transLectures

Transcription and Translation of Video Lectures

D6.1.3: Final report on scientific evaluations

UPVLC, XEROX, JSI-K4A, RWTH, EML and DDS

Distribution: Public

transLectures
Transcription and Translation of Video Lectures

ICT Project 287755 Deliverable D6.1.3

October 31, 2014

SEVENTH FRAMEWORK PROGRAMME

Project funded by the European Community under the Seventh Framework Programme for Research and Technological Development.

Information Society Technologies

Project ref no.	ICT-287755
Project acronym	transLectures
Project full title	Transcription and Translation of Video Lectures
Instrument	STREP
Thematic Priority	ICT-2011.4.2 Language Technologies
Start date / duration	01 November 2011 / 36 Months

Distribution	Public
Contractual date of delivery	October 31, 2014
Actual date of delivery	October 31, 2014
Date of last update	October 31, 2014
Deliverable number	D6.1.3
Deliverable title	Final report on scientific evaluations
Type	Report
Status & version	v1.0
Number of pages	11
Contributing WP(s)	WP6
WP / Task responsible	UPVLC
Other contributors	
Internal reviewer	Jorge Civera, Alfons Juan
Author(s)	UPVLC, XEROX, JSI-K4A, RWTH, EML and DDS
EC project officer	Susan Fraser

The partners in transLectures are:

Universitat Politècnica de València (UPVLC)
XEROX Research Center Europe (XEROX)
Josef Stefan Institute (JSI) and its third party Knowledge for All Foundation (K4A)
RWTH Aachen University (RWTH)
European Media Laboratory GmbH (EML)
Deluxe Digital Studios Limited (DDS)

For copies of reports, updates on project activities and other transLectures related information, contact:

The transLectures Project Co-ordinator
Alfons Juan, Universitat Politècnica de València
Camí de Vera s/n, 46018 València, Spain
ajuan@dsic.upv.es
Phone +34 699-307-095 - Fax +34 963-877-359

Copies of reports and other material can also be accessed via the project's homepage: http://www.translectures.eu

1 Introduction

This deliverable describes the final report on scientific evaluations related to Task 6.1. The aim of this task is to measure transLectures progress in terms of current transcriptions and translations quality. Standard evaluation metrics such as WER and BLEU are reported on the test sets defined for conventional experimental evaluation. More details about models and techniques can be found in deliverable D3.1.3.

The definition of the experimental setup for both tasks, transcription and translation, and both repositories poliMedia and VideoLectures.NET is described in deliverable D6.1.1.

2 Summary of scientific evaluations

In general, Figure 1 reflects the good progress that has been achieved for transcription and translation accuracy over the project, and more precisely in the Y3 reporting period. For automatic speech recognition (ASR) on the left side of the figure, relevant improvements with respect to M12 and M24 systems have been achieved by means of massive adaptation techniques for all languages. Relative reductions of WER with respect to M24 are reported of about 14%, 17% and 29% for English (En), Spanish (Es) and Slovenian (Sl), respectively. When more than one system exists for the same language (transcription) or language pair (translation), curve labels are appended with the first letter of the partner's name.

For machine translation (MT) on the right side of the figure, relevant advances have been yielded, particularly in the case of Slovenian pairs for Y3. Massive adaptation techniques produced improvements across most language pairs in Y3, ranging from minor gains in the translation directions involving English and Spanish (EnEs and EsEn), and English into French (EnFr), that were already providing good BLEU figures in Y2, to significant better BLEU scores in the most difficult language pairs, that is, German and Slovene. It should be noticed that following EC reviewers' recommendation a special emphasis has been put into these latter language pairs, achieving higher BLEU scores with respect to M24 of 4%, 20%, 26% relative in English into German (EnDe), Slovene into English (SlEn) and English into Slovene (EnSl), respectively.

3 VideoLectures.NET

3.1 Transcription quality

3.1.1 English (RWTH)

For the English language, RWTH has improved their WER figures with respect to their best results reported in deliverable D6.1.2 at M24 (see Table 1). The first notable improvement came from the development of an incremental decoding. Then, a new backing-off language model explained a minor improvement of 0.4 WER point. Finally, the combination of a new speaker adaptation techniques based on state-confidences and softmax adaptation (SA), and the interpolation of the conventional n-gram language model with a recurrent neural network language model (LSTM-LM) provided the final WER figure of 18.3. This means a relative reduction with respect to the best English ASR system at M24 of 13.7%.

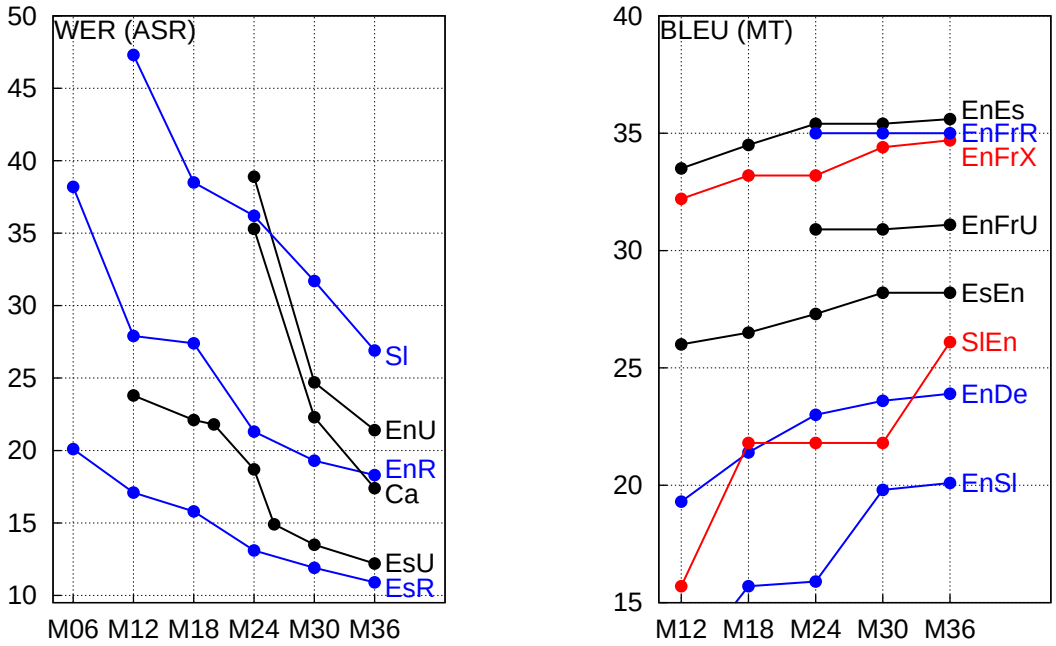

Figure 1: Progress for all languages in ASR (in the left and in terms of WER) and in MT (in the right and in terms of BLEU). The best adapted system is ploted for all partners and languages covered. Each system is codified by colors depending on the partner that owns it: UPVLC (black), RWTH (blue) and XRCE (red). Additionally, for tasks in which several partners have different systems, a suffix character is added to highlight the system owner: R (RWTH), U (UPVLC) and X (XRCE). For clarity's sake, If there is only one system, the suffix is omitted.

Table 1: WER results on the test set for the English RWTH ASR system compared to the best system in M24.

	WER
M24 system	21.3
+ incremental decoding	20.5
+ new backing-off-LM	20.1
+ state-confidences	19.3
+ LSTM-LM	18.4
+ SA	18.3

3.1.2 Slovenian (RWTH)

There have also been a significant improvement in the accuracy of the Slovene ASR system with respect to M24 as can be observed in Table 2. The first improvement came from using a unilingual hierarchical deep neural network (DNN) for bottleneck feature extraction in acoustic modelling. However, moving to a multilingual DNN trained on 800 hours of broadcast news and conversations in English, French, German and Polish turned out to be even more beneficial. Once the lecture slides were incorporated into the language model another decrease in WER was obtained. Then, the replacement of the conventional Viterbi decoding by that based on a confusion network reduced WER by almost another WER point. Finally, the interpolation of the conventional n-gram language model with a LSTM language model provided the final WER figure of 26.9. The relative reduction of WER is 29.4% with respect to M24.

Table 2: WER results on the test set for the Slovenian RWTH ASR system compared to the best system in M24.

	WER
M24 system	38.1
+ bottleneck features	33.0
+ multilingual DNN	30.4
+ lecture slides LM	29.2
+ confusion network	28.3
+ LSTM-LM	26.9

3.2 Translation quality

3.2.1 English into Spanish (UPVLC)

As reported in Table 3, the best result in M24 was obtained using the bilingual sentence selection (BSS) technique. Since then, other two well-known selection techniques were tested: Moore and Axelrod (LM+TM cross-entropy), that provided a slight improvement, but significantly reduced the computing time in comparison to BSS.

Table 3: BLEU results on the test set of the English-Spanish MT system from M24 to M36 on the VideoLectures.NET task.

	BLEU
M24	35.5
LM+TM cross-entropy	35.6

3.2.2 English into French (UPVLC-XEROX-RWTH)

At M24 RWTH deployed the best English-French MT system. Since then, as shown in Table 4, XEROX has been improving their system by applying the meta-parameter optimization, the language model array and the relevance feature approaches achieving a comparable BLEU figure to that of RWTH at M24. UPVLC attained a minor gain in BLEU with respect to M24 by moving from the BSS to the LM cross-entropy selection technique.

Table 4: BLEU results on the test set of the English-French MT system from M24 to M36 on the VideoLectures.NET task.

	BLEU
M24 - RWTH	35.0
M36 - XEROX	34.7

3.2.3 English into German (RWTH)

The English-German MT system is based on the hierarchical-phrase-based translation decoder which is part of the open-source toolkit Jane [3]. As shown in Table 5, the M24 system was first improved by incorporating a hierarchical reordering model, followed by a word-class language model that was complemented by neural networks for language and tranlation modeling. Finally, the application of discriminative maximum expected BLEU training provided the last gain in BLEU up to 23.9 points. A relative BLEU increase of 3.9%.

Table 5: BLEU results on the test set of the English-German MT system from M24 to M36 on the VideoLectures.NET task.

BLEU	
M24	23.0
+ hier. reordering	23.1
+ word class LM	23.3
+ neural network LM+TM	23.6
+ discrim. training	23.9

3.2.4 English into Slovenian (RWTH)

The M24 English-Slovenian MT system was based on the phrase-based translation decoder of the Jane toolkit and important improvements have been incorporated since then as described in Table 6. First, a manual inspection of the reference translations revealed that disfluencies were not translated. After correcting the reference translation by translating disfluencies, the BLEU score of the system increased notably, and even more after reoptimising the weights of the system. As in the English into German system, the application of discriminative maximum expected BLEU training also improve the translation quality of the system. Finally, using former and corrected translation references more robust and slightly better results were attained. To sump up, a relative BLEU increase of 20.9% is reported.

Table 6: BLEU results on the test set of the English-Slovenian MT system from M24 to M36 on the VideoLectures.NET task.

	BLEU
M24	15.9
+ corrected reference	18.2
+ reoptimization	19.3
+ discrim. training	19.9
+ 2nd reference	20.1

3.2.5 Slovenian into English (XEROX)

The Slovenian-English MT system deployed by XEROX at M24 has been notably evolved over Y3 as reported in Table 7. First, as in English-Slovenian, a corrected reference translation has been provided that increase BLEU score by 0.8 points. Then, the same technique employed by XEROX in the English-French MT system were tested here, but incorporating additional training data from reverse translation (English-Slovene) generated by RWTH in WP2. Finally, as in English-Slovenian results, adding a second reference translation produces better and more robust BLEU scores up to 26.1 points. A relative BLEU increase of 16.5%.

Table 7: BLEU results on the test set of the Slovenian-English MT system from M24 to M36 on the VideoLectures.NET task.

	BLEU
M24	21.8
+ corrected reference	22.5
+ meta-parameters optimization	22.6
+ language model array	23.5
+ reverse translation	25.2
+ relevance feature	25.3
+ 2nd reference	26.1

4 poliMedia

4.1 Transcription quality

4.1.1 Spanish (RWTH)

The Spanish RWTH ASR system has improved since the M24 system based on the combination of two systems, S1 and S2, was deployed. As shown in Table 8, both systems have been retrained, in the case of the S2 system using minimum phone error (MPE) criteria. The new system combination produced a better system with 11.9 WER points. Finally, as in VideoLectures.NET for English and Slovenian, the new LSTM language model boosted system accuracy reducing WER to 10.9. So, the relative WER decrease from M24 to M36 is 16.8%.

Table 8: Evolution of WER on the test set for the Spanish RWTH ASR system from M24 to M36.

	WER
M24	13.1
+ new S1 and S2 with MPE training	11.9
+ LSTM-LM	10.9

4.1.2 Spanish (UPVLC)

As in M24, the UPVLC Spanish automatic speech recognition (ASR) system is based on the AK toolkit [1] to train acoustic models and the SRILM toolkit [2] to deploy n-gram language models.

Table 9 reports the evolution of WER figures on the test set starting from the non-LM-adapted M24[1] to the final M36 system. The first significant gain came by tuning the unilinigual DNN. However, a multilingual DNN trained on Spanish and Catalan provided a better system. Then, the aplication of speaker adaptation techniques based on softmax adaptation turned out to slightly improve WER figures. The combination of a convolutional neural network (CNN) with the previous speaker-adapted multilingual DNN reduce WER figures by 0.2 WER points. Finally, a new LM adaptation technique based on a novel vocabulary selection algorithm produced a final boost in the system accuracy decreasing WER to 12.2 points.

Table 9: Evolution of WER (%) on the test set for the Spanish UPVLC ASR system from M24 to M36.

	WER
M24	19.5
+ unilingual DNN	17.4
+ multilingual DNN	16.8
+ softmax adaptation	16.5
+ system combination CNN	16.3
+ new LM adaptation	12.2

4.1.3 Catalan (UPVLC)

As shown in Table 10, the preliminary Catalan UPV ASR system created at M24 was significantly improved by applying the same techniques that provided good results for Spanish. Doing so, the Catalan M36 system reduced WER by half in comparison to the M24 system.

Table 10: Evolution of WER (%) on the test set for the Catalan UPVLC ASR system from M24 to M36.

	WER
M24	35.3
+ unilingual DNN	23.5
+ multilingual DNN	21.6
+ softmax adaptation	21.0
+ system combination CNN	21.0
+ new LM adaptation	17.4

4.2 Translation quality (UPVLC)

The Spanish-English MT system for poliMedia was improved in the same way than the English-Spanish MT for VideoLectures.NET described in Section 3.2.1, that is, replacing BBS by LM+TM cross-entropy techniques as show in Table 11. In this case, BLEU score has been increased by 0.9 points.

5 Conclusions

This deliverable has presented the final report of scientific evaluations for automatic transcriptions and translations on both repositories, VideoLectures.NET and poliMedia. As depicted

[1]The WER for the LM-adapted M24 system is 18.7

Table 11: BLEU figures on the test set of the Spanish-English MT system from M24 to M36 on the poliMedia task.

	BLEU
M24	27.3
LM+TM cross-entropy	28.2

in Figure 1, the evolution of WER and BLEU figures over Y3 and, more generally, over the project in all transcription languages and translation pairs is impressive. These figures prove one of the scientific and technological objectives of transLectures: the relatively small gap for the current technology on ASR and MT to achieve accurate enough results can be closed by massive adaptation.

References

[1] AK toolkit. `http://aktoolkit.sourceforge.net/`.

[2] Andreas Stolcke. SRILM – an extensible language modeling toolkit. In *Proc. of ICSLP*, 2002.

[3] David Vilar, Daniel Stein, Matthias Huck, and Hermann Ney. Jane: Open source hierarchical translation, extended with reordering and lexicon models. In *ACL 2010 Joint Fifth Workshop on Statistical Machine Translation and Metrics MATR*, pages 262–270, Uppsala, Sweden, July 2010.

A Acronyms

ASR	Automatic Speech Recognition
BLEU	Bilingual Evaluation Understudy
BNSI	Broadcast News Speech Corpus
DDS	Deluxe Digital Studios Limited
DNN	Deep Neural Network
EML	European Media Laboratory GmbH
GOS	GOvorjena Slovenscina (spoken Slovene)
JSI	Josef Stefan Institute
K4A	Knowledge for All Foundation
LM	Language Model
MLP	Multi-Layer Perceptron
MT	Machine Translation
OCR	Object Character Recognition
RWTH	RWTH Aachen University
SMT	Statistical Machine Translation
TED	Technology, Entertainment, Design
TER	Translation Error Rate
TM	Translation Model
UPVLC	Universitat Politècnica de València
WER	Word Error Rate
XRCE	XEROX Research Center Europe