

Supporting language diversity of European MOOCs with the EMMA platform

Citation for published version (APA):

Brouns, F., Serrano Martínez-Santos, N., Civera, J., Kalz, M., & Juan, A. (2015). Supporting language diversity of European MOOCs with the EMMA platform. In M. Lebrun, M. Ebner, I. de Waard, & M. Gaebel (Eds.), *Proceedings of the European MOOC Stakeholder Summit 2015* (pp. 157-165)

Document status and date:

Published: 01/05/2015

Document Version:

Publisher's PDF, also known as Version of record

Document license:

CC BY-NC-SA

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 12 Oct. 2022

Open Universiteit
www.ou.nl



Supporting language diversity of European MOOCs with the EMMA platform

Francis Brouns (*Open University of the Netherlands, Netherlands*), **Nicolás Serrano Martínez-Santos** and **Jorge Civera** (*Universitat Politècnica de València, Spain*), **Marco Kalz** (*Open University of the Netherlands, Netherlands*), **Alfons Juan** (*Universitat Politècnica de València, Spain*).

ABSTRACT

This paper introduces the cross-language support of the EMMA MOOC platform. Based on a discussion of language diversity in Europe we introduce the development and evaluation of automated translation of texts and subtitling of videos from Dutch into English. The development of an Automatic Speech Recognition (ASR) system and a Statistical Machine Translation (SMT) system is described. The resources employed and evaluation approach is introduced. Initial evaluation results are presented. Finally, we provide an outlook into future research and development.

Introduction

English is undisputable the lingua franca in the academic world and in business life. This has also been mirrored in the global open education movement. Since the first initiatives to share open educational resources (OER) have been initiated by major higher education institutions from the US, English was the primary language for OER repositories. Later, also language and cultural aspects have been taken into account (Kalz, Specht, Nadolski, Bastiaens, Leirs, & Pawlowski, 2010). The same pattern of development also applies for the fast growth of Massive Open Online Courses (MOOCs). North America as a very large English-language area has traditionally a much smaller diversity of languages compared to Europe. Crawford (2000) even discusses this as a 'war with diversity' and he provides an account how difficult bilingual education in the US is and how ideological the English-only movement is rooted in the culture.

On the contrary, Europe is a geographical area with a very high diversity of languages and cultures. This diversity is also implemented in the political, legislative and juridical system of the European Commission. The European Commission knows 24 official working languages and employs 1,750 linguists and 600 support staff plus 600 full-time and 3,000 freelance interpreters to keep and support this diversity on its highest democratic levels. Consequentially, this language diversity has also been stressed for the European Higher Education area. In the Bologna process, a balance between

national identity and mobility of learners and teachers is sought.

Moreover, in the current knowledge society citizens need to develop key competences to be able to maintain and improve their employability. The EU has identified eight key competences among which seven are in one way or another related to the multilingual and cultural issues discussed in this paper. These key competences are: communication in the mother tongue, communication in foreign languages, digital competence, learning to learn, social and civic competences, sense of initiative and entrepreneurship, cultural awareness and expression.

It is important that development of these key competences starts with young people during education and in the phase from moving from education to a working life. Throughout their lives, adults need to develop and update these skills. However, as said before, it is difficult to arrange this purely by the formal educational process. The Expert group on New Skills for New Jobs recommends actions in education and training to develop the right mix of skills in enabling key competences, such as learning to learn, digital competence, cultural awareness, and communication in foreign language (Campbell et al., 2010; Alidou, Glanz, & Nikièma, 2012). The OER and MOOC movement offer many opportunities for education providers to address these key competences, either directly or indirectly.

For the MOOC movement, many providers are into a paradoxical decision process. Depending on strategic goals aligned with their open education initiatives, they either go for entering an international market with a primarily English speaking audience, or they offer open education for their national audience and in their national language. According to data of the European MOOC Monitor from September 2014, 346 from 770 open online courses (45%) in Europe are delivered in English (see Figure 1).

Languages of European MOOCs

(n=770)

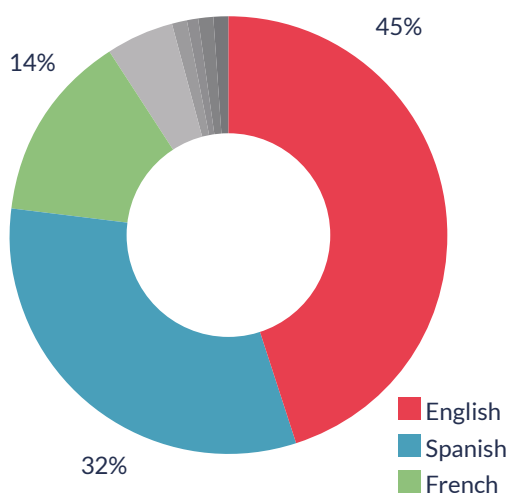


Figure 1. Languages of European MOOCs

Initially MOOC providers then will design the MOOC for the chosen market and that puts limitations both on the provider and participant. On the other hand, MOOCs are by definition open and therefore MOOCs regularly have a really worldwide audience and as a side-effect very diverse language competences of participants can be assumed (Liyaganawardena, Adams, & Williams, 2013). Consequently many participants will fail to make the best out of their MOOC learning experience. One way to deal with some of the language barriers would be to add subtitles to videos in multiple languages and translate the main content of the MOOC. In addition to addressing language barriers subtitling and translation also increases accessibility and reduces the risk of exclusion of participants with special needs. Although subtitles are commonly associated as means to assisted hearing-impaired persons, there are also very useful for participants who have not yet mastered the course language and even for participants who have some form of

mastery of the language when they for example are in a noisy environment, or assist those who prefer to read instead of watch and listen.

For MOOC providers this would have additional benefits as it would allow them to open up to niche markets or promote niche products to wider markets. For participants there can be an implicit benefit in that multilingual MOOCs will stimulate the development of the intercultural and language key competences. However, manual transcription and translation of video and content is time consuming.

In this paper we present the approach taken in the EMMA project that is based on language technology. The EMMA project (European Multiple MOOC Aggregator) aims to integrate and extend separately developed technological components to create and test an innovative learning environment for the delivery of MOOCs. While language technology has a lot of potential application examples in the educational domain (Berlanga et al, 2009) in this work we focus on software that will be used for automated generation of transcription of videos and translation of content. Results will be reported on transcription and translation of videos only.

The approach and method used by EMMA is designed to support several languages, some of which have been tested extensively. Because Dutch is a new language for the system, we will only introduce results for the automated transcription and translation of Dutch-spoken videos that are being used in our MOOCs. Last but not least we discuss the results and provide an outlook into future research and development.

Method: Cross-language exploitation of video and course content

The success of MOOCs is mostly caused by its universal and open access. However, in practice, the universal access to MOOCs is not as such for hearing-impaired people and those not knowledgeable about the language in which the course is delivered. Enriching MOOCs with the transcription and translation of their audiovisual content and the translation of textual content significantly enhances accessibility, opening MOOCs to the worldwide community. Transcription and translation of course might benefit other e-learning courses as well, but in particular in MOOCs the potential is much larger because MOOCs typically are advertised as open courses and attract participants from all over the world. Moreover, many MOOCs are designed as so-called xMOOCs, where videos form a major if not only medium for the course designer to bring across learning content. Due to the lack of interaction with the teacher

multilingual content could greatly assist the non-native participants. Even native participants can benefit from transcription as shown by Ding et al. (2014). These authors report about experiences from a bilingual MOOC in bioinformatics. Lecture videos in the MOOC were recorded in Chinese and subtitled in English. Interestingly, the English subtitles were not only beneficial to non-native Chinese students but also to Chinese native speakers.

Several different approaches are reported in the literature with regard to cross-language translation, subtitling and support. Most of the time content and videos are manually transcribed and translated by the course designer and authors of the content. When quality is important, professional translators are employed. This is a rather expensive approach. It is also known that MOOC participants have translated content and made available to others. These translated contents have then been made available by the MOOC providers in subsequent versions of the MOOC. Coursera actually recruits volunteers to translate transcripts in its Global Translator Community. A similar approach can be seen in crowdsourcing of translation. Crowdsourcing is a rather new phenomenon in which volunteers are sought via the internet to assist with a particular task. This task can be anything, from generating an idea to developing a product. Main characteristic is that it draws on existing expertise and collaborative processes. Anastasiou & Gupta (2011) for example compared machine translation with crowdsourcing translation and Hu, Resnik, & Bederson (2014) obtained good results from a multifaceted monolingual crowdsourcing approach in combination with machine translation. Although human translation is assumed to be better than machine translation, this is not always the case in crowdsourcing translation (Kunchukuttan et al. 2012). Moreover Anastasiou & Gupta (2011) found that the majority of people would need incentives for getting involved in a translation crowdsourcing process.

The manual generation, either by professional translators, course designers or through crowdsourcing, of transcriptions and translations is a rather time-consuming and expensive task. In the language processing domain however, machine translation is being used to automatically translate from one language into another.

This seems a suitable first approach is to generate transcriptions and translations for MOOC content. Unfortunately, even using the current state-of-the-art technologies, transcriptions and translations are far from perfect. Nevertheless, these automatic transcriptions and translations could be reviewed

by course designers, teachers or even volunteers to produce accurate enough materials for students will little effort. In fact, this computer-human approach has shown to reduce the effort needed from a completely manual approach (Serrano, 2014).

Automatic Speech Recognition (ASR) and Statistical Machine Translation (SMT) have made important progress over the last years achieving accurate enough results for many applications (Hinton et al., 2012, Bojar et al. 2014). Indeed, automatic transcription and translation of MOOCs define a new challenging application for ASR and SMT technology. Automatic transcription and translation of video lectures was studied in the transLectures project, in which automatic transcription and translations of videolectures were produced and post-edited via a web interface (Silvestre-Cerdà et al., 2012). User evaluations corroborated the notable increase in productivity to generate transcriptions and translations for video lectures in comparison to do it from scratch. In addition, lecturers and students show their satisfaction with this computer-assisted transcription and translation system in terms of usability.

In the current work, we describe the ASR and SMT systems that have been developed to transcribe and translate the contents of MOOCs offered on the EMMA platform. In this paper, we focus on the case of courses from educational sciences offered by the Open Universiteit in the Netherlands (OUNL). Specifically, Dutch videos extracted from MOOCs are first transcribed and then translated into English. In the following, we first describe the resources collected to create the transcription and translation systems. Next, ASR and SMT technology behind these systems is briefly described. Finally, results are discussed and prospects of future work are proposed.

Collection of resources

State-of-the-art ASR and SMT systems are usually built from a large amount of data from different domains. As a result, these are general-purpose systems that cannot properly deal with content coming from specific domains. For instance, general ASR and SMT systems will have difficulties transcribing or translating specific vocabulary included in MOOCs. Fortunately, the quality of ASR and SMT systems can be significantly improved by adapting them to the specific domain of the MOOC content in question.

As mentioned above, the first step to build ASR and MT systems is to collect audio and textual resources. These resources can be classified as in-domain and out-domain resources. In our case,

in-domain resources are those materials related to the MOOC to be transcribed and translated, while the rest of resources are considered out-domain. An alternative classification of these resources is the type of resource, instantiated as annotated speech, lexical annotation, Dutch text, English text, and

parallel Dutch-English text. The first three types of resources are used to build the ASR system, whereas the last two are employed to create the SMT system. Table 1 depicts the basic statistics for all resources obtained.

Table 1: ASR and SMT resources

	Annotated Speech	Lexical Annotation	Monolingual Dutch	Monolingual English	Parallel Dutch-English
Duration (h)	122	-	-	-	-
Sentences (M)	-	-	52.1	115.9	33.3
Running Words (M)	1.5	0.4	631	2007.3	509.3 (D) - 473.0 (E)
Vocabulary (K)	62.4	-	3956.9	7282.3	2326.5 (D) - 2836.3 (E)

In Table 1, annotated speech refers to the collection of speech documents together with its corresponding time-aligned transcriptions. These speech documents are a subset of those included in the Corpus Gesproken Nederlands (CGN) (Oostdijk, 2000), which contains over 800 hours of annotated speech. At the moment, our current Dutch transcription is trained on a selection of 122 hours, containing over 1.5 million of words from a vocabulary of 62.4 distinct words. This selection has acoustic conditions that are similar to those of the OUNL MOOCs.

Lexical annotation refers to phonetic dictionaries of Dutch, that is, a list of Dutch words with its corresponding transcription(s) at the phoneme level. Phonemes are the elemental unit of human speech. In this work, two phonetic dictionaries were involved: the CGN lexicon and the WEBCELEX lexicon.

The monolingual Dutch electronic text comes from different publicly available sources, such as the European Commission or Wikipedia. In addition, a small set of in-domain text documents were provided by the Open University of the Netherlands. A similar compilation of resources was carried out for the monolingual English electronic text.

Parallel Dutch-English refers to textual resources that contain the same sentences in Dutch and English. Most of these resources have been obtained from the OCROPUS website, but also from some European Union portals, such as Europarl TV. These parallel texts should be considered out-domain. Again, all the resources used are freely available for research and educational purposes.

Automatic Speech Recognition for Dutch

The ASR system developed for Dutch is based on a probabilistic approach to the transcription problem. Basically, given a speech signal, the system will search for the most probable transcription. This probabilistic approach to the transcription problem results in a system integrating three underlying models.

1. The acoustic model, which estimates the probability of the phonemes that are being uttered in the speech signal.
2. A lexical model, which specifies how the phonemes are built up into words.
3. A language model, which estimates the probability of the sequence of words being transcribed.

Acoustic model

The acoustic model corresponds to a hybrid between a Hidden Markov Model (HMM) and a Deep Neural Model (DNN), which correspond to the current state-of-the-art ASR systems (Hinton et al., 2012). Basically, the HMM splits the speech into segments and the DNN classifies these segments into the corresponding phonemes. This model is trained using the data shown in the second column of Table 1. The training process of this model is composed by multiple passes and employs a wide range of techniques from Pattern Recognition (PR). PR is a research area which studies and develops methods to help machines to recognise objects, logical structures (such as a language) and patterns from input signal as humans do. For the sake of clarity, we will only give a summary of this process.

First, speech files are preprocessed to reduce the noise and variability of the signal. Then, given the numerical vectors extracted from the speech signal and their transcription, the standard HMM model is trained from all samples, resulting in a universal model representing all speakers (Rabiner, 1989).

Once this standard model is trained, a speaker-adapted model is estimated using Constrained Maximum Likelihood Linear Regression (CMLLR) (Gales, 1998). Basically, this model normalises each speaker to be more homogeneous to the others. Last, a DNN is trained for each HMM trained, that is, the standard and the CMLLR. In this work, the estimation of the acoustic models has been carried out using the transcriptions UPV toolkit (TLK) (del-Agua et al., 2014), which is freely available.

Lexical Model

A lexical model provides the information of how each word in a language should be pronounced. Obviously, the lexical model to be employed in an ASR system depends on the phonetics of the language under study. In case of Dutch, the pronunciation of a word is ambiguous, i.e. the same word might be pronounced in different ways and simple pronunciation rules are not available. In order to generate the phonetic transcription of each word, a statistical grapheme-to-phoneme model (Bisani, 2008) has to be trained. This model infers the phonetic transcription of new words from a limited set of phonetically annotated words.

Language model

In this work the language model (LM) employed for Dutch ASR corresponds to a linear mixture of interpolated n-gram models trained on the textual resources available (Bellegarda, 2004). More precisely, our LM is based on n-gram models that are estimated for each available individual resource. For instance, a LM will be trained on text extracted from European Commission website, another one from Wikipedia and so on. An n-gram model is a probabilistic model which estimates the probability of a sentence as the probability of consecutive groups of up to n words (Chen and Goodman, 1996). Once, individual LMs are trained for each resource, LMs are combined using a linear mixture optimised on in-domain textual content. This was in our case text extracted from the content of the MOOCs of the EMMA platform. This language model has been trained with the SRI Language Modeling Toolkit (SRILM) toolkit (Stolcke, 2002), which is freely available for research and educational purposes.

Statistical Machine Translation for Dutch

As in ASR, state-of-the-art machine translation systems also follow a probabilistic approach to the problem. In this case, given a sentence in an input language, the system calculates which is its most probable translation. Statistical Machine Translation

(SMT) systems are composed by two models: the translation and language models. The language model corresponds to the same as described for ASR, but for the English language, since it is language to which we are translating. The translation model corresponds to a model trained using the current state-of-the-art Moses toolkit (Koehn et al., 2007), which estimates a statistical phrase-based log-linear model. This model is built by extracting bilingual phrases (understood as segments of consecutive source-target words) from word-aligned parallel text corpora. Then, several scoring models are estimated from these extracted bilingual phrases.

Again, similarly to ASR, SMT systems do not significantly improve with the inclusion of large amounts of out-domain resources. So, representative data from MOOCs are needed in order to train an effective SMT system. In the case of EMMA, this means translated MOOC-related material. However, this kind of in-domain translated material is usually scarce, and a selection of out-domain parallel sentences could be used to improve SMT performance instead.

In this regard, intelligent selection techniques have been proposed to extract from the out-domain parallel corpora those bilingual sentences that would be useful to train the in-domain SMT system and provide better translation quality. This is especially appealing in the MOOC context, where courses to be translated correspond to specific domain content that cannot be easily translated by a general-purpose SMT systems. Intelligent selection techniques are based on similarity measures computed between the in-domain and out-domain texts. Using these measures, relevant texts from the out-domain data are extracted. Finally, the SMT system will be trained on the in-domain data plus the selected set of the out-domain corpora.

Results

In this section, we describe the experiments performed to assess the quality of the described ASR and SMT systems for Dutch. These experiments have been performed on the contents of MOOCs from the Open University of the Netherlands. As evaluation approach we have chosen an empirical approach stemming from pattern recognition. This evaluation consists in an experiment in which some annotated data, i.e. a set of video lectures to be transcribed or a text to be translated, are automatically transcribed and its error estimated in comparison with the correct transcription or translation. Specifically, this annotated data is split into two sets: development and test. First, the development set is employed to tune system parameters. Then, the test set is automatically

transcribed or translated using the best parameters obtained on the development set. The transcription and translation quality is measured on both, the development and test sets. The quality on the development set corresponds to an optimistic measure of the system performance, as the system has been tuned on it. On the other hand, the quality gauged on the test set represents a more realistic performance measure, as this data set has been involved neither on the training phase, nor on the tuning phase.

Dutch ASR Evaluation

This section describes the evaluation of the ASR system developed for automatically transcribing Dutch video lectures. Concretely, four videos included in the first units of the MOOC on E-learning from the Open University of the Netherlands were selected for the evaluation. Next, these four videos were automatically transcribed using a general-purpose ASR system. A lecturer of the course volunteered to review the automatic transcriptions using post-editing with the transLectures player

interface. An example of the interface can be observed in Figure 2. Once this process was completed, reviewed transcriptions were compared to automatic transcriptions in order to automatically assess the accuracy of the ASR system.

The time devoted to the review process is measured in a measure called Real Time Factor (RTF). RTF measures the ratio between the time needed to post-edit the transcription and the total duration of the video. In our case, the review process took 6 RTF, that is, post-editing the transcription of 1-hour video requires six hours. This result is quite satisfying for a first evaluation, as the manual transcription from scratch by non-expert transcribers is usually reported to cost 10 RTF.

The four selected videos account for 1.8 hours of speech, which is a good quantity for an empirical evaluation. Next, as explained before, the videos were split into two sets: the development set and the test set. It must be noted that, the sets contain two different speakers each, resulting in speaker-independent sets. Table 2 shows basic statistics of these sets.

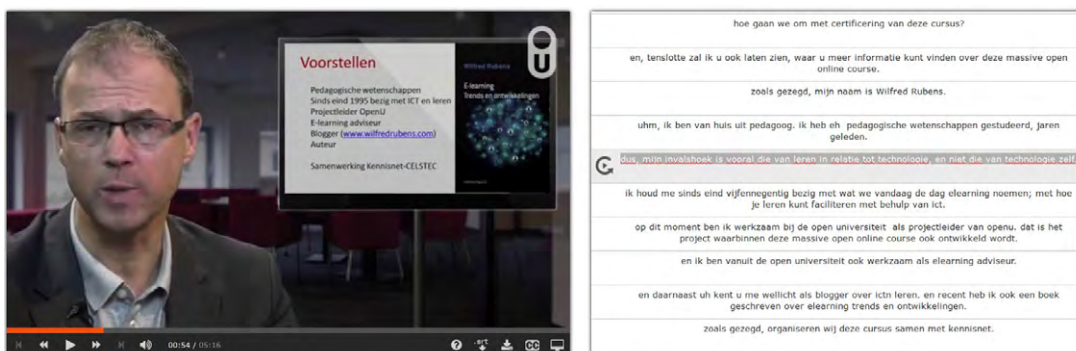


Figure 2. Example of transLectures editing interface for Dutch

Table 2. Statistics of Dutch ASR evaluation data

Set	Videos	Duration	Running Words (K)	Vocabulary(k)
Development	2	00:53:19	8.1	1.2
Test	2	00:52:51	9.4	1.1
Total	4	01:46:10	17.5	2.3

The quality of the ASR system is evaluated in terms of Word Error Rate (WER). WER is measured as the mean number of natural editions (substitutions, deletions and insertions) that have to be applied to transform the automatic transcription into the reviewed transcription. WER is thoroughly employed in the ASR literature and it has been shown to correlate well with human evaluation. Table 3 shows the results obtained in the evaluation. The

Table 3. WER results of Dutch ASR

Set	WER
Development	27.2
Test	27.5

results of Table 3 reflect that the current Dutch ASR system incorrectly recognizes one word out of four on average. This result is in line with those obtained by state-of-the-art ASR systems, such as that used in YouTube. When comparing the transcription quality of our videos being processed by the general-purpose YouTube ASR system, WER was 38.6, which is 9 points worse than the 27.5 obtained with our ASR system. This difference is mainly due to the application of adaptation techniques to the specific content of the videos, which the general-purpose YouTube ASR system does not perform.

In order to better analyse the transcription results of our system we performed an error analysis. Several transcription errors are caused by intermingling English words into the Dutch speech (the ASR system only expects Dutch words), non-native Dutch speakers (mostly Germans) and out-of-vocabulary words (unknown words for the ASR system).

Dutch to English SMT system evaluation

This section describes the evaluation of the SMT system developed for automatically translating Dutch into English for OUNL MOOCs. First of all, the needed parallel data for the evaluation was generated. We selected as the evaluation data the transcriptions of the same four videos that were reviewed in the ASR evaluation together with the introductory web texts of the two OUNL MOOCs. Similarly as in ASR, the translation of the evaluation data was performed by applying post-editing. First, a general-purpose SMT system was built with out-domain data. Next, this system was employed to automatically translate all the Dutch texts into English. Last, the lecturer reviewed the translations using the transLectures interface as shown in Figure 3.

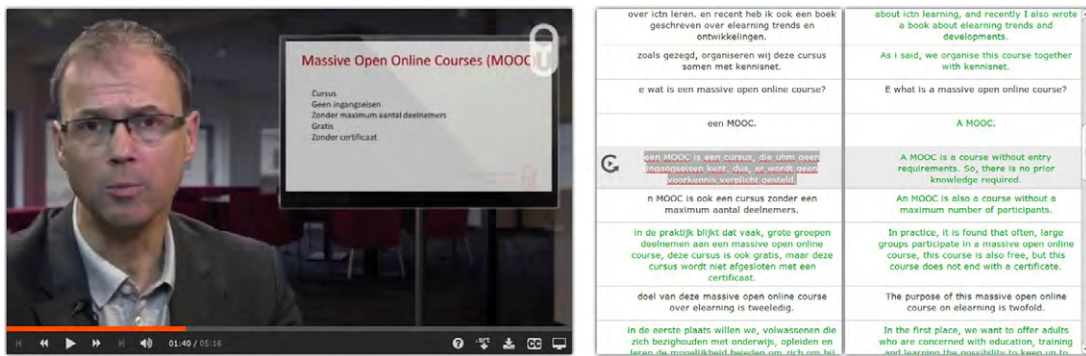


Figure 3. Example of the transLectures translation interface

Table 4. Statistic of Dutch ASR evaluation data

Set	Sentences	Running Words (K)		Vocabulary (K)	
		Nl	En	Nl	En
Development	725	16.5	14.4	2.1	1.6
Test	731	16.4	16.6	2.1	1.6
Total	1465	32.8	31.0	4.2	3.2

The translation review process took 12.2 RTF to be performed. This result is quite satisfying compared to the generation of manual translation, which costs 30 RTF on average. The main reason behind this result is the good quality of the general-purpose SMT system, which would be further improved when the in-domain data was generated during the translation review process. Similarly to ASR, the reviewed translations were split into two sets: a development set and a test set. Table 4 shows the basic statistics of these datasets.

As in ASR, the SMT system for Dutch was automatically evaluated on the development and test sets in terms of the Bilingual Evaluation Understudy (BLEU) score (Papineni, Roukos, Ward & Zhu, 2002). BLEU is the geometric mean of n-gram overlapping (precision) between the automatic and the reviewed translation, penalised by the ratio between the automatic and the reviewed translation

when the former is shorter than the latter. Several authors state that BLEU score correlates well with human judgement (Coughlin, 2003). For this reason, BLEU has become the conventional accuracy measure in SMT.

In our evaluation experiments, SMT systems based on different data selection techniques were compared in terms of BLEU score.

Table 5 depicts the results for the best performing technique on the development set and its accuracy on the test set.

Table 5. Results of the SMT system for Dutch to English

	BLEU
Development	38.5
Test	38.0

As observed in Table 5, the resulting BLEU score is at the same level of state-of-the-art and commercial systems (Bojar et al., 2014). When submitting our testset to Google Translate a BLEU score of 33.3 was obtained. Like ASR the difference in quality can be attributed to the adaption in our system to the contents of the texts. It must be noted also that, the RTF obtained in the user evaluation was 12.2 for a general-purpose SMT system which obtained a 35.7 BLEU score on the test set. Therefore, it is expected that the translations generated by the system reported in Table 5 will require less RTF to be reviewed.

Acknowledgements

This work is partially funded by the EU under the Competitiveness and Innovation Framework Program 2007- 2017 (CIP) in the European Multiple MOOC Aggregator (EMMA) project. Grant Agreement no. 621030.

References

- Alidou, H., Glanz, C., & Nikiéma, N. (2011). Quality multilingual and multicultural education for lifelong learning. *International Review of Education*, 57(5-6), 529-539. doi: 10.1007/s11159-011-9259-z
- Anastasiou, D., & Gupta, R. (2011). Comparison of crowdsourcing translation with Machine Translation. *Journal of Information Science*, 37(6), 637-659. doi: 10.1177/0165551511418760
- Berlanga, A. J., Kalz, M., Stoyanov, S., Van Rosmalen, P., Smithies, A., & Braidman, I. (2009, July). Using Language Technologies to Diagnose Learner's Conceptual Development. In *Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on* (pp. 669-673). IEEE.

Discussion

This work is a first approach to the automatic transcription and translation of Dutch MOOCs on the EMMA platform. The results obtained are encouraging, as the systems developed are in range of state-of-the-art system performance on this application. These results were also corroborated on user evaluations devoted to generate the evaluation data to automatically assess the ASR and SMT systems. User evaluations consisted in a review process based on the post-edition of automatic transcriptions and translations. In both cases, the average time devoted to review automatic transcriptions and translations was reduced by 50% with respect to do the same task from scratch. Moreover, it must be noted that user evaluations were performed on transcriptions and translations generated by general-purpose systems, it is expected that ASR and SMT systems tuned on in-domain resources produce higher quality transcriptions and translations that further reduce the review effort.

Future work includes the improvement of current ASR and SMT systems incorporating more in-domain material. In addition, as foreign words and non-native speakers have shown to be an important source of errors, a multilingual approach to ASR is needed to deal with the peculiarities of our application and improve the overall system performance.

- **Bellegarda J.** (2000). Statistical language model adaptation: review and perspectives. *Speech Communication*, 42(1), 93-108.
- **Bisani M., Ney H.** (2008) Joint-sequence models for grapheme-to-phoneme conversion. *Speech Communication*, 50(5), 434-451.
- **Bojar O., Buck C., Federmann C., Haddow B., Koehn P., Leveling J., ... and Tamchyna A.** (2014). Findings of the 2014 Workshop on Statistical Machine Translation. *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pp 12-58. Brussels: European Commission.
- **Campbell, M., Devine, J., Gonzales, J., Halasz, G., Jenner, C., Jonk, A., Hultin, G., Münz, R., Schmitz, M., & Strietska-Illina O.** (Eds). *New Skills for New Jobs: Action Now. A report by the Expert Group on New Skills for New Jobs prepared for the European Commission.*
- **Crawford, J.** (2000). At war with diversity: US language policy in an age of anxiety. *Multilingual matters* (25).
- **Coughlin, D.** (2003) "Correlating Automated and Human Assessments of Machine Translation Quality" in *MT Summit IX*, New Orleans, USA, 23-27
- **del-Agua, M. A., Giménez, A., Serrano, N., Andrés-Ferrer, J., Civera, J., Sanchis, A., & Juan, A.** (2014). The translectures-UPV toolkit. In *Advances in Speech and Language Technologies for Iberian Languages* (pp. 269-278). *Springer International Publishing.*
- **Ding, Y., Wang, M., He, Y., Ye, A. Y., Yang, X., Liu, F., ... & Wei, L.** (2014). "Bioinformatics: Introduction and Methods," a Bilingual Massive Open Online Course (MOOC) as a New Example for Global Bioinformatics Education. *PLoS Computational Biology*, 10(12), e1003955.
- **Hinton G., Deng L., Yu D., Mohamed A., Jaitly N., Senior A., Vanhoucke V., Nguyen P., Tara Sainath George Dahl, and Brian Kingsbury** (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 29(6), 82.
- **Hu, C., Resnik, P., & Bederson, B. B.** (2014). Crowdsourced Monolingual Translation. *ACM Transactions on Computer-Human Interaction*, 21(4), 1-35. doi: 10.1145/2627751
- **Kalz, M., Specht, M., Nadolski, R., Bastiaens, Y., Leirs, N., & Pawlowski, J.** (2010). OpenScout: Competence based management education with community-improved open educational resources. In Halley et al. (Eds.), *Proceedings of the 17th EDINEB Conference. Crossing Borders in Education and work-based learning* (pp. 137-146). Maastricht, The Netherlands: FEBA ERD Press.
- **Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., ... & Herbst, E.** (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions* (pp. 177-180). Association for Computational Linguistics.
- **Kunchukuttan, A., Roy, S., Patel, P., Ladha, K., Gupta, S., M. Khapra, M. and Bhattacharyya, P.** (2012). Experiences in Resource Generation for Machine Translation through Crowdsourcing. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*, 23-25. Istanbul, Turkey.
- **Liyanagunawardena, T., Adams, A. A., & Williams, S. A.** (2013). MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distance Learning*, 14(3), 202-227.
- **Oostdijk, N.** (2000). The Spoken Dutch Corpus. Overview and First Evaluation. In Zampolli, A. (Ed) *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC)*. Athens, Greece.
- **Papineni, K., Roukos, S., Ward, T., & Zhu, W. J.** (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 311-318). Association for Computational Linguistics.
- **Rabiner L.** (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2): pp. 257-286.
- **Serrano N., Gimenez A., Civera J., Sanchis A., Juan A.** (2014) Interactive handwriting recognition with limited user effort. *International Journal on Document Analysis and Recognition (IJ DAR)*. 17(1), 47-59.
- **Silvestre-Cerdà J. A., del-Agua M. A., Garcés G., Gascó, G., Giménez, A., Martínez, A., ... & Juan A.** (2012). transLectures. In *Online Proceedings of VII Jornadas en Tecnología del Habla and III Iberian SLTech Workshop (IberSpeech 2012)*, Madrid (Spain), pp. 345-351.
- **Stanley F. Chen & Joshua Goodman** (1996). An empirical study of smoothing techniques for language modeling. In *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*, pp. 310-318.
- **Stolcke A.** (2002) Srilm-an extensible language modeling toolkit. In *Proceedings International Conference on Spoken Language Processing*, pp. 257-286.