transLectures: Transcription and Translation of Video Lectures

Carlos Turró, Alfons Juan and Jorge Civera, Universitat Politècnica de València
turro@cc.upv.es, ajuan@dsic.upv.es, jcivera@dsic.upv.es
Davor Orlič, Mitja Jermol, Jožef Stefan Institute
davor.orlic@ijs.si, mitja.jermol@ijs.si

Abstract
*trans*Lectures is a FP7 project aimed at developing innovative, cost-effective solutions to produce accurate transcriptions and translations in large repositories of video lectures. This paper describes user requirements, first integration steps and evaluation plans at transLectures case studies, *VideoLectures.NET* and *poliMedia*.

1. Introduction

Online educational repositories of video lectures are rapidly growing on the basis of increasingly available and standardised infrastructure. Also, transcription and translation of video lectures is needed to make accessible those repositories to speakers of different languages and to people with disabilities. However most lectures are neither transcribed nor translated because of the lack of efficient solutions to obtain them at a reasonable level of accuracy and cost.

*trans*Lectures (translectures.eu) is a FP7 project aimed at developing innovative, cost-effective solutions to produce accurate transcriptions and translations. Our starting hypothesis is that there is only a relatively small gap for the current technology on automatic speech recognition and machine translation to achieve accurate enough results in the kind of audio-visual object collections we are considering; and that this gap can be closed by using *massive adaptation* and *intelligent interaction with users*.

We will first test our tools in two repositories: *VideoLectures.NET* and a smaller repository of Spanish video lectures, *poliMedia*. For transcription, we consider English and Slovenian in *VideoLectures.NET*, which account for more than 90% of lectures, and Spanish in *poliMedia*. For translation, we consider the language pairs: en⇆es, en⇆sl, en→fr and en→de.

Although *trans*Lectures is still in its first year, its case studies have already defined user requirements at different levels; they have already taken first integration steps, and also established evaluation plans. In this paper, after a description of *VideoLectures.NET* and *poliMedia* in Section 2, user requirements are described in Sections 3 and 4 (at a technical level, and at the level of use scenarios, respectively), Opencast Matterhorn integration and first project results in Sections 5 and 6, and evaluation plans in Section 7.

2. Case Studies

2.1 VideoLectures.net

VideoLectures.NET was founded in 2001 as an internally-funded project and is now run by the dedicated Center for Transfer in Information Technologies at the Josef Stefan Institute (JSI), Ljubljana, Slovenia. It is a free and open access repository of video lectures mostly filmed by people from JSI at major conferences, summer schools, workshops and science promotional events from many fields of Science. Indeed, *VideoLectures.NET* is being used as an educational platform for several EU funded research projects; different open educational resources organisations such as The OpenCourseWare Consortium, MIT OpenCourseWare and Open Yale Courses; as well as other scientific institutions like CERN. In this way,

VideoLectures.NET collects high quality educational content which is recorded with also high-quality, homogeneous standards.

There are more than 10000 unique users visiting *VideoLectures.NET* every day, with more than 15.000 registered users. All content in *VideoLectures.NET* falls under Creative Commons 3.0 license. In addition, for more than 70% of the content, *VideoLectures.NET* holds the written and signed consent of authors to reuse materials.

2.2. poliMedia

poliMedia is a recent, innovative service for creation and distribution of multimedia educational content at the Universitat Politècnica de València (UPV). It is mainly designed for UPV professors to record courses on video blocks lasting 10 minutes at most. It serves more than 36000 students and 2800 professors and researchers. Started only 4 years ago, it is being exported to several universities in Spain and America for free.

As in VideoLectures.NET, video blocks are accompanied with time-aligned slides though, in contrast to VideoLectures.NET, video recordings are carried out at specialised studios under controlled conditions to ensure maximum recording quality and homogeneity. Indeed, professors are filmed against a constant-colour background to postproduce presentations in which only the professor's body is shown, properly scaled, together with slides. As of June 2011, poliMedia catalogue included more than 5700 videos accounting for more than 1000 hours.

3. Technical requirements

At a technical level, *trans*Lectures tools will satisfy five basic requirements at least:

1. *Accuracy estimation for each transcription and translation:* Clearly, our main requirement is that project tools need to be as much accurate as possible. However, as it is unrealistic to produce highly accurate transcriptions and translations for every video lecture, *trans*Lectures tools should provide a global estimation of the accuracy of each transcription and translation. This can be done using confidence measures (Ueffing and Ney, 2007; Sanchis et al., 2007; Wessel, 2001). In this way, case studies will define an accuracy threshold to be reached for a transcription or translation to be delivered to the user.

2. *Adjustable computational behaviour:* Project tools for massive adaptation and intelligent interaction may be highly demanding in terms of computational cost. However, case studies computational resources are limited, and thus tools have to be efficient and adjustable in terms of computational behaviour.

3. *Output constrained to user preferences and corrections:* User interaction imposes additional technical requirements, since the system should be able to constrain output in accordance with user preferences and corrections. The user may decide to partially supervise a transcription or translation by entering sequences of words that should appear in the final translation. Then, the system should be able to propose the best translation taking into account these users' constrains.

4. *Fast learning from user corrections:* User corrections on transcriptions and translations need to be immediately incorporated into the underlying statistical models. To this purpose, incremental and on-line training techniques are explored to provide the best user experience and flexibility, since users can directly observe how their corrections are propagated to later suggested translations.

5. *User accessibility*: User accessibility is a must in *trans*Lectures, so pilot systems will be deployed in HTML5. Pilot systems are basically video players with special subtitling and multilingual functionality.

4. User roles

At the level of use scenarios, *trans*Lectures tools will consider five user roles:

1. Viewer: A viewer is just a user looking at a particular translation or translation. So she will have a simple set of features, allowing her to view a synchronized subtitles track if the confidence level of that video is above the level set by the editor or the author. The player should allow a user to become collaborative user.
2. Collaborative viewer: Users with that role will have access to an advanced set of features from the transLectures engine; for them the interface will provide a confidence level for the overall transcription/translation, and also will provide alternate translations for difficult parts. The confidence level required to display a translation for a collaborative user will be much lower than the required for a standard viewer. It is expected that collaborative users can be asked through the interface for simple tasks, like transcribing some parts of the video. Input from collaborative users will add information to the translation, but will not replace the content, and will not trigger an update on the acoustic model.
3. Expert (e.g. professional translators): An expert is a collaborative viewer with expertise in that field. So, experts can replace transcriptions and translations in parts of the video. Also an input from an expert will trigger an adjustment on the acoustic model.
4. Author: An author is the owner of a video, and will be always taken as an expert for his uploads. Also he can decide if subtitles can be displayed by viewers, based on the average confidence level.
5. Editor: An editor can set a confidence level for a whole site.

5. Integration into Opencast Matterhorn

Matterhorn is a free, open-source platform to support the management of educational audio and video content. Institutions use Matterhorn to generate lecture recordings, manage existing video, serve designated distribution channels, and provide user interfaces to engage students with educational videos. The Opencast Community and its Matterhorn project will provide the scientific expertise, organisational capacity, and primary international channel to develop, deploy, and test *trans*Lectures tools in both controlled and open settings so as to enable real-life evaluation. Moreover, Matterhorn provides a framework of services around the management of academic video that institutions can customise to meet their individual needs.

6. First project results

First project results include about 800 lectures transcribed and preliminary integration steps of translectures into poliMedia for several users' profiles.

In Figure 1, a poliMedia player with transLectures integration is shown for three user roles: viewer (left), collaborative viewer (middle) and author (right). In all cases, languages with available transcriptions/translations are indicated together with a colour code for their quality.
Figure 1. transLectures player for viewer (left), collaborative viewer (middle) and author (right).

In the case of a collaborative viewer, the player allows to select alternative transcriptions/translations by clicking on the text. Corrections from collaborative viewers will

be submitted to the transLectures engine to update current transcriptions/translations and improve underlying statistical models. Finally, in the author case, any part of the text can be selected and corrected if needed. It is worth noting that the author also has tooltips associated with difficult parts (in red).

7. Evaluation

Internal and external evaluations will carried out. More precisely, two internal, user evaluations will be organised at each case study site so as to evaluate models, tools and integration progress in a real-life yet controlled setting. The user groups and evaluation procedures will be planned in accordance with each site requirements.

8. Conclusions

*trans*Lectures is a FP7 project aimed at developing innovative, cost-effective solutions to produce accurate transcriptions and translations in large, Matterhorn-related repositories of video lectures. In this paper, we have described its case studies, user requirements, first Matterhorn integration steps and evaluation plans, and also we have taken a peek at the implementation of transLectures for the poliMedia case study.

References

Nicola Ueffing and Hermann Ney (2007). Word-level confidence estimation for machine translation. *Computational Linguistics*, 33(1):9–40, March 2007.

Albert Sanchis, Alfons Juan, and Enrique Vidal (2007). Estimation of Confidence Measures for Machine Translation. In *Proc. of the Machine Translation Summit XI*, pages 407–412, Copenhagen (Denmark), September 2007.

Frank Wessel, Ralf Schlüter, Klaus Macherey, and Hermann Ney (2001). Confidence measures for large vocabulary continuous speech recognition. *IEEE Transactions on Speech and Audio Processing*, 9(3):288–298, March 2001.

License