Machine Translation System for the Industry Domain and Croatian Language

Dunđer, Ivan

Source / Izvornik: **Journal of Information and Organizational Sciences, 2020, 44, 33 - 50**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

https://doi.org/10.31341/jios.44.1.2

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:131:280136

DOI: https://doi.org/10.31341/jios.44.1.2
Open Access
UDC 005.336.3:004.912=163.42
Original Scientific Paper

Machine Translation System for the Industry Domain and Croatian Language

Ivan Dunđer *ivandunder@gmail.com*
Faculty of Humanities and Social Sciences
University of Zagreb, Zagreb, Croatia

Abstract

Machine translation is increasingly becoming a hot research topic in information and communication sciences, computer science and computational linguistics, due to the fact that it enables communication and transferring of meaning across different languages. As the Croatian language can be considered low-resourced in terms of available services and technology, development of new domain-specific machine translation systems is important, especially due to raised interest and needs of industry, academia and everyday users. Machine translation is not perfect, but it is crucial to assure acceptable quality, which is purpose-dependent. In this research, different statistical machine translation systems were built – but one system utilized domain adaptation in particular, with the intention of boosting the output of machine translation. Afterwards, extensive evaluation has been performed – in form of applying several automatic quality metrics and human evaluation with focus on various aspects. Evaluation is done in order to assess the quality of specific machine-translated text.
Keywords: statistical machine translation, domain adaptation, automatic quality metrics, human quality evaluation, error classification, Croatian language, information and communication sciences

1. Introduction and motivation

Multilingual communication has become a top priority in today's globalized world. As human translation is a time-consuming, expensive and non-efficient way of satisfying the needs of the translation industry, one of the possible solutions to this problem is to apply the paradigms of automatic machine translation. Automatic machine translation systems today exist for widely spoken languages, while for less spoken languages they are less developed. As Croatia has recently joined the European Union, development of such systems and resources is of significant importance for inclusion into European research projects, academic cooperation and industry.

The main goal of this research was to set up three different machine translation systems that are capable of translating from English into Croatian (EN>HR). Despite the recent advances in neural machine translation (NMT) [1], the author decided to build statistical machine translation systems (SMT) – more specifically, phrase-based

statistical machine translation systems. Such systems segment source sentences into phrases, translate each phrase, reorder every phrase and compose target sentences from these phrase translations [2].

The foundations of statistical machine translation and the corresponding methods and techniques [3] originate in the studies of, mainly, statistics, artificial intelligence and natural language processing, and are therefore, also applicable in higher education and for purposes of academic curricula – especially in selected courses in information and communication sciences, computer science and computational linguistics, that deal, among others, with machine translation in general, translation memories and digital resources, various aspects of natural language understanding and generation, computational language analyses, knowledge and information extraction, intelligent machine behavior, human-computer interaction etc.

In this research, machine translation trials were conducted on the *general domain* and the industry domain, so-called *in-domain*, which in this case considered specific content regarding development and use of various computer software. Also, the following research questions were raised:

- How good are English-Croatian (EN>HR) statistical machine translation systems developed in this research for the general domain and the in-domain in terms of automatic quality evaluation scores and human judgment?
- Are the relatively small datasets used in this research enough to build well-performing English-Croatian (EN>HR) statistical machine translation systems?
- Do multiple phrase translation tables improve machine translation system performance?

2. Related work

Extensive research in the field of statistical machine translation with focus on the Croatian language, and especially with regard to domain adaption techniques has been done by [4].

Another research applied additional morphological knowledge in form of pseudo-lemmatization in a Croatian-English statistical machine translation system [5].

As Croatian can be considered a low-resource language with an evident lack of massive high-quality corpora, one research has focused on providing more resources with help of digitization and subsequent linguistic description [6].

Another research tried a gamification approach in order to increase user engagement in a specially built crowdsourcing platform for collecting parallel corpora for Croatian [7].

Evaluating sentence alignment has also been shown to be important, as demonstrated on an example of Croatian-English parallel corpora [8].

Also, various obstacles in the process of obtaining high-quality sentence-aligned English-Croatian parallel corpora were identified and analyzed [9].

One paper tried to combine automatic speech recognition (ASR) and machine translation in the business correspondence domain for the English-Croatian language pair [10].

Evaluation of machine translation has been done in many different research papers. One paper focused on applying automatic quality metrics on machine translations in the sociological-philosophical-spiritual domain [11].

Another paper demonstrated the use of automatic and human evaluation on English-Croatian legislative tests [12], also with special focus on the BLEU metric [13].

Automatic and human evaluation of online machine translation services has also been done for English/Russian-Croatian [14], [15].

3. Research

The following subsections describe the chosen and analyzed dataset, applied methodology for conducting experiments and the available experimental resources. The first subsection deals with the quantitative examination of the dataset and the preprocessing phase. All of the research steps derive from the field of natural language processing (NLP), and can be employed in higher education and for purposes of academic curricula, e.g. for efficient data analysis training, teaching information extraction from specific corpora and subsequent analyses etc. Natural language processing, if applied correctly, can be used for various objective, precise and cost-effective analyses.

The second subsection discusses the applied research approach and methods for building different solutions, i.e. machine translation systems for the English-Croatian (EN>HR) language pair. Here the author concentrates mainly on exploring the possibilities of utilizing domain adaptation in form of multiple phrase translation tables for the purpose of increasing machine translation quality for the Croatian language.

The third subsection presents the computational and human resources that were used in this research.

3.1. Dataset

The statistical machine translation systems were trained on two different parallel corpora, which were obtained from the internet: a general domain parallel corpus and an industry-specific domain (in-domain) corpus. All datasets used in training (training set), tuning (development set) and testing (test set) processes derived from the collected parallel corpora.

The general domain dataset did not contain any dominant content – on the contrary, it consisted of subtitles from a collection of movies, covering a variety of different types of terminology, such as professional and everyday communication, human interaction, weather, sports, jurisdiction, weaponry, food, slang etc. Furthermore, the general domain dataset did not consist only of Croatian sentences, but also of Bosnian, Slovenian and Serbian sentences, which induced noise in the training process, due to differences in diacritics, vocabulary, morphology etc. Also, a lot of sentences were written in Serbian Cyrillic letters. Very poor translations were also represented, often with certain parts missing in the translation or without any

diacritics. However, nothing was excluded from the original collected corpus, as the intention of the author also was to examine the impact of low-quality corpora on machine translation output.

The domain-specific dataset (in-domain corpus) belonged to the computer software domain, consisting of technical documentation and user guides, various manuals covering specific terminology, such as graphical user interface elements, keyboard shortcuts, information technology acronyms and technological terms.

All datasets were in parallel corpus (paragraph) form, stripped of any text formatting, and in raw plain text format with UTF-8 file encoding, which was used in order to ensure that the whole dataset was saved and represented in a correct manner.

The datasets were computationally analyzed and processed with Python and Perl. Table 1 shows that a relatively small amount of data was used in the machine translation experiments, due to the lack of large and freely available (domain-specific) English-Croatian parallel corpora on the internet.

Language	Sentences	Words	Number of unique words	Max. sentence length (words)	Average sentence length (words)
Training set for general domain machine translation system					
English	289080	2105795	94647	220	7.28
Croatian	289080	1633166	185281	64	5.65
Initial in-domain dataset					
English	20118	210182	15673	86	10.45
Croatian	20118	194321	23609	73	9.66
Training set for in-domain machine translation system					
English	18118	196559	14988	86	10.85
Croatian	18118	181910	22498	73	10.04
Tuning set for general and in-domain systems					
English	1000	4661	1445	27	4.66
Croatian	1000	4265	1818	21	4.26
Test set for general domain, in-domain and combined systems					
English	1000	8962	2028	39	8.96
Croatian	1000	8146	2686	36	8.15

Table 1. Statistics of the used datasets.

Tuning and test sets were extracted and excluded from the initial in-domain corpora: the first 1000 sentences for the development set, and the last 1000 sentences for the test set – this was chosen completely arbitrary. The remaining sentences constituted the training sets. In order to preprocess and prepare datasets, standard natural language processing (NLP) steps were applied: data was tokenized and

truecased, while sentences exceeding 80 words in length were removed from the corpus.

3.2. Experiments

The author chose a phrase-based approach [2] for building statistical machine translation systems for generating English-Croatian (EN>HR) translations. Also, the author focused on domain adaptation [4] as a means of increasing the quality of machine translation system output.

The first statistical machine translation system was trained using the general domain dataset, after which it was tuned with in-domain data (tuning set) in order to fine-tune the weights of the different system models [3] towards the in-domain content.

The second statistical machine translation system was trained with domain-specific data (in-domain corpus with computer software content) and tuned afterwards with the very same tuning set, as used in the first statistical machine translation system, i.e. in the general domain experiment.

The third machine translation system was a combined machine translation system, i.e. a combination of the two earlier mentioned machine translation systems. Namely, both phrase translation tables (general domain + in-domain) were used for scoring, which means that every translation option was collected from each phrase translation table and scored by each phrase translation table. All other machine translation system model features [3], e.g. weights of the reordering model were based on the in-domain machine translation system settings. Also, a model back-off approach was used ("decoding-graph-back-off") [16], so the in-domain phrase translation table was preferred, while the general domain phrase translation table was used only if no translations (for unigrams) were found in the in-domain phrase translation table. In addition, an experiment without "decoding-graph-back-off" was carried out, so that a log-linear interpolation of both (general domain + in-domain) phrase translation tables (with alternative decoding paths) was performed [3].

In all of the machine translation systems in this research, the n-gram language models were trained with the IRSTLM toolkit [17] with the order set to 3, and smoothing with improved Kneser-Ney [3] was applied. In all cases, GIZA++ [18] was used for word alignment, while the "grow-diag-final-and" algorithm [19] was used as the symmetrization method for obtaining word alignments from GIZA++ output. Machine translation system training, i.e. phrase extraction and scoring, generating of phrase translation tables and lexicalized reordering tables was done with Moses [16]. The machine translation systems were always tuned using Minimum Error Rate Training (MERT) [20], in order to increase the BLEU score [21] (based on n-gram precision and brevity penalty) for the given development set, and consequently, the test set.

All three systems were tested using the same test set (see Table 1).

3.3. Resources

The author of this paper decided to deliberately conduct the experiments on an ordinary low-resource, but out of date personal computer with dual-core CPU, only 6 GB of RAM, no GPU and Ubuntu 12.04 operating system.

The process of data preprocessing and preparation took about 3 hours, training and tuning of machine translation systems took about 35 hours, experimenting with the combined machine translation system took cca. 5 hours, whereas automatic and human evaluation of machine translation quality took about 15 hours. Human evaluation with focus on adequacy and fluency was done by a native Croatian speaker and included also an analysis of error typology.

4. Results and discussion

This section deals with the experiment results and discusses the various experimental aspects. This part of the paper analyses not only the quantitative aspects of the resulting machine translations, but also evaluates the machine translation output on a qualitative level.

Namely, here the author presents the results of the automatic and human machine translation quality evaluations. In addition, the author presents and discusses the different obstacles and drawbacks of the conducted experiments, as well as its implications. In the third subsection some machine translation examples are shown in order to point out the various downsides and limitations of the experiments.

4.1. Automatic machine translation evaluation

As human evaluation is time-consuming, expensive and subjective, automatic quality evaluation metrics try to approximate human evaluation as much as possible. Automatic machine translation evaluation is based on machine translation system output, i.e. so-called hypothesis sentences, and reference sentences, i.e. correct translations, which are regarded as the so-called "gold standard".

In this research, automatic evaluation of machine translation output was performed using four different metrics: BLEU [21], $METEOR_{ex}$ [22], GTM-1 [23] and TER [24]. Automatic evaluation metrics differ in many ways, but, basically, the more similar a hypothesis sentence is to the correct translation, the better the translation is scored. The results are presented in Table 2.

Machine translation system	**direction**	**Automatic evaluation metrics**			
		BLEU *	$METEOR_{ex}$ *	GTM-1 *	TER **
general domain	EN>HR	0.053	0.097	0.286	0.821
in-domain	EN>HR	0.319	0.290	0.622	0.483
combined system	EN>HR	model back-off approach ("decoding-graph-back-off")			

(general domain + in-domain)		0.102	0.132	0.357	0.774
		log-linear interpolation of two phrase translation tables			
		0.311	0.283	0.612	0.492
Remarks: * higher is better, ** lower is better					

Table 2. Results of automatic evaluation of machine translation quality.

Table 2 shows that the general domain machine translation system scored very low (BLEU score of 5.3). This might be due to the initial observation that the corresponding training data induced a lot of noise in the system training process. Also, the diverse nature of training (general domain) and tuning sets (in-domain, i.e. computer software content) implies a high rate of out-of-vocabulary words (OOV), rarely seen words and different word alignments, large vocabulary in general, and differences in sentence length.

Moreover, the Croatian and English language have significant structural differences, which do also influence the quality of machine translation from English into Croatian. Croatian is a morphologically rich language with flexible word order, whereas English follows a certain "SVO" (subject-verb-object) linguistic pattern with fairly fixed word order and simple morphology. The "SVO" pattern is the most common pattern in Croatian, but other aesthetic or archaic patterns are also common, especially in certain literary styles, such as poetry. For example, the "SVO" construction *Marko čita novu knjigu* (Eng. Marko reads a new book) is relatively freely transformed into the less common "OSV" (object-subject-verb) construction *Novu knjigu Marko čita* (Eng. A new book is read by Marko). In other words, whole phrasal categories (NP, VP, PP etc.) are easily reordered without changing the meaning of a sentence. Such inversions and other types of permutations are characterized as stylistic, but possible due to morphological richness of Croatian. In fact, Croatian words are highly inflectional and have declensional endings, indicating number, case, gender, direct and indirect objects. Still, a syntactic category, i.e. part of speech (noun, verb, preposition etc.) that appears at the beginning of a sentence is more emphasized. Put differently, translating from a less linguistically complex language (here, EN) to a more complex language (here, HR) should also be accounted for low quality of machine translations, as issues with addressing the necessary reordering produce low scores in the different statistical machine translation systems.

The in-domain system gave the best results in terms of all automatic quality metrics, i.e. BLEU, $METEOR_{ex}$, GTM-1 and TER score. High quality and homogeneity of training and tuning datasets contributed to the evaluation results of machine translation system output. This confirms that with even such a small training dataset good translations can be generated, which is also reflected in the error typology analysis (see Table 4).

The combined system (general domain + in-domain) showed an improvement in machine translation quality when compared to the general domain system, regardless the fact that the combined system's components were not separately tuned, i.e. tuning weights of the general and the in-domain machine translation systems were used. The

model back-off approach ("decoding-graph-back-off") did not perform as expected. In contrary, due to the fact that the training sets of the general domain and in-domain systems were so dissimilar, giving preference to in-domain translation options was not considered suitable for boosting translation quality, i.e. doubling BLEU score (see Table 2). Also, log-linear interpolation of two phrase translation tables (general domain + in-domain) in the combined system experiment did not outperform the in-domain system but did much better than the model back-off approach – i.e. BLUE score was tripled. Therefore, for later-stage human evaluation only the combined machine translation system with log-linear interpolation of two phrase translation tables was considered.

Obviously, the worst results with regard to BLUE, $METEOR_{ex}$, GTM-1 and TER are scored for the general domain machine translation system, whereas the combined system (with both trials: model back-off approach and log-linear interpolation of two phrase translation tables) scored better, but much lower than the in-domain system (except log-linear interpolation). Still, such an approach showed that combining two (and possibly more) phrase translation tables can increase machine translation system performance for the English-Croatian language pair.

4.2. Human machine translation evaluation

Human quality evaluation was conducted in terms of fluency and adequacy [4] on the first 200 machine-translated sentences from the test dataset, by one native speaker of the Croatian language.

Fluency captures to what extent the translation is well-formed grammatically, contains correct spelling, adheres to common use of terms, titles and names, is intuitively acceptable and can be sensibly interpreted by a native speaker [25]. Adequacy captures to what extent the meaning in the source text is also expressed in the translation [25].

This means that adequacy measures how much meaning is transferred from the source sentence to its translation, while fluency indicates how natural the machine translation sounds to a native speaker of the target language.

Both fluency and adequacy were scored on a [1, 4] scale (more is better), and a general average score was also calculated for each machine translation system. The results of fluency and adequacy evaluation are shown in Table 3.

Human evaluation: fluency					
Machine translation system	**Average fluency**	**Flawless (4)**	**Good (3)**	**Disfluent (2)**	**Incomprehensible (1)**
general domain	2	21	7	87	85
in-domain	4	78	62	50	10
combined system (general	2	49	26	72	53

domain + in-domain)					
Human evaluation: adequacy					
Machine translation system	**Average Adequacy**	**Everything (4)**	**Most (3)**	**Little (2)**	**None (1)**
general domain	2	25	43	71	61
in-domain	4	120	53	21	6
combined system (general domain + in-domain)	4	68	38	59	35

Table 3. Results of human evaluation of fluency and adequacy.

In terms of human evaluation of fluency, the in-domain system scored best. Table 3 indicates that in this system incomprehensible sentences appeared only 10 times (5%) in the evaluated test set (out of 200 sentences), whereas flawless sentences appeared 78 times (almost 40%).

The general domain system scored worst: 85 sentences were rated as "incomprehensible" (42.5%), whereas as "flawless" only 21 times (10.5%). The combined system scored relatively well when compared to the general domain system: "flawless" and "good" sentences amounted to 37.5%, whereas "disfluent" and "incomprehensible" to 62.5% of all evaluated hypothesis sentences.

In terms of human evaluation of accuracy, the in-domain system scored best again: in 86.5% of all cases, "everything" and "most" of the meaning in the source sentence was also expressed in the corresponding translation, indicating high quality of machine translation for the computer software domain. Again, as expected the general domain machine translation system scored worst with regard to accuracy.

Furthermore, an error typology analysis was conducted in order to count the different types of machine translation errors, according to the machine translation error classification taken from an analytic metric [25] (see Table 4).

Accuracy	• Incorrect interpretation of source text – mistranslation • Incorrect/misunderstanding of technical concept • Ambiguous translation • Omission (essential element in the source text missing in the translation) • Addition (unnecessary elements in the translation not originally present in the source text)

	• 100% match not well translated or not appropriate for context • Untranslated text
Language	• Grammar – syntax: non-compliance with target language rules • Punctuation: non-compliance with target language rules • Spelling: errors, accents, capital letters
Terminology	• Non-compliance with company terminology • Non-compliance with 3rd party or product/application terminology • Inconsistent
Style	• Non-compliance with company style guides • Inconsistent with other reference material • Inconsistent within text • Literal translation • Awkward syntax • Unidiomatic use of target language • Tone • Ambiguous translation
Country standards	• Dates • Units of measurement • Currency • Delimiters • Addresses • Phone numbers • Zip codes • Shortcut keys • Cultural references • Tone …

Table 4. Machine translation error classification.

The results of the error typology analysis, as presented in Table 5, show that the general domain system was rated worst: almost 600 errors were detected, out of which 364 (61.6%) were accuracy errors. In the evaluation of the combined system, accuracy errors (298) were mostly represented, followed by language errors (48). In total, more than 240 errors were found in the evaluation set of the in-domain system: 106 language errors (44%), 84 accuracy errors (35%), 29 country standards errors (12%), 14 style errors (5%) and 10 terminology errors (4%).

Human evaluation: error typology						
Machine translation system	**Accuracy**	**Language**	**Terminology**	**Style**	**Country standards**	**Total**
general domain	364	138	43	21	25	591
in-domain	84	106	10	14	29	243
combined system (general domain + in-domain)	298	48	15	11	25	397

Table 5. Results of human evaluation of error typology.

The most prominent errors in total were accuracy errors (54%), followed by language errors (24%), which pointed out the difficulties with processing morphologically rich languages like Croatian. These two error categories had the most effect on the perception of translation quality.

4.3. Machine translation examples

Some examples of machine translations generated by the three different machine translation systems are shown below (Table 6).

No.	Machine translation system		Sentence
1	Source (English)		move to next search result and highlight it in the document
	Reference (Croatian)		premještanje na sljedeći rezultat pretraživanja i njegovo isticanje u dokumentu
	Hypothesis (Croatian)	general domain	se sljedeći pretražiti rezultat i highlight to u dokument
		in-domain	prešli na sljedeći rezultat pretraživanja i ga istaknuli unutar dokumenta
		combined system (general domain + in-domain)	sljedeći pretražiti da move rezultat i highlight ga u document
2	Source (English)		move bookmarks out of a nested position

	Reference (Croatian)		premještanje knjižnih oznaka iz ugniježđenog položaja
	Hypothesis (Croatian)	general domain	se bookmarks iz nested položaj
		in-domain	premještanje knjižne oznake izvan ugniježđenog položaj
		combined system (general domain + in-domain)	move van bookmarks od nested položaj
3	Source (English)		you can use http , ftp , and mailto protocols to define your link .
	Reference (Croatian)		da biste definirali vezu , možete koristiti protokole http , ftp i mailto .
	Hypothesis (Croatian)	general domain	moћељ koristiti http , ftp , i mailto protocols da definirali tvoj kariku .
		in-domain	možete koristiti http , FTP i mailto protocols da biste definirali vaš vezu .
		combined system (general domain + in-domain)	možete upotrijebiti http , FTP i mailto protocols da definirali vaš link .

Table 6. Translation examples for each machine translation system.

The source sentence represents the English input that was used for generating machine translations, whereas the reference sentence represents the correct (desired) translation, i.e. the "gold standard". A hypothesis sentence corresponds to the generated machine translation system output in the Croatian language.

Here, three hypothesis sentences per translation example are shown – one hypothesis sentence for each machine translation system. Here, the in-domain machine translation system produced the best output for all translation examples in terms of quality when compared to the corresponding reference translation.

5. Future research and additional directions

In order to increase the quality of machine translation, and due to the fact that data sparsity severely affects the system output, more training data of high quality is required for machine translation research that involves morphologically rich and syntactically complex languages, such as Croatian, and should therefore be crawled from the internet or other sources.

More extensive human evaluation on a larger test set should be done, possibly in combination with a more detailed evaluation framework, such as Multidimensional

Quality Metrics (MQM) [26]. Also, other automatic quality metrics should be examined, especially rank-based metrics.

Furthermore, experiments shall be repeated, but for the Croatian-English (HR>EN) language pair, using same or different parameters. Different types of domain adaptation should also be tested. Training of higher order n-gram language models should be performed as well.

Experiments with machine translation systems trained on concatenated datasets (general domain and in-domain content) might also give valuable insights. Joint parallel corpora consisting of texts written in different Slavic languages, but in Latin script might also be used for system training and comparison of combined systems with emphasis on post-processing techniques and effort.

Experiments with neural machine translation are also planned. Benchmarking of developed machine translation systems against general domain systems, like Google Translate or Yandex.Translate should be conducted, as this always represents valuable feedback.

Possibilities of using placeholders during training and decoding processes should also be investigated. This might be particularly useful for domain-specific corpora (like computer software domain) where shortcuts, numbers, tags, acronyms and abbreviations are encountered very frequently, and which induce noise in the machine translation model. In order to perform statistical significance testing, bootstrapping for NIST or BLEU confidence intervals should be utilized.

When it comes to additional suggestions and directions for future research, some of the possible research tasks related to machine translation are: enhancing the machine translation system model with supplementary features, such as word embeddings in form of vector representation of words [27]; integrating machine translation into a Croatian speech synthesis system [28] with additional word-level evaluation [29] or domain-specific evaluation [30]; analyzing the affective states of machine translation output in comparison to the emotions expressed in the corresponding reference translations by applying sentiment analysis [31]; generating concordances from machine translation output using a novel concordance search algorithm [32], and analyzing the resulting concordances computationally [33]; extracting key terminology out of machine translation output and creating new resources using language-independent methods, which could then be used for e.g. rule-based machine translation (RBMT) [34], [35]; or computationally analyzing domain-specific word occurrences and distributions [36] in machine translation output with regard to more than one reference translation for one hypothesis sentence.

6. Conclusion

In this research, the author experimented with phrase-based statistical machine translation for the English-Croatian (EN>HR) language pair. The processes of data preparation, system training and tuning, testing and evaluating took about 60 hours. Quality of parallel corpora, heterogeneity of data used for training and morphological richness of the Croatian language had large impact on the quality of machine translation and its perception.

This is especially reflected in the general domain system experiment. Namely, this system scored worst in all aspects of evaluation. The in-domain system scored best in terms of automatic and human evaluation. The machine translation trials clearly showed that the in-domain statistical machine translation system is capable of producing good-quality English-Croatian (EN>HR) translations for the computer software domain, despite the lack of large parallel corpora: according to human review on a translation sample, 40% of all sentences were rated as flawless with no post-editing required, while in 86.5% of all cases there is no loss of information in translations. The combined system that used log-linear interpolation of two phrase translation tables showed a quality improvement when compared to the general domain machine translation system. Most frequent types of errors were inaccurate translations and language errors, which is mainly due to language complexity of the target language.

The author is convinced that the applied research methodology can be adopted in various scenarios and for different purposes, and that machine translation of Croatian has a huge potential, especially since numerous problems and experimental drawbacks have been identified, out of which all can be taken into consideration in future research. Despite the relatively low results of the chosen domain adaptation approach in this research, the results still look promising. Further investigation should yield definitive conclusions on what domain adaptation method is appropriate for translating industry-specific texts, from English into Croatian, and vice versa.

References

[1] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches," In Proc. SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, 2014, pp. 103-111.

[2] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation," In Proc. 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (NAACL '03), vol. 1, 2003, pp. 48-54.

[3] P. Koehn, *Statistical Machine Translation*. New York, NY: Cambridge University Press, 2010.

[4] I. Dunđer, *Sustav za statističko strojno prevođenje i računalna adaptacija domene* (*Statistical Machine Translation System and Computational Domain Adaptation*). Doctoral dissertation, University of Zagreb, Zagreb, 2015.

[5] M. Brkić, M. Matetić, and S. Seljan, "Pseudo-lemmatization in Croatian-English SMT," In Proc. 2014 Central European Conference on Information and Intelligent Systems CECIIS, 2014, pp. 242-249.

[6] S. Seljan, I. Dunđer, and A. Gašpar, "From Digitisation Process to Terminological Digital Resources," In Proc. 36th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2013), 2013, pp. 1329-1334.

[7] R. Jaworski, S. Seljan, and I. Dunđer, "Towards educating and motivating the crowd – a crowdsourcing platform for harvesting the fruits of NLP students' labour," In Proc. 8th Language & Technology Conference – Human Language Technologies as a Challenge for Computer Science and Linguistics, 2017, pp. 332-336.

[8] S. Seljan, Ž. Agić, and M. Tadić, "Evaluating sentence alignment on Croatian-English parallel corpora," In Proc. 6th International Conference on Formal Approaches to South Slavic and Balkan Languages (FASSBL 2008), 2008, pp. 101-108.

[9] M. Brkić, M. Matetić, and S. Seljan, "Towards Obtaining High Quality Sentence-Aligned English-Croatian Parallel Corpus," In Proc. 4th IEEE International Conference on Computer Science and Information Technology (ICCSIT 2011), 2011, pp. 241-247.

[10] S. Seljan, and I. Dunđer, "Combined Automatic Speech Recognition and Machine Translation in Business Correspondence Domain for English-Croatian," In Proc. International Conference on Embedded Systems and Intelligent Technology (ICESIT 2014) – International Journal of Computer, Information, Systems and Control Engineering, vol. 8, 2014, pp. 1069-1075.

[11] S. Seljan, and I. Dunđer, "Automatic Quality Evaluation of Machine-Translated Output in Sociological-Philosophical-Spiritual Domain," In Proc. 10th Iberian Conference on Information Systems and Technologies (CISTI'2015), vol. 2, 2015, pp. 128-131.

[12] M. Brkić, S. Seljan, and T. Vičić, "Automatic and Human Evaluation on English-Croatian Legislative Test Set," In Proc. 4th International Conference (CICLing 2013) "Computational Linguistics and Intelligent Text Processing", Part I – Intelligent Text Processing and Computational Linguistics, Lecture Notes in Computer Science - LNCS, Springer, 2013, pp. 311-317.

[13] S. Seljan, M. Brkić, and T. Vičić, "BLEU Evaluation of Machine-Translated English-Croatian Legislation," In Proc. Eighth International Conference on Language Resources and Evaluation (LREC'12), 2012, pp. 2143-2148.

[14] S. Seljan, M. Tucaković, and I. Dunđer, "Human Evaluation of Online Machine Translation Services for English/Russian-Croatian," In Proc. WorldCIST'15 – 3rd World Conference on Information Systems and Technologies (Advances in Intelligent Systems and Computing – New

Contributions in Information Systems and Technologies), 2015, pp. 1089-1098.

[15] S. Seljan, and I. Dunđer, "Machine Translation and Automatic Evaluation of English/Russian-Croatian," In Proc. International Conference "Corpus Linguistics – 2015" (CORPORA 2015), 2015, pp. 72-79.

[16] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, "Moses: Open Source Toolkit for Statistical Machine Translation," In Proc. 45th Annual Meeting of the Association for Computational Linguistics – Companion Volume Proceedings of the Demo and Poster Sessions, 2007, pp. 177-180.

[17] N. Bertoldi, "A tutorial on the IRSTLM library," In. Proc. Second Machine Translation Marathon (MTM 2008): Open Source Convention, Tool demonstration documentation. May 2008, p. 34.

[18] F. J. Och, and H. Ney, "A Systematic Comparison of Various Statistical Alignment Models," *Computational Linguistics*, vol. 29, no. 1, March, pp. 19-51, 2003.

[19] B. Haddow. Lecture, Topic: "Word Alignment." Institute of Formal and Applied Linguistics Faculty of Mathematics and Physics, Charles University, Third Machine Translation Marathon (MT Marathon 2009), 2009, p. 3.

[20] F. J. Och, "Minimum error rate training in statistical machine translation," In Proc. 41st Annual Meeting of the Association for Computational Linguistics, ACL 2003, vol. 1, 2003, pp. 160-167.

[21] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation," In Proc. 40th Annual Meeting of the Association for Computational Linguistics, ACL 2002, 2002, pp. 311-318.

[22] A. Agarwal, and A. Lavie, "METEOR, M-BLEU and M-TER: Evaluation Metrics for High-Correlation with Human Rankings of Machine Translation Output," In Proc. Third Workshop on Statistical Machine Translation (StatMT '08), ACL 2008, 2008., pp. 115-118.

[23] J. Turian, L. Shen, and I. D. Melamed, "Evaluation of machine translation and its evaluation," In Proc. 2003 Machine Translation Summit (MT Summit IX), 2003, pp. 386-393.

[24] M. Snover, B. Dorr, R. Schwartz, J. Makhoul, L. Micciula, and R. Weischedel, "A Study of Translation Error Rate with Targeted Human Annotation," University of Maryland, College Park, Tech. Report. LAMP-TR-126/CS-TR-4755/UMIACS-TR-2005-58. July 2005, p. 17.

[25] A. Görög, "Quality Evaluation Today: the Dynamic Quality Framework," In Proc. Translating and The Computer 36, 2014, pp. 155-164.

[26] J. van der Meer, A. Görög, D. Dzeguze, and D. Koot, "Measuring Translation Quality – From Translation Quality to Business Intelligence," TAUS BV, De Rijp, The Netherlands, Report (White paper). June 2017, p. 18.

[27] I. Dunđer, and M. Pavlovski, "Through the Limits of Newspeak: an Analysis of the Vector Representation of Words in George Orwell's 1984," In Proc. 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2019), 2019, pp. 0691-0696.

[28] I. Dunđer, "CroSS: Croatian Speech Synthesizer - design and implementation," In Proc. 16th International Multiconference INFORMATION SOCIETY - IS 2013 / Collaboration, Software and Services in Information Society (CSS'2013), vol. A, 2013, pp. 257-260.

[29] S. Seljan, and I. Dunđer, "Automatic Word-Level Evaluation and Error Analysis of Formant Speech Synthesis for Croatian," In Proc. 4th European Conference of Computer Science (ECCS '13) – Recent Advances in Information Science (Recent Advances in Computer Engineering Series 17) / Image, Speech and Signal Processing, 2013, pp. 172-178.

[30] I. Dunđer, S. Seljan, and M. Arambašić, "Domain-Specific Evaluation of Croatian Speech Synthesis in CALL," In Proc. 7th European Computing Conference (ECC '13) – Recent Advances in Information Science (Recent Advances in Computer Engineering Series 13) / Language and Text Processing, 2013, pp. 142-147.

[31] I. Dunđer, and M. Pavlovski, "Behind the Dystopian Sentiment: a Sentiment Analysis of George Orwell's 1984," In Proc. 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2019), 2019, pp. 0685-0690.

[32] R. Jaworski, I. Dunđer, and S. Seljan, "Usability Analysis of the Concordia Tool Applying Novel Concordance Searching," In Proc. 10th International Conference on Natural Language Processing (HrTAL2016) – Springer Lecture Notes in Computer Science (LNCS), 2016, p. 6, in print.

[33] I. Dunđer, and M. Pavlovski, "Computational Concordance Analysis of Fictional Literary Work," In Proc. 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2018), 2018, pp. 0644-0648.

[34] S. Seljan, I. Dunđer, and H. Stančić, "Extracting Terminology by Language Independent Methods," In Proc. 2nd International Conference on Translation and Interpreting Studies "Translation Studies and Translation

Practice" (TRANSLATA II) – Peter Lang series "Forum Translationswissenschaft", vol. 19, 2014, pp. 141-147.

[35] I. Dunđer, S. Seljan, and H. Stančić, "Koncept automatske klasifikacije registraturnoga i arhivskoga gradiva (The concept of the automatic classification of the registry and archival records)," In Proc. 48. savjetovanje hrvatskih arhivista (HAD) / Zaštita arhivskoga gradiva u nastajanju, 2015, pp. 195-211.

[36] M. Pavlovski, and I. Dunđer, "Is Big Brother Watching You? A Computational Analysis of Frequencies of Dystopian Terminology in George Orwell's 1984," In Proc. 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2018), 2018, pp. 0638-0643.