the available corpora, we select relevant sentences limiting the corpus to a specific size.

Table 2 illustrates the amount of data used to train, tune and evaluate our translation models. The upper part of the table shows the number of parallel entries used to train the translation models, considering the data selection approach (cf. Subsection 3.1.). While corpora for the English-Slavic language pairs consist of different domains, e.g. legal, medical, financial, IT, parallel data between Slavic language pairs consist mostly out of the OpenSubtitles corpus (Lison and Tiedemann, 2016).[2]

4.2. Evaluation Datasets

The dataset used for evaluating the translation performance consists of around 2.000 sentences for each language pair of various domains.[3] When translating from or into English, sentences from different corpora[4] were added to the evaluation dataset (isolated from the training dataset). The data used for evaluating translations between the Slavic languages consist mostly out of the OpenSubtitles corpus since this corpus builds the largest part (≈95%) of the data used to train the translation models.

4.3. Machine Translation tools

For our SMT translation task, we use the statistical translation toolkit **Moses** (Koehn et al., 2007), where the word alignments were built with the GIZA++ toolkit (Och and Ney, 2003). The KenLM toolkit (Heafield, 2011) was used to build a 5-gram language model.

OpenNMT (Klein et al., 2017) is a generic deep learning framework mainly specialised in sequence-to-sequence (seq2seq) models covering a variety of tasks such as machine translation, summarisation, image to text, and speech recognition. We used the default OpenNMT parameters, i.e. 2 layers, 500 hidden bidirectional LSTM[5] units, input feeding enabled, batch size of 64, 0.3 dropout probability and a dynamic learning rate decay. We train the networks for 13 epochs and report the results in Section 5.

In addition to the default setting, we perform for the English-Slovene language pair experiments on larger neural networks, extending LSTM to 1,000 hidden units and increase the network to 4 layers.

4.4. Evaluation Metrics

The automatic translation evaluation is based on the correspondence between the SMT output and reference translation (gold standard). For the automatic evaluation

[2] `http://www.opensubtitles.org/`

[3] The evaluation set can be obtained under: `http://server1.nlp.insight-centre.org/asistent/data/asisten_evaluation_set.tar.gz`

[4] DGT, EMEA, Europarl, KDE and OpenSubtitles for English-Slovene; DGT, hrenWaC, KDE, OpenSubtitles and SETimes for English-Croatian; KDE, OpenSubtitles and SETimes for English-Serbian

[5] LSTM -Long Short Term Memory

	SMT			NMT		
	BLEU	Meteor	chrF	BLEU	Meteor	chrF
English → Slovene	37.35	29.92	60.74	27.41	24.83	53.02
Slovene → English	46.02	37.42	64.61	27.80	29.28	52.76
English → Croatian	32.44	28.03	57.46	20.94	21.69	49.73
Croatian → English	37.49	35.26	60.66	23.73	26.67	50.41
English → Serbian	31.30	27.33	55.28	17.44	20.24	45.89
Serbian → English	32.49	34.18	57.82	20.64	25.00	48.02
Slovene → Serbian	19.37	22.09	41.52	19.95	21.35	40.99
Serbian → Slovene	21.51	22.82	43.46	21.68	22.30	43.10
Slovene → Croatian	21.29	22.93	43.97	19.54	21.01	40.60
Croatian → Slovene	25.94	25.39	47.60	24.77	24.12	45.71
Serbian → Croatian	68.96	48.24	79.96	61.06	42.76	76.85
Croatian → Serbian	68.15	46.70	78.14	64.25	43.84	76.83

Table 3: Automatic evaluation of translation quality for all targeted language pairs using two million sentences, selected based on new vocabulary.

we used the BLEU (Papineni et al., 2002), METEOR (Denkowski and Lavie, 2014) and chrF (Popović, 2015) metrics.

BLEU (Bilingual Evaluation Understudy) is calculated for individual translated segments (n-grams) by comparing them with a dataset of reference translations. Those scores, between 0 and 100 (perfect match), are then averaged over the whole *evaluation dataset* to reach an estimate of the translation's overall quality.

METEOR (Metric for Evaluation of Translation with Explicit ORdering) is based on the harmonic mean of precision and recall, whereby recall is weighted higher than precision. Along with exact word (or phrase) matching it has additional features, i.e. stemming, paraphrasing and synonymy matching. In contrast to BLEU, the metric produces good correlation with human judgement at the sentence or segment level.

chrF3 is a character n-gram metric, which has shown very good correlations with human judgements on the WMT2015 shared metric task (Stanojević et al., 2015), especially when translating from English into morphologically rich(er) languages.

The approximate randomization approach in MultEval (Clark et al., 2011) is used to test whether differences among system performances are statistically significant with a p-value < 0.05.

5. Evaluation

In this Section, we report the translation quality based on the evaluation datasets generated with the SMT and NMT models. Additionally, we perform experiments extending the training data to five millions entries as well as extending the neural architecture for the English-Slovene language pair.

5.1. Translation Evaluation Based on Two Million Relevant Sentences

As a first evaluation, we automatically compare the translations generated by SMT and subword NMT models, trained on two million selected relevant sentences. As seen in Table 3, the results show a better performance of the SMT system in almost all translation directions. Only when translating from Slovene into Serbian, the NMT system performs statistically significantly ($p < 0.05$) better in comparison to the SMT generated sentences. The lower translation quality generated by the NMT system can be explained due to the relevant sentence and vocabulary selection (see Table 5), while SMT can better handle the high vocabulary density in the parallel corpus of two million sentence used for training. On the other hand, we observed that data selection can be beneficial within the SMT approach. If a corpus of two million random sentences was selected to train an SMT model, the BLEU score drops from 37.35 to 33.43, when translating from English to Slovene and from 46.02 to 40.71, when translating from Slovene into English.

	BLEU	Meteor	chrF
Slovene-Serbian	5.30	11.48	25.61
Serbian-Slovene	5.24	11.97	26.61
Slovene-Croatian	4.80	11.51	26.34
Croatian-Slovene	4.76	11.98	27.80
Serbian-Croatian	66.78	46.70	78.53
Croatian-Serbian	67.20	46.08	77.85

Table 4: Language similarities based on the BLEU, Meteor and chrF metric.

The high evaluation scores in Table 3 between Croatian and Serbian can be explained due to the language similarities between these two languages. Table 4 illustrates the similarity based on the vocabulary, as well as on the character level for the three south Slavic languages. The scores were calculated in a manner that the source text of the evaluation dataset was treated as the generated translation of the translation system and compared with the target side of the evaluation set. We observed that Slovene is expected less similar to Serbian and Croatian, whereby Serbian and Croatian show high similarity even without any translation approaches. The SMT generated translations are, neverthe-

	2M			5M		
	Sentences	Tokens	Types	Sentences	Tokens	Types
random (English)	2,000,000	25,496,017	243,512	4,980,012	61,802,915	386,260
random (Slovene)	2,000,000	21,175,388	415,965	4,980,012	51,219,305	640,471
compressed (English)	2,299,805	37,849,280	631,114	5,003,508	85,675,760	618,112
compressed (Slovene)	2,299,805	33,379,920	587,018	5,003,508	72,868,553	964,763

Table 5: Statistics on the two and five million parallel corpus used to train the English-Slovene translation systems (tokens = running words; types = unique words).

	2M			2M++			5M			5M++		
English → Slovene	BLEU	Meteor	chrF	BLEU	Meteor	chrF	BLEU	Meteor	chrF	BLEU	Meteor	chrF
random	35.49	29.47	60.52	35.26	29.27	59.52	36.60	29.94	60.33	38.01	30.75	61.54
compressed	27.41	24.83	53.02	29.30	25.75	54.51	37.56	30.57	61.75	38.75	31.19	62.58
Slovene → English	BLEU	Meteor	chrF	BLEU	Meteor	chrF	BLEU	Meteor	chrF	BLEU	Meteor	chrF
random	38.77	35.29	61.15	36.86	33.90	59.31	36.72	33.90	59.58	38.02	34.08	59.70
compressed	27.80	29.28	52.76	28.76	29.53	53.19	37.71	34.38	60.09	40.30	35.61	62.16

Table 6: Automatic translation evaluation based on different parallel corpora and network architecture.

less, statistically significantly better than a direct comparison.

5.2. Translation Evaluation on Extended Neural Networks

Due to the low performance of the subword unit NMT models, we experimented with extending the training data for the English-Slovene language pair to five million sentences. Within this experiment, we first randomly selected five million sentences and secondly identified five million relevant sentences with a high vocabulary density, as described in Section 3.1. Furthermore, we extended the LSTM neural network architecture to 1,000 nodes and use 4 layers in the network. Table 5 illustrates the vocabulary change of the datasets based on randomly selected sentences and the identification of relevant sentences based on newly seen vocabulary. As seen, the approach increases the vocabulary of the two million corpus (2M in Table 5) from around 200,000 to more than 600,000 unique words (types) for English, and from 400,000 to almost 590,000 unique words for Slovene. Similarly, we observed a vocabulary increase for the five million entries corpus (5M).

Table 6 shows the results of the automatic translation evaluation between the different corpora described in Table 5 and the network architecture. In summary, we observed that with the usage of the parallel corpus of two million sentences, the network architecture does not have a large impact on the translation quality improvement. In the case when translating from Slovene into English, the translation quality even decreases for the randomly selected parallel corpus. A comparison between randomly selected sentences and identified relevant sentences, we learned that due to the high density of the vocabulary, the network cannot store all the provided information, therefore the neural models trained on the random sample performed better than the relevant sentence corpus (relevant sentences). Due to this, we increase the corpus to five million sentences. In this setting the larger network architecture (5M++)[6] allows translation quality improvement in terms of the BLEU scores. Furthermore, the relevant sentence corpus outperforms the neural model trained on randomly selected sentences, since the neural network is large enough to handle the dense vocabulary within the training dataset.

6. Conclusion

In this paper, we compared the performance of SMT and NMT approaches between English and the morphological rich south Slavic languages, Slovene, Serbian and Croatian. Although SMT performs better on the reduced training dataset, we observed translation quality improvement can be achieved with a parallel corpus containing selected sentences (in comparison to randomly selected sentences) for the NMT approaches, if the networks are enlarged in terms of the LSTM nodes and the number of hidden neural layers.

Our ongoing work focuses further on the selection techniques to reduce the parallel resources while preserving the translation quality. Furthermore, we continue focusing on the subword unit segmentation for terminological expressions and named entities.

Acknowledgement

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289 (Insight).

[6] 1,000 LSTM nodes, 4 hidden layers

7. References

Mihael Arčan, Maja Popović, and Paul Buitelaar. 2016. Asistent – A Machine Translation System for Slovene, Serbian and Croatian. In Tomaž Erjavec and Darja Fišer, editors, *Proceedings of the Conference on Language Technologies and Digital Humanities*, pages 13–20, Ljubljana, Slovenia. Academic Publishing Division of the Faculty of Arts.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473.

Jonathan Clark, Chris Dyer, Alon Lavie, and Noah Smith. 2011. Better Hypothesis Testing for Statistical Machine Translation: Controlling for Optimizer Instability . In *Proceedings of the Association for Computational Lingustics*.

Marta R. Costa-Jussà and José A. R. Fonollosa. 2016. Character-based neural machine translation. *CoRR*, abs/1603.00810.

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*.

Thierry Etchegoyhen, Lindsay Bywood, Mark Fishel, Panayota Georgakopoulou, Jie Jiang, Gerard Van Loenhout, Arantza Del Pozo, Mirjam Sepesy Maučec, Anja Turner, and Martin Volk. 2014. Machine Translation for Subtitling: A Large-Scale Evaluation. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC14)*, Reykjavik, Iceland, May.

Marcello Federico, Alessandro Cattelan, and Marco Trombetti. 2012. Measuring user productivity in machine translation enhanced computer assisted translation. In *Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas (AMTA)*.

Philip Gage. 1994. A new algorithm for data compression. *C Users J.*, 12(2):23–38, February.

Spence Green, Jeffrey Heer, and Christopher D Manning. 2013. The efficacy of human post-editing for language translation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 439–448. ACM.

Kenneth Heafield. 2011. KenLM: faster and smaller language model queries. In *Proceedings of the EMNLP 2011 Sixth Workshop on Statistical Machine Translation*, pages 187–197, Edinburgh, Scotland, United Kingdom, July.

Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent continuous translation models. Seattle, October. Association for Computational Linguistics.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. 2017. Opennmt: Open-source toolkit for neural machine translation. *CoRR*, abs/1701.02810.

Filip Klubička, Gema Ramírez-Sánchez, and Nikola Ljubešić. 2016. Collaborative development of a rule-based machine translator between Croatian and Serbian. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation (EAMT)*, volume 4, Riga, Latvia. Baltic Journal of Modern Computing.

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, and 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.

Simon Krek. 2012. *Slovenski jezik v digitalni dobi – The Slovene Language in the Digital Age*. META-NET White Paper Series. Georg Rehm and Hans Uszkoreit (Series Editors). Springer. Available online at `http://www.meta-net.eu/whitepapers`.

Samuel Läubli, Mark Fishel, Gary Massey, Maureen Ehrensberger-Dow, and Martin Volk. 2013. Assessing post-editing efficiency in a realistic translation environment. In *Machine Translation Summit XIV*.

Wang Ling, Isabel Trancoso, Chris Dyer, and Alan W. Black. 2015. Character-based neural machine translation. *CoRR*, abs/1511.04586.

Pierre Lison and Jörg Tiedemann. 2016. Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).

Nikola Ljubešić, Petra Bago, and Damir Boras. 2010. Statistical machine translation of Croatian weather forecast: How much data do we need? In Vesna Lužar-Stiffler, Iva Jarec, and Zoran Bekić, editors, *Proceedings of the ITI 2010 32nd International Conference on Information Technology Interfaces*, pages 91–96, Zagreb. SRCE University Computing Centre.

M. Manojlović, L. Dajak, and M. B. Bakarić. 2017. Idioms in state-of-the-art croatian-english and english-croatian smt systems. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1546–1550, May.

M. Sepesy Maučec and J. Brest. 2017. Slavic languages in phrase-based statistical machine translation: a survey. *Artificial intelligence review*, ??:1–41.

Mirjam Sepesy Maučec, Janez Brest, and Zdravko Kačič. 2006. Slovenian to English Machine Translation using Corpora of Different Sizes and Morpho-syntactic Information. In *Proceedings of the 5th Language Technologies Conference*, pages 222–225, Ljubljana, Slovenia, October.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29.

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 on Association for Computational Linguistics*, ACL '02, pages 311–318.

Maja Popović and Mihael Arcan. 2015. Identifying main obstacles for statistical machine translation of morphologically rich south slavic languages. In *18th Annual Conference of the European Association for Machine Translation (EAMT)*.

Maja Popović and Nikola Ljubešić. 2014. Exploring cross-language statistical machine translation for closely related South Slavic languages. In *Proceedings of the EMNLP14 Workshop on Language Technology for Closely Related Languages and Language Variants*, pages 76–84, Doha, Qatar, October.

Maja Popović, David Vilar, Hermann Ney, Slobodan Jovičić, and Zoran Šarić. 2005. Augmenting a Small Parallel Text with Morpho-syntactic Language Resources for Serbian–English Statistical Machine Translation. pages 41–48, Ann Arbor, MI, June.

Maja Popovic, Mihael Arcan, and Filip Klubicka. 2016a. Language related issues for machine translation between closely related south slavic languages. In *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects, VarDial@COLING 2016, Osaka, Japan, December 12, 2016*, pages 43–52.

Maja Popović, Kostadin Cholakov, Valia Kordoni, and Nikola Ljubešić. 2016b. Enlarging scarce in-domain english-croatian corpus for smt of moocs using serbian. In *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3)*, pages 97–105. The COLING 2016 Organizing Committee.

Maja Popović. 2015. chrf: character n-gram f-score for automatic mt evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal, September. Association for Computational Linguistics.

Miro Romih and Peter Holozan. 2002. Slovensko-angleški prevajalni sistem (a Slovene-English translation system). In *Proceedings of the 3rd Language Technologies Conference (in Slovenian)*, Ljubljana, Slovenia.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2015. Neural machine translation of rare words with subword units. *CoRR*, abs/1508.07909.

Miloš Stanojević, Amir Kamran, Philipp Koehn, and Ondřej Bojar. 2015. Results of the WMT15 Metrics Shared Task. In *Proceedings of the 10th Workshop on Statistical Machine Translation (WMT-15)*, pages 256–273, Lisbon, Portugal, September.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, NIPS'14, pages 3104–3112, Cambridge, MA, USA. MIT Press.

Víctor M. Sánchez-Cartagena, Nikola Ljubešić, and Filip Klubička. 2016. Dealing with data sparseness in SMT with factored models and morphological expansion: a Case Study on Croatian. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation (EAMT)*, volume 4, Riga, Latvia. Baltic Journal of Modern Computing.

Jörg Tiedemann. 2012. Character-based pivot translations for under-resourced languages and domains. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 141–151, Avignon, France, April.

Antonio Toral, Raphael Rubino, Miquel Esplà-Gomis, Tommi Pirinen, Andy Way, and Gema Ramirez-Sanchez. 2014. Extrinsic Evaluation of Web-Crawlers in Machine Translation: a Case Study on Croatian–English for the Tourism Domain. In *Proceedings of the 17th Conference of the European Association for Machine Translation (EAMT)*, pages 221–224, Dubrovnik, Croatia, June.

Antonio Toral, Raphael Rubino, and Gema Ramírez-Sánchez. 2016. Re-assessing the Impact of SMT Techniques with Human Evaluation: a Case Study on English-Croatian. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation (EAMT)*, volume 4, Riga, Latvia. Baltic Journal of Modern Computing.

Jerneja Žganec Gros and Stanislav Gruden. 2007. The voiceTRAN machine translation system. In *Proceedings of the 8th Annual Conference of the International Speech Communication Association (INTERSPEECH 07)*, pages 1521–1524, Antwerp, Belgium, August. ISCA.