



transLectures: Transcription and Translation of Video Lectures

Funding agency: European Commission

Funding call identification: FP7-ICT

Type of project: STREP

Project ID number: 287755

<http://www.translectures.eu/>

List of partners
Universitat Politècnica de València, Spain (coordinator)
Xerox Research Centre Europe, France
Jozef Stefan Institute and Knowledge for All Foundation, Slovenia and UK
RWTH Aachen University, Germany
European Media Laboratory GmbH, Germany
Deluxe Digital Studios Limited, UK

Project duration: November 2011 — October 2014

Summary

Online educational repositories of video lectures are rapidly growing. An example of this is *VideoLectures.NET*, a free and open access educational video lectures repository. Transcription and translation of video lectures in *VideoLectures.NET* is needed to make them accessible to a wider audience. However, similarly to other repositories, most lectures are neither transcribed nor translated.

The aim of **transLectures** is to develop innovative, cost-effective solutions to produce accurate transcriptions and translations in *VideoLectures.NET*, with generality across other Matterhorn-related repositories. The 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 object collections we are considering; and that it can be closed by achieving the following three objectives:

1. *Improvement of transcription and translation quality by massive adaptation.*
2. *Improvement of transcription and translation quality by intelligent interaction.*
3. *Integration into Matterhorn to enable real-life evaluation.*

The main result of **transLectures** will be a set of cost-effective tools to produce accurate transcriptions and translations in *VideoLectures.NET* and other Matterhorn-related repositories. Indeed, these tools will be tested in *VideoLectures.NET* and in a smaller repository of Spanish video lectures, *poliMedia*. For transcription, we will consider English and Slovenian in *VideoLectures.NET* (more than 90%) and Spanish in *poliMedia*. For translation, we will consider the language pairs: en↔es, en↔sl, en→fr and en→de.

Upon successful achievement of the project, its techniques will probably spread over many educational repositories, enabling them to overcome language barriers and reach wider audiences.