

# MLLP Transcription and Translation Platform

Alejandro Pérez González de Martos, Joan Albert Silvestre-Cerdà, Juan Daniel Valor Miró, Jorge Civera, and Alfons Juan

Universitat Politècnica de València,  
Camino de Vera s/n, 46022 Valencia, Spain  
{aperez, jsilvestre, jvalor, jcivera, ajuan}@dsic.upv.es

**Abstract.** This paper briefly presents the main features of MLLP's Transcription and Translation Platform, which uses state-of-the-art automatic speech recognition and machine translation systems to generate multilingual subtitles of educational audiovisual and textual content. It has proven to reduce user effort up to 1/3 of the time needed to generate transcriptions and translations from the scratch.

**Keywords:** video lectures, automatic speech recognition, machine translation, efficient video subtitling, post-editing, document translation

## 1 Introduction

Multilingual access to educational content is needed to overcome language barriers, improve accessibility to educational resources and contribute to emerging trends in online education. A remarkable example on these trends are Massive Open Online Courses (MOOCs), usually consisting of video lectures, readings, and arrangements sets. Thousands of students enroll every year in these courses, offered by different institutions such as EdX [3], Coursera [2], Iversity [4], etc. However, according to the European MOOC Monitor, most of them are delivered in a limited set of languages (45% in English, 32% in Spanish, 14% in French and the remaining 9% in other languages). This fact keeps a significant portion of potential students from taking these courses. The manual generation of multilingual subtitles for audiovisual content, or the manual translation of text resources is a time-consuming task that requires an important human effort. Automatic technologies, such as automatic speech recognition (ASR) and machine translation (MT), still far from perfect, can help to reduce effort on having course contents accurately translated [8]. In addition of making this content available to speakers of different languages, subtitles can be exploited to incorporate features such as searchability, content summarisation, recommendation, etc. The lack of efficient solutions for translating multimedia content has led the MLLP research group to build a web-based automatic Transcription and Translation Platform (TTP).

## 2 TTP

TTP [6] is an online transcription and translation platform to create accurate transcriptions and translations from multimedia content. Using top-notch ASR and MT technologies [1], TTP offers fine-tuned systems able to exploit the most typical characteristics of the educational domain content to obtain the best possible automatic results [5]. In addition, it includes user-friendly interfaces that have been carefully designed to minimise user effort on the error supervision process.

There are two main workflows depending on which kind of material is going to be translated: audio and video files on one side, and text and documents on the other.

### 2.1 Multilingual subtitling of audiovisual content

Figure 1 shows TTP Ingest Service workflow, which is responsible of generating automatic transcriptions and translations. Video and audio files can be uploaded to the platform in a Media Package form to be automatically transcribed (Transcription Generation) and translated (Translation(s) Generation) to all supported languages. As of June 2015, TTP includes ASR systems for English, Spanish, Catalan, French, Estonian, Italian, Dutch, Portuguese and German. As for translations, all ASR supported languages can be translated into English, English into French and Italian, and any combination of the trio {English, Spanish, Catalan}. Text-to-speech synthesised tracks can be generated from translations as an additional final step (Text-To-Speech Track(s) Generation).

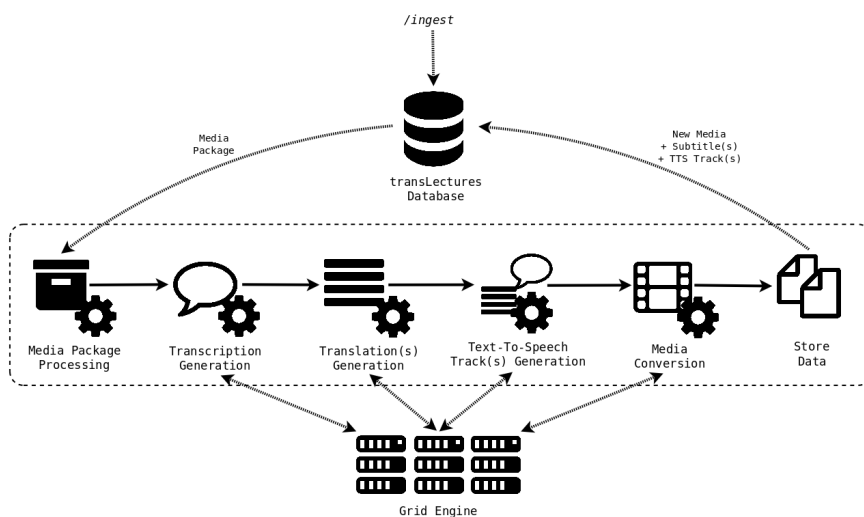


Fig. 1. Ingest Service.

Nonetheless, accurate translations are unlikely to be achieved through fully-automated approaches alone. In order to reach the desired levels of accuracy, we must consider user interaction. For that purpose, TTP includes a caption editor that allows users to easily supervise and correct transcription and translation errors on their videos (see Figure 2). On the left, the video/audio file is played showing the current subtitles, which can be aligned in time and modified using the mouse on the bottom part (enlarge, reduce, join, split, etc.). On the right side, subtitles text is shown synchronously with the video, and can be edited simply clicking on top of it.

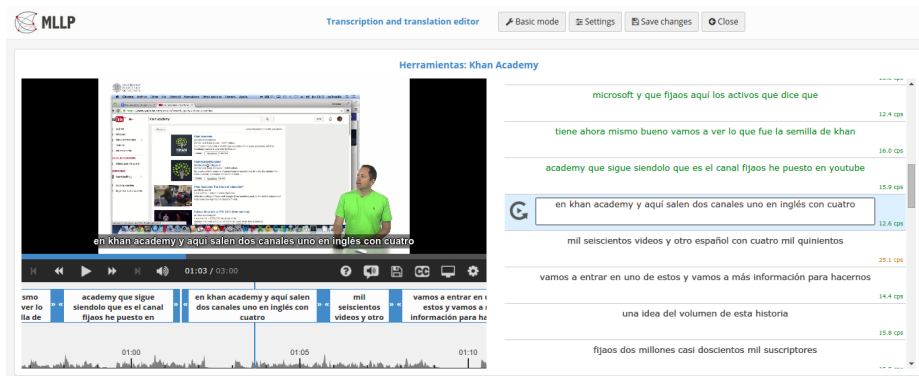


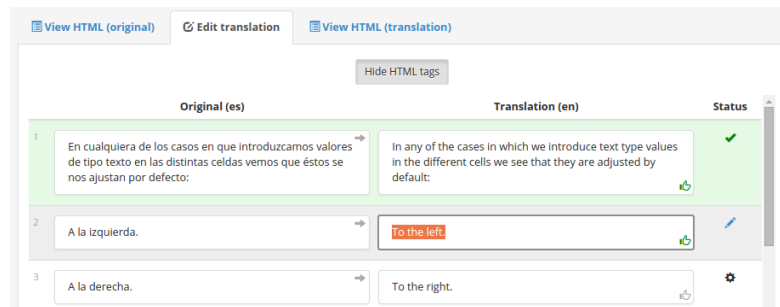
Fig. 2. TTP Caption Editor.

## 2.2 Document translation

MOOCs often include documents (text materials) that have to be also translated. These documents can be uploaded to TTP to be translated into different languages. As in the case of captions, the platform includes a document translation editor that allows users to correct the translations sentence by sentence in an efficient way (see Figure 3). On top, user can select three different views: the source document (left), the target document (right), and the editing view (center). When editing, source sentences (left) are shown aligned with their corresponding translations (right).

## 2.3 Integration into other platforms

Every interaction with the platform (uploading media, transcribing, translating, downloading subtitles, etc.) can be programmed using TTP API REST services. There is detailed documentation and different libraries available (PHP, Python) to simplify the integration process into any external repository [7].



**Fig. 3.** TTP Document Translation Editor.

### 3 Conclusions

We have briefly presented MLLP’s Transcription and Translation Platform, which offers organisations the possibility to have all their educational content accurately transcribed and translated into several languages. It has proven to reduce user effort on generating transcriptions and translations of multimedia material up to 1/3 of the time needed to do it from the scratch [9].

### Acknowledgments

The research leading to these results has received funding from the European Union FP7/2007-2013 under grant agreement no 287755 (transLectures) and ICT PSP/2007-2013 under grant agreement no 621030 (EMMA), and the Spanish MINECO Active2Trans (TIN2012-31723) research project.

### References

1. del Agua, M.A., et al.: The transLectures-UPV toolkit. In: Proc. of IberSpeech (2014)
2. coursera.org: Take the World’s Best Courses, Online, For Free. [coursera.org](http://coursera.org) (2011)
3. EdX: Free online courses from the world’s best universities. [edx.org](http://edx.org) (2012)
4. Iversity: Study Anywhere. [iversity.org](http://iversity.org) (2013)
5. Martínez-Villaronga, A., et al.: Language model adaptation for video lectures transcription. In: Proc. of ICASSP. pp. 8450–8454 (2013)
6. MLLP Research Group: MLLP Transcription and Translation Platform (TTP). <http://ttp.mllp.upv.es> (2015)
7. MLLP Research Group: MLLP Transcription and Translation Platform (TTP) Documentation. <http://ttp.mllp.upv.es/doc> (2015)
8. Valor Miró, J.D., et al.: Evaluating intelligent interfaces for post-editing automatic transcriptions of online video lectures. *Open Learning: The Journal of Open, Distance and e-Learning* 29(1), 72–85 (2014)
9. Valor Miró, J.D., et al.: Efficient generation of high-quality multilingual subtitles for video lecture repositories. In: Tenth European Conference on Technology Enhanced Learning (EC-TEL 2015). *Lecture Notes in Computer Science*, Springer (2015)

## 1 TTP presentation at EC-TEL 2015 booth

The platform will be shown in a laptop with Internet access. An overall view of the platform will be given, and the two different workflows will be shown in detail. Assistants will be able to see how the platform works, what is the platform performance and accuracy on generating automatic subtitles and translations, and the different interfaces that have been specifically developed for post-editing purposes (see Figures 1 and 2).

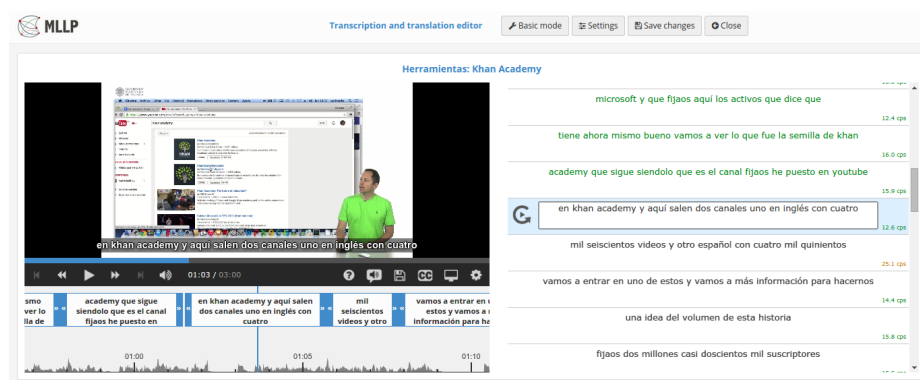


Fig. 1. TTP Caption Editor.

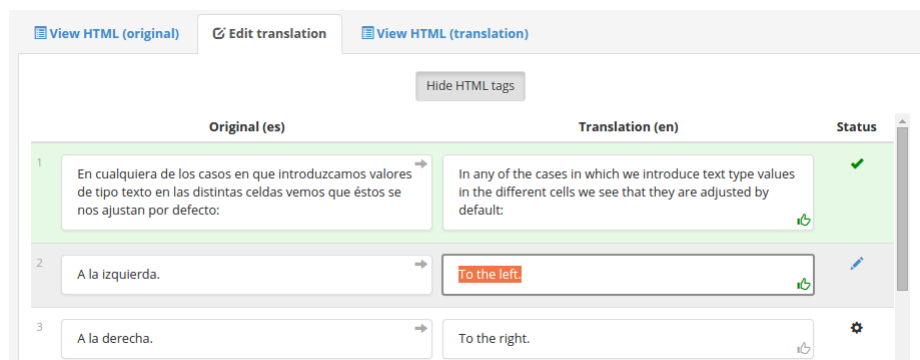


Fig. 2. TTP Document Editor.

It is also our intention to record a few portion of a presentation given by any of the EC-TEL 2015 participants (upon approval of the person being recorded) and use TTP for assistants to assess transcription and translation results.

The idea is to record few minutes of the participant talk with a mobile phone. Then, the video will be uploaded to TTP to be automatically transcribed and translated (see Figures 3 and 4).

The screenshot shows the 'Media information (Step 1 / 3)' form. On the left sidebar, 'Upload media' is highlighted with a red circle. The main form area contains the following fields and instructions:

- Title\*:** A text input field with a note: "The title will be used to search for related documents on the web to improve the quality of the transcription."
- Media file\*:** A section with a red asterisk note: "\* Only if Topic adaptation is enabled in Step 2". It includes a text input for "Enter URL (YouTube and Vimeo URLs are supported)" and a file selection button labeled "Seleccionar archivo" with the text "Ningún archivo seleccionado" and "File extensions supported" below it.
- Media language\*:** A dropdown menu with the instruction "Select the language spoken on the media file." and a "Next →" button at the bottom right.

**Fig. 3.** TTP Upload Form (Step 1).

The screenshot shows the 'Translation and text-to-speech (Step 3 / 3)' form. The left sidebar remains the same. The main form area contains the following fields and instructions:

- Translate to:** A multi-select dropdown menu with options: "Español (English-Spanish MT System)", "Català (English-Catalan MT System)", "Italiano (English-Italian MT System)", and "Français (English - French MT System)". A note below says "Hold down Ctrl/Cmd key to select multiple languages".
- Generate TTS tracks:** A checkbox labeled "Generate text-to-speech audio tracks" which is checked. Below it, it says "Available for: English (only female)".
- Buttons for "← Back" and "Upload" are located at the bottom of the form.

**Fig. 4.** TTP Upload Form (Step 3).

Then, once transcriptions and translations are ready, we will access to the video to check their quality. We will show, in addition, how the post-editing interface can be used to correct transcription and translation errors.