Combined Automatic Speech Recognition and Machine Translation in Business Correspondence Domain for English-Croatian

Sanja Seljan, Ivan Dunđer

Abstract—The paper presents combined automatic speech recognition (ASR) of English and machine translation (MT) for English and Croatian and Croatian-English language pairs in the domain of business correspondence. The first part presents results of training the ASR commercial system on English data sets, enriched by error analysis. The second part presents results of machine translation performed by free online tool for English and Croatian and Croatian-English language pairs. Human evaluation in terms of usability is conducted and internal consistency calculated by Cronbach's alpha coefficient, enriched by error analysis. Automatic evaluation is performed by WER (Word Error Rate) and PER (Position-independent word Error Rate) metrics, followed by investigation of Pearson's correlation with human evaluation.

Keywords—Automatic machine translation, integrated language technologies, quality evaluation, speech recognition.

I. INTRODUCTION

COMBINATION of automatic speech recognition (ASR) and machine translation (MT) aims to increase efficiency and quality of communication. They can be used as stand-alone solutions or combined into an information workflow process. ASR and MT technologies are often exploited for business or travelling, but also in education, in Computer-assisted Language Learning (CALL), in e-learning systems, in assistive technologies, in web-based learning tools with conversational interface [1] etc. They can be applied for information search on mobile platforms [2], via smartphones, tablets etc. and integrated with GPS or optic technologies. Such systems are particularly important for communication in less-resourced languages, due to augmented interest for cooperation, tourist visits, business or education. They can be also used in the situation when a user does not speak the foreign language (e.g. patient in the hospital, foreign traveller asking for directions etc.), for faster communication (e.g. in conditions of big multilingual associations), for impaired persons, as presented in [3], where lectures are synchronised and text captions produced by use of automatic speech recognition or in the learning process.

Such integrated technologies have moved from word-to-word translation towards conversational natural language communication and unconstrained domain.

The interest in integrated technologies shows that even in

Sanja Seljan and Ivan Dunđer are with the University of Zagreb, Faculty of Humanities and Social Sciences, Department of Information and Communication Sciences (phone: +385-1-409-2350; fax: +385-1-409-2431; e-mail: sanja.seljan@ffzg.hr, ivandunder@gmail.com).

the recession period, translation industry has retained its popularity, focusing on integrated linguistic tools, its impact on economic productivity and better communication with customers, with employees on medical translations [4], for better translation workflow or in localisation industry.

According to the research presented in [5], spoken language translation, unifying speech recognition, machine translation and speech synthesis, belongs to one of 10 emerging technologies that will change the world communication.

Applications using integrated technologies are rarely available for not widely spoken languages, due to restricted training resources. Their integration would be of interest for global market, for traveling, information retrieval, for language preservation and better communication.

Integrated speech and machine-translated tools are also point of interest for less spoken languages, especially when integrated into larger social, political and cultural communities.

As formal written communication represents widely popular everyday need in business communication, the research is made in the domain of business correspondence integrating ASR and MT technologies.

After the related work, the research on two experiments will be presented, using the commercial tool Dragon Naturally Speaking for automatic speech recognition (ASR) of English and free online tool for machine translation (MT) for English-Croatian and then Croatian-English language pairs. ASR is made for English texts and evaluated by error analysis. MT is performed on the set of sentences and evaluated by human and by automatic evaluation metrics. Human evaluation is made according to criteria of usability and for internal consistency Cronbach's alpha is calculated. Automatic evaluation is made using WER (Word Error Rate) and PER (Position-independent word Error Rate) metrics, followed by Pearson's correlation with human evaluation. In the end, the results are discussed and followed by conclusion.

II. RELATED WORK

Automatic speech recognition is affected by variables such as background noise, vocabulary size, speaker dependency, fluency and clearness of speech, type of utterances, performed training or possibility of accuracy improvement (by adding/spelling names or digits). Automatic speech recognition tools distinguish regarding the speaker's adaptiveness, need or not for training, vocabulary size, by domain specific continuous speech or unlimited spontaneous

World Academy of Science, Engineering and Technology
International Journal of Industrial and Systems Engineering
Vol:8, No:11, 2014

dialogue, by command recognition or by predefined way of formulating requests etc.

Various researches have been conducted using different types of speech technologies, as integrated or stand-alone solutions and evaluated by automatic evaluation metric, as in [6] where ASR is presented as integrated component in computer-assisted translation in order to increase the productivity in the translation process.

In the paper presented in [7] two types of speech recognition tools differing by speaker-dependency, vocabulary size and type of utterances/commands have been evaluated using WER (Word Error Rate) and SER (Sentence Error Rate) metrics.

Google also offers various features of speech recognition and machine translation, but also for generating captions on YouTube [8]. The research by [9] presents integration of ASR and MT for generation of high-quality closed captions and subtitles for live broadcasted TV shows, but still with a human responsible for post-editing and quality assurance.

The research in [10] elaborates on combined approach of speech-to-text system and machine translation using hybrid phrase-based statistical machine translation (SMT) system evaluated by WER (Word Error Rate), TER (Translation Error Rate) and CER (Character Error Rate) in processing of Broadcast News (BN) or Broadcast Conversations (BCs).

Searching the web by voice is presented in [11] using speech technology for up-to-minute information retrieval. Reference [12] analysed the application of integrated ASR and SMT technologies, in the scenario where a human dictates the spoken language translation into speech dictation system, which is then passed to SMT module. The main idea is to reduce WER – Word Error Rate of ASR technologies by incorporating knowledge from SMT.

Reference [13] points out the role of adaptive/assistive technologies for impaired persons having various disabilities.

Speech input recognition system, especially a trained one, allows users to communicate by speech, or speech output system which can read screened text. Various adaptive/assistive technologies could be used in educational and non-educational processes by using multi-sensory input, interaction, individualised training or repetition with positive affective attitudes [14]. The limitations of speech control as assistive technology are presented in [15].

Reference [16] presents relevance of spoken language translation technology from social and economic point of view. In the past researches, the analyses of ASR were mostly technology-oriented and performed under controlled conditions within limited application domains. Today, researches are more oriented to interactive speech translation, not limited to specific domains; they are more user-independent and often integrated into other systems.

Machine translation and computer-assisted translation technologies for Croatian include research on various aspects from the user perspective as presented in [17]. Speech technologies for Croatian include development of formant speech synthesis tool and evaluation among four domains of hotel reservation, weather forecast, insurance and automobile

industry as described in [18], [19] and analysis of recent research efforts in [20].

The paper [21] presents the use of speech synthesis technology in Computer-assisted Language Learning (CALL) domain, pointing out various applications, such as talking dictionaries offering pronunciation of mostly headwords or in some cases whole phrases, talking texts, text dictation, pronunciation training and dialogue partner. The paper [22] summarises in the research main points of speech synthesis systems. They may assume three different roles in CALL: reading machine, pronunciation model and conversational partner. Speech synthesis can be integrated into learning environments which provide controlled interactive speaking practice outside the classroom [23].

Almost the only projects integrating machine translation and speech technologies for Croatian language were two projects conducted at Language Technologies Institute, Carnegie Mellon University. The intention was to develop new technologies for new language pairs using a data-driven approach.

In the paper presented by [24] the basic versions of the system DIPLOMAT were developed for Croatian, Korean, Spanish and Haitian Creole, but could be adapted to new languages. The system was built for very restricted domain and in cooperation with the US Army Chaplain School. It was supposed to communicate with local people about non-military issues such as medical supplies, refugees etc. For the DIPLOMAT project, four speech models were developed: acoustic models and language models for each language.

Speech synthesis included modules of text analysis (expanding numbers, abbreviations, symbols), then lexicon to find pronunciation of words and sound rules, prosody models and waveform synthesis strings of phonemes, which are converted into waveforms. The aim of projects was not to build the prototype, but to investigate efforts in building new systems for new language pairs in speech-to-speech translation system.

The TONGUES project presented by [25] targeted only Croatian language and implemented only Croatian speech. The system integrated speech recognition for English and Croatian, speech synthesizer for English and Croatian, translation system in both directions, and interface allowing active communication. The average grade was OK (among bad, OK, good) with most of grammar/case errors and problems with loudspeakers, followed by translation, recognition and synthesis errors.

III. Research

A. Methodology and Data Set

For the research purposes, two types of interrelated experiments were conducted integrating speech recognition software with online machine translation, all in business correspondence domain.

The first part of experiment included the process of automatic speech recognition of English text by use of Dragon Naturally Speaking Home 12. The system was firstly regularly

World Academy of Science, Engineering and Technology
International Journal of Industrial and Systems Engineering
Vol:8, No:11, 2014

trained with inbuilt texts for approximately 3 hours. Then four sets of 50 sentences were firstly dictated without specific training. After the specific training of mistakes, each test was again dictated and error analysis was made.

The second part of experiment included machine translation research. The total of 70 sentences was machine-translated from English to Croatian by Google Translate in February 2014 and the same set of Croatian reference translations was machine-translated into English. Testing is made on sentences of different length, from phrases (*Dear Sir/Madame, Yours faithfully*) up to sentences of 33 words. The average number of words in English sentences was 12.74.

The human evaluation was performed on a scale ranging from 1 to 5 (more is better) according to the criteria of usability.

The lemmatisation of the whole test set was performed manually, due to requirements of Hjerson tool. It included lemmatisation of English and Croatian reference translation and lemmatisation of automatically translated texts, saved in separate files. In English texts, verbs were transformed into infinitive forms; plural was transformed into singular, comparative and superlative into positive form. In Croatian texts all nouns, pronouns, adjectives and numbers were transformed into nominative case, gender masculine and number singular. The changes were made for the purpose of automatic evaluation by WER and PER metrics, which are then correlated with human evaluation.

B. Metrics

ASR was analysed by quantitative error evaluation and error type analysis. Data set was then spell checked, machine translated and evaluated by human and automatic metrics. Although human evaluation is expensive and time consuming [26], subjective and tiresome work, it is useful to correlate it with automatic measures during the system tuning. Automatic evaluation is performed by Word Error Rate (WER) and Position independent word Error Rate (PER), which have shown to be valuable tools in ASR and MT technologies, for comparing different machine translation systems, as well as for evaluating improvements within one system [27].

C. Human Evaluation of Translated Sentences

The human evaluation of machine translated sentences is performed using the criteria of usability, integrating understandability, adequacy, fluency and satisfaction.

According to [28] usability integrates functionality, educational and entertainment value. Most of usability measures are subjective and include success rate, task completion time, turn correction ratio and number of interaction problems.

In earlier periods the criterion of usability was task-oriented, but moves today to non-task oriented area in human-computer interaction systems pointing out user's satisfaction. In this research the human evaluation was performed using the criterion of usability on the scale ranging from 1 to 5.

D. Cronbach's Alpha

In order to measure the level of internal consistency of human evaluation among evaluators, Cronbach's alpha metric is used, according to standard scale where $\alpha \geq 0.9$ indicate excellent consistency, $0.8 \leq \alpha < 0.9$ good consistency, $0.7 \leq \alpha < 0.8$ acceptable, $0.6 \leq \alpha < 0.7$ questionable, $0.5 \leq \alpha < 0.6$ poor and $\alpha < 0.5$ unacceptable consistency. In the presented equation, K is the sum of components, σ_x^2 is the variance of the observed total test score and $\sigma_{Y_i}^2$ is the variance of K components (1).

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K}\sigma_{Y_i}^2}{\sigma_x^2}\right) \tag{1}$$

E. Word Error Rate (WER) and Position-Independent Word Error Rate (PER)

The human evaluation is correlated with WER (Word-error metric) and PER (Position-Independent word Error Rate) automatic metrics, widely used in speech recognition and machine translation systems.

WER and PER calculation in this experiment was performed using Hjerson, a tool for automatic classification of errors in machine translation output [29]. As input, the tool requires reference translation(s) and hypothesis along with their corresponding base forms.

The Word Error Rate (WER) is based on the Levenshtein distance [30], which performs at character level, while WER metrics is based on misrecognised items on the word level.

It is the minimum number of insertions, deletions and substitutions that have to be performed to convert the generated machine translation (hypothesis) into the reference text. Every word in the hypothesised sentences is compared with reference translation and every word which does not match (inserted, deleted or substituted) is counted as an error and divided by total number of words in reference translation. The WER of the hypothesis hyp with respect to the reference ref is calculated as in

$$WER = \frac{1}{N_{ref}^*}\sum_{k=1}^{K}\min_r d_L(ref_{k,r}, hyp_k) \tag{2}$$

where $d_L(ref_{k,r}, hyp_k)$ is the Levenshtein distance between the reference sentence $ref_{k,r}$ and the hypothesis sentence hyp_k (2). In other words, the sum of lexical items which differ from lexical items in a reference sentence (all substituted (S), deleted (D) and inserted (I) words) is divided by the total number of words in the reference sentence (N), as shown in (3).

$$WER = \frac{S+D+I}{N} \tag{3}$$

The main disadvantage of WER is the fact that it does now take permutations of words into consideration, i.e. the word order of the hypothesis translation cannot be different from the word order of the reference, even if the translation is correct.

World Academy of Science, Engineering and Technology
International Journal of Industrial and Systems Engineering
Vol:8, No:11, 2014

In order to overcome this problem, the Position independent word Error Rate (PER) compares the words in the two sentences without taking the word order into account. PER is always lower than or equal to WER [27].

HPER refers to the set of words in a hypothesis sentence which do not appear in the reference sentence, while RPER denotes the set of words in a reference sentence which do not appear in the hypothesis sentence. In other words, main goal of HPER and RPER is to identify all words in the hypothesis which do not have a counterpart in the reference, and vice versa. HPER (4) and RPER (5) measures can be calculated by following equations:

$$HPER(p) = \frac{1}{N_{hyp}} \sum\nolimits_{k=1}^{K} n(p, herr_k) \qquad (4)$$

$$RPER(p) = \frac{1}{N_{ref}^*} \sum\nolimits_{k=1}^{K} n(p, rerr_k) \qquad (5)$$

$herr_k$ refers to the set of words in the hypothesis sentence k which do not appear in the reference sentence k. Analogously, $rerr_k$ represents the set of words in the reference sentence k which do not appear in the hypothesis sentence k. In order to acquire base forms, manual lemmatisation of corpora was conducted. Lemmatisation is an important part of textual pre-processing and aims to reduce the complexity of the vocabulary of the documents by normalising morphological variants [31]. Lemmatisation is the task of grouping together word forms that belong to the same inflectional morphological paradigm and assigning to each paradigm its corresponding canonical form, called lemma [32]. This is a complex task, especially for inflectionally rich languages, like Croatian. E.g., word forms *pjeva, pjevao, pjevamo, pjevajte, pjevaše, pjevaju, pjevahu* constitute a single morphological paradigm which is assigned the lemma *pjevati* (eng. *sing*).

F. Pearson's Correlation

Pearson's correlation is used as a correlation measure between two variables, in this case between the automatic metric and human evaluation, giving scores between [-1, 1]. When the correlation is negative, it suggests that one variable increases in value (WER or PER), while the other variable decreases (human evaluation).

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Results of speech recognition and machine translation experiments are shown in the following sections.

A. Results of Speech Recognition Experiment

Fig. 1 presents results performed by speech recognition software Dragon Naturally Speaking Home 12. After the general training of several hours, four sets of 50 sentences were dictated. Results significantly improved after the specific training of non-recognised units. The specific training

included transformation of dates into numerals, lexical training and transforming of abbreviated phrases into normalised text (*Dear Mme* into *Dear Madame*) and training of abbreviations. After the specific training, average number of mistakes for all 200 sentences has decreased from 0.41 to 0.12 per sentence. Number of mistakes decreased for each new set of 50 sentences, but also after each specific training, by adding new word and phrases into vocabulary.
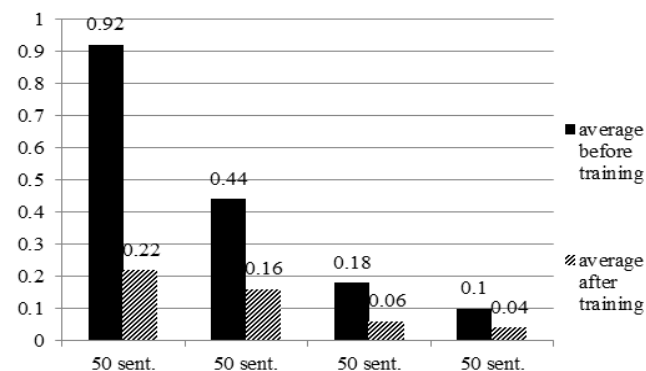

Fig. 1 Average number of mistakes before and after training for each new set of 50 sentences

Fig. 2 shows several types of errors (ordinary words, dates and numbers, phrases) of speech recognised sentences. User specific training significantly improves results for each error type. Distribution of error type is the same regardless its number: lexical errors (*rates, room, hotel, week, fair, various names etc.*) are mostly represented with 66-67%, followed by dates and numbers with 25% , while phrases consisting of 2-4 words are represented with 7-8%.
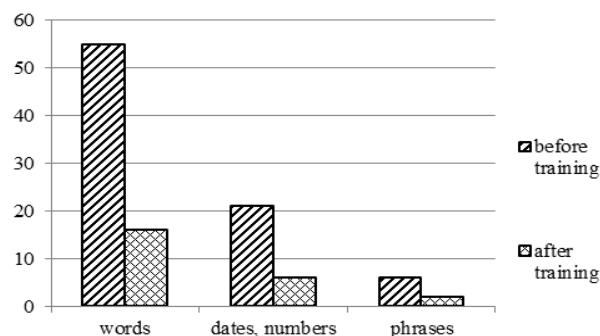

Fig. 2 Average number of mistakes before and after training for each new set of 50 sentences

B. Results of Machine Translation Experiment

In the second experiment, Google Translate statistically-based online translation service was used as the machine translation system, supporting also Croatian language. In this experiment the total of 70 sentences, which were recognised by Dragon Naturally Speaking were translated by Google Online Translation Tool for the English-Croatian and Croatian-English language pairs. The translated sentences were evaluated by Croatian native speakers having academic education and skills in business communication. The evaluation was performed for both language pairs, using the

World Academy of Science, Engineering and Technology
International Journal of Industrial and Systems Engineering
Vol:8, No:11, 2014

criteria of usability and the error analysis. In order to measure the level of internal consistency among evaluators, Cronbach's alpha metric is used.

Fig. 3 shows a histogram of grades given to English-Croatian and Croatian-English translations performed by Google Translate, indicating linear growth of better grades for Croatian-English translation direction, and normal distribution of grades for English-Croatian direction. In the English-Croatian language pair, the middle grade (3) is represented with 32%, while anterior/posterior grades (2 and 4) appeared less frequently (18-23%), while in Croatian-English language pair there are the most of excellent grades, followed by grade very good.

Average grade for English-Croatian is 3.24 and for Croatian-English 4.09 showing that average grade is generally higher for English for 0.7-1.0 score, due to morphological variants in Croatian and relatively free word order.
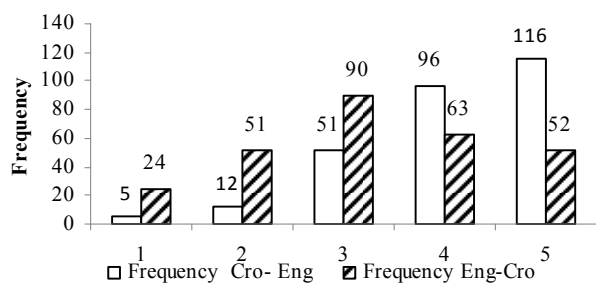

Fig. 3 Histogram of grades for machine-translated sentences

The human evaluation is performed on the scale ranging from 1 to 5, by Croatian native speakers, but very fluent in English business correspondence. Cronbach's alpha coefficient shows excellent consistency (0.92) for translations to Croatian and good consistency for translations to English (0.86).

In Croatian machine-translated sentences there is high number of morphological errors, due to case agreement between pronouns or adjectives with nouns (e.g. *za prikladnim hotelu -> za prikladnim hotelom*), in prepositional phrases (e.g. *u svoj hotel -> u vašem hotelu*) or multi-word units (e.g. *uvjete najma -> uvjeti najma*), followed by lexical errors. There is also lower number of syntactical errors and ommitted words, (e.g. *dvostruko-sobu -> dvokrevetnu sobu*, *potvrdili svoj telefonski poziv -> potvrdili vaš telefonski poziv*).

Fig. 4 presents average grades per evaluator and per language pair showing better grades for translation into English than into Croatian due to morphological differences and word order in the sentence. Scores for English sentences are generally better for cca 0.7-1.0 grade.
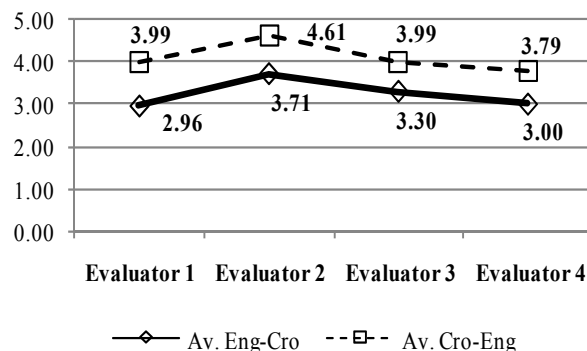

Fig. 4 Average grades per evaluator for Eng-Cro and Cro-Eng

Table I presents average values of automatic metrics WER, HPER and RPER. Results of automatic metrics show better grades for English, i.e. lower error rates, than for Croatian, for 12%-19%.

Conclusively, this indicates that the number of erroneous word orders is higher in the Croatian output. WER results are lower due to high number of morphologically different word forms in Croatian. The second reason is that Google Translate is probably more trained and more suitable for English as the target language.

TABLE I
AVERAGE VALUES OF AUTOMATIC METRICS - WER, HPER AND RPER

	# of sent.	WER	HPER	RPER
Eng.	70	49.00	37.62	40.37
Cro.	70	67.89	55.81	52.11

Table II presents results of Pearson's correlation between automatic metrics and human evaluation. Comparing the two languages, the correlation is much better for English language than for Croatian, with the best correlation for Position-Independent word Error Rate (PER), i.e. for PER metric.

TABLE II
CORRELATION BETWEEN AUTOMATIC METRICS AND HUMAN EVALUATION

Pearson's correlation	WER – Human eval.	HPER – Human eval.	RPER – Human eval.
Eng.	-0.47	-0.33	-0.60
Cro.	-0.13	-0.06	-0.18

V. CONCLUSION

The paper presents results of combined automatic speech recognition (ASR) for English and machine translation (MT) for English-Croatian/Croatian-English language pairs in the domain of business correspondence.

Results of ASR show significant improvements after specific purpose training of ASR system, with significant decrease of number of mistakes in the domain. Average number of mistakes has decreased from 0.41 to 0.12, with the largest proportion of unrecognised whole lexical units.

Results of the machine translation experiment performed by Google Translate tool show that average results of usability for Croatian-English sentences (4.09) is higher than English-Croatian (3.24). Cronbach's alpha shows excellent consistency (0.92) in the evaluation of Croatian target sentences,

World Academy of Science, Engineering and Technology
International Journal of Industrial and Systems Engineering
Vol:8, No:11, 2014

performed by native speakers, and good consistency (0.86) in the evaluation of English target sentences. Translation into English is generally better graded than Croatian for cca 0.7-1.0.

Automatic evaluation conducted by WER and PER metrics, showed better results for English sentences. When comparing the two automatic metrics, better scores were obtained by PER metric, which is more suitable for languages with relatively free word order. The correlation between automatic metrics and human evaluation is better for English than for Croatian, offering best results for PER metric.

The results show possible implementation of combined automatic speech recognition and machine translation technologies for not widely spoken languages (Croatian), although this requires specific training for ASR and improvements in MT use.

The main limitation of this research is relatively small test set in the domain of business correspondence. Further improvements would include an enlargement of test sets, evaluation in other domains and for other language pairs.

REFERENCES

[1] S.-A. Selouani, T.-H. Lê, C. Moghrabi, B. Lanteigne, and J. Roy, "Online Collaborative Learning System Using Speech Technology," *WASET, International Journal of Social Sciences*, 2(2), 2008, pp. 665-670.

[2] S.-Y. Suk, and H.-Y. Chung, "A speech and character combined recognition engine for mobile devices," *International Journal of Pervasive Computing and Communications*, 4(2), 2008, pp. 232-249.

[3] M. Wald, "Creating accessible educational multimedia through editing automatic speech recognition captioning in real time," *Interactive Technology and Smart Education*, 3(2), 2006, pp. 131-141.

[4] M. Miyabe, T. Fukushima, T. Yoshino, and A. Shigeno, "Development of Circulating Support Environment of Multilingual Medical Communication using Parallel Texts for Foreign Patients", International Conference on Health and Medical Informatics (ICHMI 2010), World Academy of Science, Engineering and Technology – WASET (4), 2010, pp. 212-216.

[5] S. Yamamoto, "10 Emerging Technologies That Will Change Your World. *Engineering Management Review*," *IEEE*, 32(2), 2004, pp. 32-51.

[6] E. Vidal, F. Casacuberta, L. Rodriguez, J. Civera, and C. D. M. Hinarejos, "Computer-assisted translation using speech recognition," *Audio, Speech, and Language Processing, IEEE Transactions*, 14(3), 2006, pp. 941-951.

[7] L. Frädrich, and D. Anastasiou, "Siri vs. Windows Speech Recognition," *Translation Journal*, 16(3), 2012.

[8] K. Harrenstien, "Automatic captions in You Tube, *Google Official Blog*," Retrieved on 3rd of April 2014 from: http://googleblog.blogspot.com/2009/11/automatic-captions-in-youtube.html

[9] H. Sawaf, "Automatic Speech Recognition and Hybrid Machine translation for High-Quality Closed-Captioning and Subtitling for Video Broadcast," *Proceedings of Association for Machine Translation in the Americas – AMTA*. San Diego, United States of America, 2012.

[10] M. J. F. Gales, X. Liu, R. Sinha, P. C. Woodland, K. Yu, S. Matsoukas, T. Ng, K. Nguyen, L. Nguyen, J.-L. Gauvain, L. Lamel, and A. Messaoudi, "Speech Recognition System Combination for Machine Translation," *Proceedings of the International Conference On Acoustics,* *Speech and Signal Processing.* Honolulu, *United States of America,* 2007, pp. 1277-1280.

[11] J. Schalkwyk, D. Beeferman, F. Beaufays, B. Byrne, C. Chelba, M. Cohen, M. Garret, and B. Strope, "Google Search by Voice: A case study," *Advances in Speech Recognition.* Mobile Environments, Call Centers and Clinics, 2010, pp. 61-90.

[12] A. Reddy, R.-C. Rose, and A. Désilets, "Integration of ASR and Machine Translation Models in a Document Translation Task," *Proceedings of the International Conference of the International Speech Communication INTERSPEECH,* Antwerp, Belgium, 2007, pp. 2457-2460.

[13] S. Burgstahler, "*Working Together: People with Disabilities and Computer Technology,*" Seattle, University of Washington, 2012.

[14] B. Andresen, "Literacy, assistive technology and e-inclusion," *Journal of Assistive Technologies*," 1(1), 2007, pp. 10-14.

[15] S. Judge, Z. Robertson, and M. Hawley, "The limitations of speech control: perceptions of provision of speech-driven environmental controls," *Journal of Assistive Technologies*, 5(1), 2011, pp. 4-11.

[16] F. Casacuberta, M. Federico, H. Ney, and E. Vidal, "Recent efforts in spoken language translation," *Signal Processing Magazine, IEEE*, 25(3), 2008, pp. 80-88.

[17] S. Seljan *et al.*, "*Computational Language Analysis: Computer-Assisted Translation and e-Language Learning*", (reprint of published papers), Zagreb: Department of Information and Communication studies of Faculty of Humanities and Social Sciences Zagreb, 2012, ch. 1-5.

[18] I. Dunđer, S. Seljan, and M. Arambašić, "Domain-specific Evaluation of Croatian Speech Synthesis in CALL," *Recent Advances in Information Science,* Athens, WSEAS Press, 2013, pp. 142-147. (7th European Computing Conference Dubrovnik, Croatia).

[19] S. Seljan, and I. Dunđer, "Automatic word-level evaluation and error analysis of formant speech synthesis for Croatian," *Recent Advances in Information Science - Recent Advances in Computer Engineering Series 17.* Athens, WSEAS, 2013, pp. 172-178. (4th European Conference of Computer Science, Paris, France).

[20] D. Boras, and N. Lazić, "Aspects of a Theory and the Present State of Speech Synthesis," *Proceedings of the 29th International Convention MIPRO: Computer in Technical Systems,* Rijeka, Croatian Society for Information and Communication Technology, Electronics and Microelectronics – MIPRO, *2006,* pp. 187-190.

[21] Z. Handley, "Is text-to-speech synthesis ready for use in computer-assisted language learning?," *Speech Communication*, 51(10), 2009, pp. 906-919.

[22] Z. Handley, and M.-J. Hamel, "Establishing a Methodology for Benchmarking Speech Synthesis for Computer-Assisted Language Learning (CALL)," *Language Learning & Technology*, 9(3), 2005, pp. 99-120.

[23] F. Ehsani, "Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm," *Language Learning & Technology*, 2(1), 1998, pp. 54-73.

[24] A. Black, R. Brown, R. Frederking, K. Lenzo, J. Moody, A. Rudnicky, R. Singh, and E. Steinbrecher, "Rapid Development of A Speech-to-Speech Translation Systems," *Proceedings of the International Conference on Spoken Language Processing in Denver, United States of America, 2002.*

[25] A. Black, R. Brown, R. Frederking, R. Singh, J. Moody, and E. Steinbrecher, "TONGUES: Rapid Development of a Speech-to-Speech Translation System," P*roceedings of the 2nd International Conference on Human Language Technology Research – HLT,* San Diego, United States of America, *2002*, pp. 183-186.

[26] D. Jurafsky, J. Martin, and A. Kehler, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* New Jersey: Prentice Hall, 2000.

[27] M. Popović, and H. Ney, "Word Error Rates: Decomposition over POS Classes and Applications for Error Analysis," *Proceedings of the Second Workshop on Statistical Machine Translation, Association for Computational Linguistics,* Prague, Czech Republic, 2007, pp. 48-55.

[28] L. Dybkjær, N. O. Bernsen, and W. Minker, "Overview of Evaluation and Usability," *Spoken Multimodal Human-Computer Dialogue in Mobile Environments: Text, Speech and Language Technology*, 28, 2005, pp. 221-246.

[29] M. Popović, "Hjerson: An Open Source Tool for Automatic Error Classification of Machine Translation Output," *The Prague Bulletin of Mathematical Linguistics*, 96(1), 2011, pp. 59-67.

World Academy of Science, Engineering and Technology
International Journal of Industrial and Systems Engineering
Vol:8, No:11, 2014

[30] V. I. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," *Soviet Physics Doklady*, 10(8), 1966, pp. 707-710.
[31] J.-M. Torres-Moreno, "Beyond Stemming and Lemmatization: Ultra-stemming to Improve Automatic Text Summarization," 2012. arXiv:1209.3126 (cs.IR).
[32] A. Gesmundo, and T. Samardžić, "Lemmatisation as a Tagging Task," *Association for Computational Linguistics,* 2, 2012, pp. 368-372.

Sanja Seljan, PhD, Full Professor at the Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb. Her research interests are computer-assisted translation (CAT) and machine translation (MT), natural language processing (NLP), Lexical-Functional Grammar (LFG) and Computer-Assisted Language Learning (CALL). She has published more than 60 research papers, participated in more than 70 conferences and workshops, gave 7 invited lectures and 6 workshops. She has participated in four international and two domestic projects. She was the leader of two national projects: "Information Technology In Computer-Assisted Translation (CAT) of Croatian and in e-Language Learning (eLL)" and "Digital document classification by machine translation techniques (eClass)" funded by the Croatian Ministry of Science, Education and Sport of the Republic of Croatia. She has published a book as a project result "Computational Language Analysis: Computer-Assisted Translation and e-Language Learning". She is a member of different associations related to language technologies. She speaks English and French fluently and uses Italian and German.

Ivan Dunđer, PhD researcher, Research and Teaching Assistant at the Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb. He has graduated Informatics from the same Department with the best grade and has received the Faculty prize for Excellency. His research interests are natural language processing, machine translation and evaluation, language and speech technologies, knowledge management, along with modelling and development of databases and information systems. He has published 17 research papers and speaks fluently German and English.