D7.3: Final report on dissemination

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

Distribution: Public

transLectures
Transcription and Translation of Video Lectures

ICT Project 287755 Deliverable D7.3

October 31, 2014
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 Media Europe (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, 46022 València, Spain
aju@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

© 2014, The Individual Authors

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.
Abstract

The purpose of this deliverable is to report on all dissemination activities and tools developed during the third and final year of the transLectures project.

Contents

1 Executive summary 4
  1.1 Highlights from Y3 4

2 Dissemination strategy 5
  2.1 Review of Y1 5
  2.2 Review of Y2 5
  2.3 Strategy going into Y3 5

3 Dissemination activities M25-36 6
  3.1 Event organisation 6
    3.1.1 Internet of Education conference (IoE2013) 6
    3.1.2 OpenCourseWare Consortium Global Conference 2014 6
    3.1.3 transLectures satellite workshop 7
    3.1.4 UNESCO Chair Workshop 8
    3.1.5 Pre-conference workshop at Online Educa Berlin 8
  3.2 Publications 8
  3.3 Open source tools 10
  3.4 JSI-K4A dissemination campaign 11
    3.4.1 Opening Up Slovenia 13
    3.4.2 The UNESCO Chairs Programme 13
    3.4.3 VideoLectures.NET 14
  3.5 Dissemination Timeline M24-M36 14
    3.5.1 Confirmed events beyond project end 18

4 Business plan and exploitation 18
  4.1 Summary of D7.5 18
  4.2 Current deployment 19
  4.3 Possible partnerships 20

5 Evaluation of dissemination activities 21
  5.1 References to transLectures in other publications 21
  5.2 Web analytics 22
  5.3 Subjective feedback 23

6 Conclusions 24

A Appendix 28
  A.1 Events timeline M1-M36 28
  A.2 Publications M1-M36 31
  A.3 Dissemination material 33
1 Executive summary

The aim is to develop a set of tools that can be used to automatically generate quality transcriptions and translations for these video lectures and, by doing so, make their content accessible to the deaf and hard-of-hearing, to speakers of other languages and to language learners. They will also allow them to be watched in noisy and noise-restricted environments. The vast amounts of text data generated using these tools can be used to derive advanced digital content management features, such as lecture classification, recommendation, fragmentation, etc., thereby improving user experience above and beyond accessibility to the content of a given video.

At the scientific level, the goals of transLectures are to advance the state-of-the-art in model adaptation (to the domain, to the speaker, and using title searches and text data extracted from the presentation slides) and intelligent human-machine interaction (including the use of confidence measures).

From the beginning, transLectures has had a strong practical focus, with two case studies and multiple external evaluation phases. Built into the Description of Work is the goal of producing a set of cost-effective transcription and translation tools for use on the two case study sites (VideoLectures.NET and poliMedia), as well as integration into the Matterhorn platform for educational audiovisual management. In fact, project technologies are already integrated into the two case studies. Moreover, transLectures technologies are currently being used to maintain the live poliMedia repository, where subtitles are available in Spanish, Catalan and English.

The aim of dissemination is to enable and raise awareness of all of the above among the key groups of stakeholders established at various stages of the project. In this Final Dissemination Report (D7.3) we will report all dissemination tools and activities that have been developed during the final year of transLectures. First we review strategy over the three years of the project, in Section 2. As in previous deliverables, we present a timeline of all dissemination activities carried out in Y3, plus sections on event organisation, publications, our open source tools, external implementations and other lines of dissemination (all in Section 3).

We recap the Final Business Plan (D7.5) as delivered in M32, as well an overview of the current situation as regards the deployment and uptake of transLectures in its various forms, in Section 4. An appraisal of some performance indicators can be found in Section 5. We end with some conclusions, in Section 6.

1.1 Highlights from Y3

- OCWC2014 and transLectures workshop (Nos. 14, 15 in the Dissemination Timeline 3.5)
- First Internet of Education conference (No. 1)
- UNESCO Chair workshop (No. 25)
- Launch of the Try our Tools (ToT) service (No. 9)
- Release of the transLectures Platform (TLP) (No. 20)
- First formal agreements for the transfer of project technologies (No. 16, 27)
- Two joint publications accepted at AMTA and IberSpeech 2014. (Nos. 29, 33)
2 Dissemination strategy

2.1 Review of Y1

In previous deliverables we identified project stakeholders as researchers, practitioners and other players in the following three broad fields:

- MT, ASR, NLP, machine learning.
- Lecture capture, audiovisual repository management.
- Education sector.

The main idea behind dissemination activities in Y1 was to bring the project and its potential applications to the attention of as many people within these communities as possible.

In this period, ties were established and consolidated with related projects and organisations (META-NET, REC:all, LT-Innovate, SUMAT and PASCAL2). Prof. Harald Höge (Siemens; MT, ASR, ML and NLP expert), Meena Hwang (OCWC) and Olaf Schulte (Opencast) accepted our invitations to form part of the advisory board. Partners attended events organised by Opencast, OCWC, UNESCO and ELRA. transLectures enjoyed considerable local press coverage in two partner countries marking its launch. At the end of this first year, JSI-K4A organised their first event featuring transLectures, the successful “Co-creation of Emerging Trends in Academia” conference.

See the First Report on Dissemination for more details.

2.2 Review of Y2

In Y2 we further refined dissemination strategy, identifying two main and immediate uses for project technologies: transLectures for subtitle generation, and transLectures as text data provider. Accordingly, we focused our attention on the potential user communities. This shift reflected both recommendations from the end-of-year review and work being carried out in WP6. The idea was reach out to these user communities and generate interest in using the tools being developed.

In this period, UPVLC produced two demo videos, revamped the project website, released the first version of the transLectures-UPV toolkit (TLK), also taking the first steps towards establishing external pilots at other Spanish universities. It also led the IASD challenge in partnership with PASCAL2. JSI-K4A, meanwhile, led a strong joint dissemination campaign with X-Like and MediaMixer, fellow EU projects which would fall within transLectures' second use category, bringing the project to the attention of UNESCO, Slovenian government officials, and universities in China and India. This was the year that VideoLectures was declared the best educational product of the decade at the World Summit Awards. Much was also done by these partners to strengthen the relationship with the OCWC, culminating with their winning bid to host the OCWC 2014 global conference in Y3.

See the Second Report on Dissemination for more details.

2.3 Strategy going into Y3

Going into the final year of the project, OCWC2014 and further open source technology releases were to be the cornerstones of work in WP7. Now more than ever the idea was to pave the way for uptake of transLectures tools. There was to be renewed focus on securing non-case study implementations of project technologies, among other things to help inform the Final Business Plan (D7.5) delivered in M32. We also set the goal of getting joint papers published and, as in previous reporting periods, scientific dissemination was to continue throughout.

All activities carried out in Y3 are reported in the following sections.
3 Dissemination activities M25-36

In this section we present all activities carried out in WP7 in this reporting period, highlighting the major areas of work: event organisation, open source releases, external implementations of project technologies and outreach to policy-makers. A chronological list of all activities can be found at the end of this section, complemented in Appendix A.1 by a list of all activities carried out since project kick-off.

3.1 Event organisation

What follows is a list of the events organised by transLectures partners in this reporting period. These events, if we recall, are in addition to JSI-K4A’s Co-creation of Emerging Trends workshop [28] [14] in Y2, plus the IASD Challenge [10] organised by the UPVLC (also in Y2).

3.1.1 Internet of Education conference (IoE2013)

The first Internet of Education [6] conference was held in Ljubljana, Slovenia, organised by JSI-K4A and attended by UPVLC.

Held under UNESCO patronage, the goal of this event was to provide a forum for researchers and policy-makers to explore means of improving the effectiveness of a MOOC-based education, as well as the implications of new edutech trends for the classical university paradigm. transLectures was co-sponsor, alongside the MediaMixer and X-like projects. It came as a continuation of the hugely successful Workshop on Co-Creation of Emerging Trends in Academia [19] [28] and was co-organised by the University College London (UCL). Speakers included academics, researchers and other professionals from UCL, JSI, Coursera, SMEs, and MOOC start-ups.

Gonçal Garcés (UPVLC) presented transLectures, making key contacts in industry and research. For more information, see his comprehensive review [12] of the event, complete with links to the transLectures presentation slides and video [15] (published on VideoLectures.NET).

3.1.2 OpenCourseWare Consortium Global Conference 2014

The culmination of a strong dissemination campaign aimed, among others, at this community, the 2014 OCWC global conference (OCWC2014) [7] was held in Ljubljana, Slovenia, organised by JSK-K4A, and attended by UPVLC and RWTH.

As anticipated at the end of Y2, this was one of the main foci of dissemination activities during this reporting period, and substantial time and resources were devoted to it accordingly.

K4A led the organisation of this conference, overseeing much of the local organisation and the role of programme chair. With the theme “Open Education for a Multicultural World”, the aim of OCWC2014 was to promote the need for technical developments in support of open educational resources (OER), the value of multilingual tools in increasing access to open education, in particular the tools developed within transLectures, and the importance of collaboration across different levels of education, from primary to higher education.

These goals are very well-aligned with UNESCO’s vision and activities to build an inclusive, people-centred, development-oriented Information Society where everyone can create, access and share information and knowledge. Over the past decade, UNESCO has led global initiatives to promote OERs and thereby support both developed and developing nations in reaching internationally-agreed targets in education by 2015. Having already brought transLectures to their attention in previous actions, OCWC2014 was held under UNESCO patronage.
JSI-K4A’s Opening Up Slovenia initiative (see Section 3.4.1 for more details) was launched at this event by the Slovenian Minister of Education, Jernej Pikalo, and the European Commissioner for Education, Culture, Multilingualism and Youth. As silver sponsor, transLectures was also presented as part of the conference’s Dissemination Track by project coordinator Alfons Juan, and Carlos Turrò (both UPVLC).

Our various experiences at OCWC2014 strengthened our sense that we are creating something for which there is both a need and a market. Between the organisation, talk and workshop, project representatives were present throughout the event and were approached by many attendees keen to find out more about the project and how they can make our subtitling tools work in their setting.

A full write-up of which can be found in the project blog [18]. More information regarding this event can be found at the conference website [7].

K4A is currently in the process of publishing the 1st Video Journal of Open Education Abstracts Volume 1 with transLectures tools.

3.1.3 transLectures satellite workshop

A pre-conference workshop was organised between JSI-K4A and UPVLC to showcase project technologies to an interested audience. RWTH also attended the event.

Special effort was taken to prepare talks which outlined the key areas of research of transLectures in a way that was accessible to a non-specialist audience, spelling out exactly what the latest scientific results mean in practice for the end user.

First, transLectures project coordinator, Alfons Juan (UPVLC), gave an overview of the project, its motivations and the contexts in which its results will be applied. Next, Muhammed Ali Tahir (RWTH) gave his talk, “Domain adaptation for subtitles you can take seriously”, in which he discussed the ways in which we are adapting transcription and translation models for specific use in the video lecture domain. Then, Matjas Rihtar (JSI) spoke about the process of integrating project technologies into VideoLectures.NET and Carlos Turrò (UPVLC) spoke of poliMedia’s experience with transLectures subtitles, including the UPVLC’s transLectures Player, its “Try our Tools” service, and its experience with internal and external evaluations.

The highlight for us was the talk by our guest speaker, Stuart Phillipson, media technologies coordinator at the University of Manchester, who talked about what has been achieved at this university; namely, the largest working deployment of Matterhorn technologies in Europe. Phillipson expressed an interest in future collaboration. We were also pleased at the turnout by consortium members, with speakers from JSI-K4A, RWTH and UPVLC.

Following the event, the UPVLC received several new registrations for its transLectures-TOT, as well as new leads for future exploitation and external case studies. In particular, contact was made with the rector of the African Virtual University, Dr. Bakary Diallo, and the executive director of the OEC, Mary Lou Forward. The UPVLC have since followed-up on these contacts, announcing the launch of ToT, and Bakary has confirmed forwarding our press release onto the relevant teams.

See Appendix A.3 for the workshop agenda and the transLectures flyer for OCWC2014.
3.1.4 UNESCO Chair Workshop

The first UNESCO Chair Workshop on Open Technologies for Open Educational Resources and Open Learning was held in UNESCO Headquarters, in Paris, France.

Co-organised by JSI-K4A, this workshop comes on the back of the success of transLectures case study, VideoLectures.NET, which was recognised by the UN and UNESCO as one of the most outstanding examples of creative and innovative e-Content in the world in the last decade [17] [2]. It preceded the signature of the agreement between UNESCO and JSI to formally establish a UNESCO Chair on Open Technologies for OER and Open Learning at JSI.

The aims of the workshop were to share and disseminate knowledge, information and experiences in state-of-the-art internet-based technologies and services that can be used to organize, present, browse, analyse, remix and reuse OERs; to promote cooperative and collaborative activities to support the core themes of the Chair; and, by doing so, seek out synergies for Horizon 2020 work programme proposals.

See the workshop [programme] plus the news item on the [UNESCO website] [23].

3.1.5 Pre-conference workshop at Online Educa Berlin

Held in Berlin, Germany, Online Educa Berlin is a major event in the global field of e-learning and technology. JSI-K4A held a half-day workshop entitled “Artificial Intelligence Methods for Online-based Education” at this 19th edition of the conference. transLectures was presented as one of the innovative solutions currently being developed; as were MediaMixer, and AI application Curious Cat. The aim of the workshop was to discuss how these technologies could be incorporated into higher education to the benefit of students and educators alike. Joining JSI-K4A in the panel was the director of iVersity, Hannes Klopper, who expressed interest in the project.

See our [full write-up] on the project website.

3.2 Publications

Among the targets set for Y3, the main achievement in publications for this reporting period was the submission and acceptance of two joint publications, the first by RWTH, UPVLC and Xerox (No. 5) and the second by UPVLC and JSI (No. 9). This reporting period also saw the project’s first journal publication (No. 3). What follows is a list of all papers published between M24-M36:


2. *Assessing Quick Update Methods of Statistical Translation Models.*

3. *Evaluating intelligent interfaces for post-editing automatic transcriptions of online video lectures.*
4. **Evaluación del proceso de revisión de transcripciones automáticas para vídeos Polimedia.**  

5. **What can VideoLectures.NET and transLectures do for Opening Higher Education to the Multicultural World**  

6. **Comparison of Data Selection Techniques for the Translation of Video Lectures.**  

7. **Translation Modelling with Bidirectional Recurrent Neural Networks.**  

8. **Language model adaptation for lecture transcription by document retrieval**  

9. **The transLectures-UPV toolkit**  


11. **Statistical text-to-speech synthesis of Spanish subtitles**  

12. **Incrementally updating the SMT reordering model**  

A patent is being filed to protect the research presented in this paper, with the same title and author as the paper itself.

A full list of all publications since project start can be found in Appendix A.2 as well as in the new [Publications] section on the project website.

It is also worth mentioning that the transLectures-UPV toolkit has been used for other tasks by UPVLC researchers, specifically Arabic handwritten text recognition, leading to publications in international competitions and journals.
3.3 Open source tools

The development, release and update of a set of open source transLectures tools was a key line of work at the UPVLC for this reporting period. Here we give a brief description of each, plus the latest download figures.

- transLectures-UPV Toolkit (TLK)
  A full description of TLK can be found in Deliverables D7.5 [27], as well as on the project website, where it can also be downloaded. Briefly, it contains tools for building, training and applying acoustic models that can be used, by individuals or groups with a background in basic programming, to generate transcripts for video lectures. It is available for download on our website under an open source licence (specifically, Apache 2.0 [5]), along with full documentation, in English. As a minimum, basic command-line abilities are required to use these tools.

  To disseminate this toolkit, we made use of specialist mailing lists, the project website and social media accounts, and LinkedIn groups, with the latter proving particularly fruitful. We also promote it at events. In fact, two downloads were the direct consequence of our presentations at the UCL’s Digital Pedagogies event, attended by JSI-K4A, and the Science Dissemination event in Trieste, attended by UPVLC, both in Y2. New users are also known to have been recruited at IOE2013 and OCWC2014.

  Figure 1 gives an indication of TLK uptake (including only direct source downloads, as apt-get downloads requested from Debian-based Linux systems are not recorded). As an additional reference figure, we have recorded 723 unique visitors to the TLK website (excluding visits from UPVLC). Even with the launch of ToT in M29, there has been continued interest in TLK. The majority of downloaders are researchers in the field of NLP, and media staff at universities concerned with issues of accessibility.

Table 1: TLK releases and recorded direct source downloads

<table>
<thead>
<tr>
<th>M19</th>
<th>M26</th>
<th>M28</th>
<th>M33</th>
<th>M36</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLK release</td>
<td>1.0.0</td>
<td>1.1.0</td>
<td>1.2.0</td>
<td>1.3.0</td>
</tr>
<tr>
<td>Downloads</td>
<td>30</td>
<td>22</td>
<td>46</td>
<td>35</td>
</tr>
</tbody>
</table>

- transLectures Platform (TLP)
  A full description of TLP can be found in the User Guide and Documentation [22] available on the project website, where it can also be downloaded. Briefly, it is a self-contained piece of software developed at the UPVLC which includes everything you need in order to integrate transLectures transcription and translation technologies into a video lecture repository. Its main components are the transLectures Database, Web Service, Ingest Service and Player. Available under the same Apache Licence as TLK, TLP was released in M33. Perhaps most importantly up until now, TLP was installed remotely at UC3M, the UPVLC’s first external deployment site (see Section 4.2).

- transLectures Matterhorn (TLM)
  A full description of our plugin for Opencast Matterhorn can be found in D5.4 transLectures Deployment [8]. Briefly, it enables seamless integration of TLP and TLK into the Matterhorn Platform, allowing Matterhorn user institutions to ingest their videos for automatic transcription and translation. This is important because, by making our technologies compatible with Matterhorn, we vastly broaden the immediate user base of our open source software.
A similar plugin has also been developed for the Universidad de Vigo’s platform for the automatic online publication of audiovisual content generated at universities and institutions, PuMuKIT [4]. See our Section 4.2 on deployment for more details.

- **transLectures-Try our Tools (TOT)**

A full description of how transLectures-TOT works can be found in Deliverable 7.5 [27], plus the FAQ-type blog entry [16] on the project website. Briefly, it is an online service developed by the UPVLC where registered users can upload their videos and download transLectures subtitles in English, Spanish, Catalan, Italian, Portuguese and Dutch (these latter thanks to EMMA and local research efforts). It is free to register, and up to five videos may be uploaded per user. The aforementioned deliverable also gives an idea as to how this service might be adapted and expanded to form the basis of future exploitation plans.

Created as a tool for dissemination, it is proving invaluable as a means of showing would-be users exactly what kind of quality they can expect from our tools. For instance, Belgian company Speechware uploaded a number of test dictations, each time reporting positively the absolute and comparative quality against their current system (Nuance-based). The same happened with UC3M and UNED, who we invited to upload videos during discussions about their taking up transLectures tools (see Section 4.2 for more details).

Following OCWC2014 we sent a press release to new contacts at OER (formerly OCWC) and the African Virtual University, who forwarded it via internal mailing lists (See A.3). Already, a number of prominent figures from the education and accessibility sector have registered to ‘try our tools’. UPVLC is currently making sure to engage with each such registered user on a one-to-one basis to find out their needs and expectations. Institutional users so far include University Carlos III (UC3M), who are using transLectures-UPV technology to transcribe and translate their MOOC courses, and the Spanish National Open University (UNED). Additionally, the UPVLC are in talks with International Centre for Theoretical Physics (ICTP) (Italy), whose event “Science Dissemination and Online Certification for All” the UPVLC attended in M24, though nothing has been agreed on this front as yet.

So far, TOT has been used by 22 organisations from both the private and public sectors. Four of these correspond to universities from the EMMA project consortium: Universidade Aberta (Portugal), University of Naples Federico II (Italy), Open Universiteit Nederland (Netherlands), and University of Leicester (UK). Besides UC3M, UNED, ICTP and Speechware, all mentioned in more detail in other sections of this deliverable, TOT has also been trialled by the US Distance Learning Association and the Heidelberger Institut fur Theoretische Studien. It has also reached stakeholders in Asia. In total 73 hours of video have been transcribed and translated, in six different languages and over 10 different countries. See Table 2 for a summary.

### 3.4 JSI-K4A dissemination campaign

JSI-K4A have led a strong joint dissemination campaign together with the X-Like and MediaMixer projects, which began with their first event, Co-Creation of Emerging Trends in Academia [19], in M13. In Y2 their VideoLectures.NET was recognised at the World Summit on Information Society 10 Year Review Meeting (WSIS+10) as the best online educational product of the decade [17]. They also met with the Director-General of UNESCO and attended the Neth-ER seminar Reflections on “Opening Up Education” (reported in D7.2 [26]). In Y3 this was followed up with their Internet of Education conference (IoE2013) [6], and OCWC2014 [7], both discussed in depth in Section 3.1.
Table 2: Summary of main users of the transLectures-TOT public trial service.

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Country</th>
<th>No. of videos</th>
<th>Length (h)</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univ. Carlos III de Madrid</td>
<td>Spain</td>
<td>203</td>
<td>32</td>
<td>ES</td>
</tr>
<tr>
<td>Univ. Nacional de Educación a Dist. (UNED)</td>
<td>Spain</td>
<td>21</td>
<td>15</td>
<td>ES</td>
</tr>
<tr>
<td>Delft University of Technology</td>
<td>Netherlands</td>
<td>5</td>
<td>1.2</td>
<td>EN</td>
</tr>
<tr>
<td>Supercomputing Centre of Galicia (CESGA)</td>
<td>Spain</td>
<td>5</td>
<td>0.2</td>
<td>ES</td>
</tr>
<tr>
<td>Intl. Centre for Theoretical Physics (ICTP)</td>
<td>Italy</td>
<td>3</td>
<td>5</td>
<td>EN</td>
</tr>
<tr>
<td>SpeechWare</td>
<td>Belgium</td>
<td>3</td>
<td>3</td>
<td>ES</td>
</tr>
<tr>
<td>RWTH Aachen</td>
<td>Germany</td>
<td>2</td>
<td>0.1</td>
<td>EN</td>
</tr>
<tr>
<td>Centre for Dev. of Adv. Computing (C-DAC)</td>
<td>India</td>
<td>1</td>
<td>1</td>
<td>EN</td>
</tr>
<tr>
<td>UMH Occupational Observatory</td>
<td>Spain</td>
<td>1</td>
<td>1</td>
<td>ES</td>
</tr>
<tr>
<td>WBAC Technological College</td>
<td>Thailand</td>
<td>1</td>
<td>0.2</td>
<td>EN</td>
</tr>
<tr>
<td>VESIT Inst. of Technology Mumbai</td>
<td>India</td>
<td>1</td>
<td>0.2</td>
<td>EN</td>
</tr>
<tr>
<td>Paradise Valley Community College, Arizona</td>
<td>USA</td>
<td>1</td>
<td>0.1</td>
<td>EN</td>
</tr>
<tr>
<td>Heidelberger Institut für Theoretische Studien</td>
<td>Germany</td>
<td>1</td>
<td>0.1</td>
<td>EN</td>
</tr>
<tr>
<td>University of Michigan</td>
<td>USA</td>
<td>1</td>
<td>0.1</td>
<td>EN</td>
</tr>
<tr>
<td>Universidad de Cantabria</td>
<td>Spain</td>
<td>1</td>
<td>0.1</td>
<td>ES</td>
</tr>
<tr>
<td>US Distance Learning Association (USDLA)</td>
<td>USA</td>
<td>1</td>
<td>&lt;0.1</td>
<td>EN</td>
</tr>
<tr>
<td>UPV Linguistic Standardisation Service</td>
<td>Spain</td>
<td>1</td>
<td>&lt;0.1</td>
<td>CA</td>
</tr>
<tr>
<td>Aseryla project</td>
<td>Spain</td>
<td>1</td>
<td>&lt;0.1</td>
<td>ES</td>
</tr>
<tr>
<td>University of Naples Federico II</td>
<td>Italy</td>
<td>53</td>
<td>4</td>
<td>IT,EN</td>
</tr>
<tr>
<td>Universidade Aberta</td>
<td>Portugal</td>
<td>10</td>
<td>7</td>
<td>PT,EN</td>
</tr>
<tr>
<td>University of Leicester</td>
<td>UK</td>
<td>9</td>
<td>0.5</td>
<td>EN,ES,CA</td>
</tr>
<tr>
<td>Open Universiteit Nederland</td>
<td>Netherlands</td>
<td>4</td>
<td>2</td>
<td>NL,EN</td>
</tr>
<tr>
<td>Totals:</td>
<td></td>
<td>22</td>
<td>330</td>
<td>73, 6</td>
</tr>
</tbody>
</table>
In this section we report on the other initiatives and achievements of JSI-K4A in project dissemination.

### 3.4.1 Opening Up Slovenia

JSI-K4A’s recent “Opening Up Slovenia” initiative heads up activities aimed at policy-makers. Inspired by their involvement in transLectures, it is a new initiative in which Slovenia becomes one of the first European member states to attempt to create a nationwide research environment for open education. At the heart of this initiative is a platform for the development and uptake of intelligent tools, OERs and courseware. All of Slovenia’s universities are involved, as well as some compulsory and vocational education institutions, and partners from research and industry. Launched on 25 April 2014 at the OCWC global conference, the idea is to create an open education system in parallel to formal education.

At its launch, Commissioner Vassiliou, from the European Commission for Education, Culture, Multilingualism and Youth, commented:

“I am delighted that Slovenia has adopted a national strategy to promote open educational resources. This complements the Commission’s own “Opening up Education” initiative. Our shared objective is not to replace traditional textbooks or face-to-face teaching, but to make the most of opportunities created by digitisation and new technologies. Europe cannot afford to lag behind its international rivals if we want to ensure our young people have the skills needed in the modern world. I hope that this Slovenian initiative is successful and inspires similar schemes in other member states”.

### 3.4.2 The UNESCO Chairs Programme

Launched in 1992, the UNESCO Chairs Programme, involving over 812 institutions in 128 countries and 600 chair holders, promotes international inter-university cooperation and networking to enhance institutional capacities through knowledge sharing and collaborative work.

On 30 September 2014, UNESCO Director-General Irina Bokova and the director of JSI, Prof. Jadran Lenarcic, signed an agreement establishing a UNESCO Chair on Open Technologies for Open Educational Resources and Open Learning at JSI. With this prestigious title, transLectures partners JSI has proven that its vision of free and open access to academic knowledge in the form of open online videos and smart technologies is an effective approach for addressing the global digital divide in education knowledge and transfer.

The Chair will operate with an initial set of assets in the form of EU FP7 research projects such as transLectures, MediaMixer, xLiMe, X-Like and VideoLectures.NET. In its work, it will continue to disseminate transLectures tools in synergy with the Opening up Slovenia initiative (see [3.4.1]) and K4A, promoting an integrated system for research, training, information and documentation in online learning, OERs and open learning. It will facilitate collaboration between high-level, internationally-recognised researchers and teaching staff at JSI and other institutions around the world. Welcoming the establishment of this Chair, UNESCO Director-General commented:

“We are very pleased to be expanding our cooperation in this forward-looking area that offers the opportunity to increase the quality of learning. This is essential in building knowledge-based societies.”
3.4.3 VideoLectures.NET

Discussions have taken place between JSI and the UPVLC regarding this latter university using its technology to generate transcriptions and translations for the VideoLectures.NET repository; specifically, for the videos uploaded since the project began and which are therefore not included in the current transLectures repository. Based on the snapshot generated by JSI on 1st May 2014, there are a total of 3479 new lectures (3649 videos, 2244 hours) to be subtitled.

As of 21 October 2014, the UPVLC had transcribed and translated some 696 lectures (478 hours) published between 8 October 2013 and 29 April 2014 using its local infrastructure. It will continue this task until project end. EML has also joined this final effort to transcribe and translate the most recent VideoLectures.NET content.

3.5 Dissemination Timeline M24-M36

1. 11/12.11.13: Organised the first Internet of Education 2013 conference - Ljubljana (Slovenia) (JSI-K4A, UPVLC)
   Organised by JSI-K4A under UNESCO patronage, the goal of this event was to provide a forum for researchers and policy-makers to explore means of improving the effectiveness of a MOOC-based education, as well as the implications of new edutech trends for the classical university paradigm.
   Gonçal Garcés (UPVLC) presented transLectures.
   See Section 3.1.1 for more details.

2. 5/6.12.13: Attended the International Workshop on Spoken Language Translation (IWSLT2013) - Heidelberg (Germany) (RWTH, XEROX)
   transLectures researchers from both RWTH and Xerox attended this event, presenting their papers (Nos. 1, 2 in Section 3.2).

3. 4.12.13: Organised pre-conference workshop at Online Educa Berlin - Berlin (Germany) (JSI-K4A)
   See Section 3.1.5 for more details.

4. 11.12.13: Attended REC:all workshop - Leuven (Belgium) (UPVLC)
   Carlos Turro (UPVLC) presented transLectures at this workshop, entitled “Beyond the pilot stage: large-scale implementation of lecture capture in European Higher Education”.

5. 1.14: Yota Georgakopoulou interviewed for JoSTrans, including a section on transLectures, and on machine translation in subtitling in general.
   See the full video here.

6. 1.14: Yota Georgakopoulou co-authored article for Multilingual magazine, in which she talks broadly about the use of automatic technologies in the subtitling industry and the need for industry, academia and language professionals to collaborate. She cites the transLectures project by way of example.
   The article introduction can be found here, though the full article can only be viewed by subscribers.

7. 7.1.14: Released the transLectures-UPV toolkit (TLK) v.1.1.0. (UPVLC)
   TLK version 1.1.0 for Linux and Mac introduced new high-level scripts to make preprocessing, training and recognition tasks simpler, with a new tutorial [24]. As a minimum, basic command-line abilities are required to use these tools.
   See our blog post.
8. 6/8.1.14: JSI-K4A held the 1st Winter School on Multimedia Processing and Applications - Dublin (Ireland) (MediaMixer)
   See the official [website](#) for more details.

9. 13.3.14: Try our Tools (ToT) service made available to the general public (UPVLC)
   In March the UPVLC opened up the transLectures-ToT service, allowing registered users to upload their videos for transcription and translation. In M31, Italian and Dutch were added to the portfolio of languages (En, Es, Ca).
   Please consult our [FAQ-type](#) blog entry for more details; also, Section 3.3 of this deliverable.

10. 11.3.14: Multiple presentations by K4A at the European Commission: to Ana Carla Pereira, Head of Unit, and Ricardo Ferreira, Policy Officer in charge of Open Educational Resources, European Commission DG Education and Culture, Unit A3: Skills and Qualifications, and to Zoran Stančić, Deputy Director General, DG CONNECT - Brussels (Belgium) (JSI-K4A)

11. 11.3.14: Spring workshop on Mining and Learning (SMiLe) workshop - Ostende (Belgium) (K4A)

12. 3.4.14: K4A presentation at the Ministry of Education, Science and Sport, International Cooperation and European Affairs Service, EU Coordination Unit - Brussels (Belgium) (K4A)

13. 4.4.14: New landing page for the website, with a more current and mobile-friendly design.
   In M30 the UPVLC designed a new landing page for the website, with a more current and mobile-friendly design (see Fig. 1). Above all the idea was to give the Try our Tools (ToT) service greater prominence, emphasising the end results or “products” of the transLectures project. Other minor structural changes that have been made over the past few months include a new tabs configuration, making publications, reports, demos and tools more immediately visible.

![Figure 1: transLectures.eu new look landing page.](#)
14. 22.04.14: Organised and attended OCWC 2014 **transLectures** workshop (JSI-K4A, RWTH, UPVLC)

   See Section [3.1.3](#) for further details.

   Also available for consultation are the [conference website](http://example.com) and our [blog entry](http://example.com).

15. 23/25.04.14: Organised the **OCWC global conference 2014** - Ljubljana (Slovenia) (JSI-K4A)

   As silver sponsor for the event, project coordinator Alfons Juan (UPVLC) and Carlos Turrò (also UPVLC) attended to present the project, a [video](http://example.com) of which can be viewed on VideoLectures.NET. The idea pursued here was to disseminate **transLectures** more as a product than a project.

16. 6.14: Formal agreement drawn up between UPVLC and Universidad Carlos III de Madrid for their uptake of **transLectures**-UPV technologies. See Section [4.2](#) for more details.

17. 18/19.6.14: Attended ClipFlair - Barcelona (Spain) (Deluxe)

   Maria Gialama (Deluxe) presented **transLectures** at this conference on the topic of using audiovisual material in the classroom for language learning purposes, a potential use of the **transLectures** player that this partner have endeavoured to promote. The event was mainly attended by audiovisual translation academics, and university academics who have implemented [ClipFlair Studio](http://example.com) (recommend watching the demo videos) for language learning. Gialama’s presentation attracted a lot of interest and questions regarding **transLectures** tools, what languages are being developed and the data used. UAB showed particular interest, and were sent a follow-up email with project details.

18. 27.6.14: Attended International Conference on Computer Systems and Technologies (CompSysTech’14) - Ruse (Bulgaria) (JSI-K4A)


20. 15.7.14: Released the **transLectures** Platform “TLP” version 1.0.0. (UPVLC)

21. 29.7.14: Released the **transLectures**-UPV toolkit “TLK” version 1.3.0. (UPVLC)

22. 1.9.14: **transLectures** mention in the EMMA newsletter, sent to the EMMA mailing list. See Appendix [A.3](#) (UPVLC)

23. 25.9.14: UPV (i.e. not just **transLectures**) met with Johannes Heinlein, Vice-President of Strategic Partnerships at edX. (UPVLC)

   A team of representatives presented project technologies and current exploitations to Heinlein, who was very interested in the idea of a **transLectures** plugin for OPENedX along the same lines as the UPVLC is currently developing for UNED (see Section [4.2](#)). A similar plugin for EdX itself is not beyond the realms of possibility.

24. 28.9.14: Visit from Helene Gram, pedagogical designer at the University of Stavanger (UPVLC)

   Learning of **transLectures** through EMMA, Gram has expressed interest in a Norwegian service.

25. 30.9.14: Organised UNESCO Chair Workshop on Open Technologies for Open Educational Resources and Open Learning - UNESCO Headquarters, Paris (France) (JSI-K4A)

   See Section [3.4.2](#) for more details.
transLectures was disseminated to the user and prosumer community at the UPVLC via the poliFormaT homepage, as shown in Figure 3. Besides announcing that subtitles are available for the videos published on poliMedia, it attributed them to the transLectures project, and explained the process by which both lecturers and any poliMedia user can edit these subtitles, again using transLectures technology. It was featured for two weeks at the beginning of term.

Furthermore, having loaded a video to begin the post-editing process, at the top of the screen we have incorporated the transLectures logo (see Fig. 2). The aim is to forge the beginnings of a brand awareness, and the association of online subtitles and hassle-free subtitle post-editing with the transLectures project.

Figure 2: poliFormaT portal to the poliMedia repository, featuring the transLectures logo.

Figure 3: Example poliFormaT video in editing mode.
27. 10.14: transLectures plugin for PuMuKIT developed for use at Universidad Nacional de Educación a Distancia (UPVLC). See Section 4.2 for more details.

28. 21.10.14: Released the transLectures-UPV Toolkit (TLK) v.1.3.1. (UPVLC)

29. 22-26.10.14: Attended Eleventh Biennial Conference of the Association for Machine Translation in the Americas (AMTA 2014) - Vancouver, Canada (Xerox)
   S. Mirkin will present the joint transLectures paper “Comparison of Data Selection Techniques for the Translation of Video Lectures”.

30. 25/29.10.14: Attended conference on Empirical Methods in Natural Language Processing - Doha, Qatar (RWTH)

31. 29.10.14: Released transLectures-UPV Platform (TLP) v.1.0.1 (UPVLC)

32. 30.10.14: Released transLectures Matterhorn (TLM) plugin (UPVLC)

3.5.1 Confirmed events beyond project end

33. 19-21.11.14: “VIII Jornadas en Tecnología del Habla” and “IV Iberian SLTech Workshop” (IberSpeech 2014) - Las Palmas de Gran Canaria (Spain) (UPVLC)
   Four papers relating to work carried out as part of transLectures will be presented here, including a joint UPVLC-JSI publication.

   Xerox will be attending to present their paper “Incrementally updating the SMT reordering model”.
   A patent is currently being filed to protect the research presented in this paper, with the same title and author as the paper itself.

4 Business plan and exploitation

In addition to the transLectures-UPV tools described in Section 3.3, conceived as tools for dissemination to demonstrate exactly what transLectures can do and establish a user community, transLectures technology is already being used at two other Spanish universities. In this section we will first recap the business plan as outlined in D7.5, a major area of work during this reporting period, and then give an account of all current exploitations of project technologies.

4.1 Summary of D7.5

Building on the draft delivered in M24, the Final Business Plan (M32) was a key area of work in WP7 this reporting period.

For it, we considered a range of possible commercial applications of project technologies, including online transcription services, dictation software, closed-captioning for TV and as a language learning tool. In the end, though, the business case was very well-defined in the Description of Work and, accordingly, transLectures tools have been developed with the video lecture use case in mind: specialised models, and interaction modes tried, tested and adapted to meet the needs and preferences of lecturers and students.

We also took a closer look at the main competitors, to see what they were bringing to the market and how. Based on this research, we determined that there is indeed a place for
transLectures technologies in the education sector, especially given the kind of results we are achieving, at least in the case of data-rich languages (Es, En, Fr, De). The need for the kind of service, or data, that transLectures can provide has been confirmed time and again via our various dissemination channels. The ultimate confirmation, of course, is the fact that our tools are already being commercially exploited (see Section 4.2).

By way of overview, the transLectures business plan is in fact a set of partner-specific plans for commercialisation and/or further research, though there is of course scope for continued collaboration.

For its part, the UPVLC intends to commercialise and otherwise distribute its tools in the form of open source tools releases, and a combination of a transLectures-UPV web service and technology transfers. So far this has been a case of adapting what it has to offer and the way it offers it to meet the needs of the receiving institution. Through its work on EMMA and other local research efforts, it hopes to extend the current service portfolio to include new languages and new features (for instance, text-to-speech).

JSI-K4A have projects under way that will make use of the underlying technologies to produce derivative video management functions (cross-lingual search, video fragmentation and recommendation, etc.) that can be commercialised or otherwise used to give VideoLectures.NET a competitive advantage. Furthermore, with the UNESCO Chair on Open Technologies, they are also well-positioned within the open educational resources sector to not only have an impact at institutional level, but promote transLectures implementations. Their Opening Up Slovenia initiative also offers great potential for transLectures exploitation.

EML will incorporate transLectures technologies into its existing products and services portfolio. Namely, its EML Transcription Server, which can be easily integrated into platforms hosting video lectures, and professional services including language development, language model adaptation, integration support and training on all of the above. It will also offer access to the EML Transcription Server and language components for the educational market to transLectures case study sites for a twelve month introduction period (maximum of 100 hours per month per site) in exchange to access to audio and derived data to improve their commercial speech recognition models.

RWTH will be building on transLectures research into the future. Both its proprietary software toolkits, and its acoustic, language and translation models are available for free for academic research and under a licence agreement for commercial applications.

Xerox, similarly, will continue to build on transLectures research into the future. Additionally it has pledged to make XRCE-hosted APIs available for use by project partners until October 2015 for the Slovene-English language pair. After this one-year observation period, a decision will be made between commercially exploiting Xerox MT models within services offered by other transLectures partners, or integration into educational content platforms such as Matterhorn. Available since August 2014, so far JSI-K4A have expressed interest in this offer for VideoLectures.NET.

Deluxe Media Europe has expressed interest in collaborating as language services provider for post-editing services and professional consultancy on any potential exploitation of the transLectures project. They will also make available all the transcriptions and translations generated by the company as part of the project (DDS foreground) to all consortium partners for an unlimited time and for any purpose (commercial use included).

4.2 Current deployment

- Universidad Carlos III de Madrid (UC3M)
In an initial pre-agreement stage, UC3M uploaded several videos to transLectures-TOT to test the quality of the subtitles across varying video formats: one poliMedia-like format and the other a digital board format.

In June 2014, a formal technology transfer agreement was drawn up and signed between UC3M and UPVLC. After advising (UPVLC) and procurement (UC3M) of the technical requirements for installation, the UPVLC installed a virtual transLectures system (transLectures-TLP) at UC3M for the onsite transcription and translation of unlimited educational videos only (i.e. non-commercial use, no third parties). The agreement reached includes updates and maintenance on the part of the UPVLC (for a period of one year), and supervision (post-editing) of subtitles on the part of UC3M using the transLectures Player.

As of September 2014, word error rate (WER) for Spanish transcriptions was competitive for both formats, with a global WER of 14.2, while the English translations are typically taking an RTF of 13.6 to supervise (by non-native speakers).

The installation process was carried out remotely without major difficulties. We feel this method of deployment can be replicated at other, more distant locations.

- Universidad Nacional de Educación a Distancia (UNED)

The Spanish national open university have massive archives of video material that they are currently in the process of digitalising. They also want subtitles. The UPVLC have been in talks with them for some months to decide the best way to integrate transLectures into their setting to meet their needs.

In the end, a pay-as-you-go plan has been agreed, whereby UNED purchase a given number of hours of transcribed and translated video in advance, and then over a period of time submit the videos they want to be processed. According to UNED’s requirements, this has been achieved through the incorporation of a transLectures plugin into the PuMuKit platform they currently have deployed, with Universidad de Vigo as technology provider, that will allow UNED to send and receive transcriptions and translations from servers hosted at the UPVLC. Archive administrators will be able, at the click of a button, to send individual videos for transcription and translation. This way they can apply their own criteria when deciding which videos to get “translectured”. For instance, the most popular, the newest, if there are known deaf or hard-of-hearing students on a given course, etc., while they might avoid sending videos recorded in less than optimum conditions.

This is a new development as of M36 so feedback is limited to the videos they submitted using the transLectures-TOT service. The decisive factor in the quality of the subtitles was audio quality. Ultimately, the fact that their transLectures-PuMuKit plugin is up-and-running, and via their third party video hosting platform provider, speaks to their satisfaction. We hope to report further on progress at UNED at the final review meeting in December.

- European Multi MOOC Aggregator (EMMA)

Up-and-running since February 2013, the UPVLC is participating as sole technology provider in this pilot study, exploiting and building upon technologies developed as part of transLectures, including new languages. In October 2014 a first UPVLC Spanish MOOC was published on the aggregator, complete with supervised transLectures subtitles in Spanish, English and Catalan. A second will be added in the not-too-distant future.

4.3 Possible partnerships

Given the shared trajectory in transLectures, there is obviously considerable scope for partnership. As reported above in Section 4.1 both EML and Xerox have pledged to make services
available for a period of one year after project end, EML in exchange for data and Xerox as an observation period to decide between business models. As yet, no formal agreement has been reached in this regard. One partnership that is already up-and-running is that already outlined in Section 3.4.3 to get the later VideoLectures.NET videos transcribed and translated. Furthermore, future joint research efforts are not beyond the realms of possibility, funding-dependent.

5 Evaluation of dissemination activities

In this section we will evaluate the impact of dissemination activities and the project as a whole. Having already discussed the UPVLC’s open source tools and their uptake, and highlighted the key outcomes of JSI-K4A’s far-reaching joint dissemination campaign with MediaMixer, here we will analyse other evidence of the impact and scope of transLectures.

Not always quantifiable, we will begin with the more tangible, objective measures, such as references in publications and web analytics, before moving onto more subjective indicators such as feedback.

5.1 References to transLectures in other publications

References to transLectures in publications are a good indicator of the impact the project is having within academia. It does seem that the message of what transLectures makes possible is being heard and taken as something of a benchmark.

- **The AMARA Corpus: Building Resources for Translating the Web’s Educational Content**
  Authors: F. Guzman, H. Sajjad, S. Vogel, A. Abdelali, at the Qatar Computing Research Institute (Doha, Qatar). In: Proceedings of the International Workshop on Spoken Language Translation, IWSLT 2013.

  “Lecture Translation has become an active field of research in the wider area of Speech Translation (Fügen et al., 2006, Fügen at al., 2007). This is demonstrated by large scale projects like the EU-funded transLectures (Silvestre-Cerdà et al., 2012 [9]) and by evaluation campaigns like the one organized as part of the International Workshop of Spoken Language Translation (IWSLT) (Paul et al., 2010).”

- **The AMARA Corpus: Building Parallel Language Resources for the Educational Domain**

  Quote as above.

- **Amara: A Sustainable, Global Solution for Accessibility, Powered by Communities of Volunteers**
  Authors: D. Jansen, A. Alcala, F. Guzman.

  Quote unavailable (subscription needed).
Leveraging video annotations in video-based e-learning

Authors: S. Aubert, Y. Prié, C. Canellas, at the University of Nantes. In: Proc. of the 7th International Conference on Computer Supported Education, CSEDU 2014.

“Most video-based e-learning systems use only plain videos, sometimes fragmented into small independent videos, providing only basic features. In order to make these videos more accessible, e-learning platforms should commonly provide features such as transcription or chaptering. Some projects such as TransLectures aim at providing automatic or semi-automatic transcription of video, so that users may use the transcriptions as entry points into the video, either for querying and finding specific fragments, or as a simple navigation means.”

Here, no specific TransLectures publication was cited, but it does include a link to our website.

5.2 Web analytics

There have been a number of changes to the website since project start, to both the structure and the content. The first change in January 2013 introduced a more contemporary design and new content written specifically for the website with non-scientists in mind. The next major change came in April 2014 with the incorporation of a new mobile-friendly landing page, which drives home the idea of TransLectures as a product, not just a project.

Based on figures extracted from Google Analytics, we see that total traffic during the Y3 reporting period has increased 66.43% on Y2. This, after a 362% increase from Y1 to Y2. In absolute terms, we are talking about 3445 visits in Y3 versus 2070 in Y2.

Figure 5.2 shows the evolution of unique visitors to the website over the project lifetime, with Valencia-based visitors removed. This is because lecturers at the UPVLC who took part in the external evaluations at poliMedia accessed their videos via the website and should therefore be discounted as unique visitors. Broadly-speaking, we can say that the number of visitors to the site has increased year on year, particularly since the first website redesign in January 2013. Other spikes correspond to the dissemination of tools releases and the OCWC2014 global conference.

Top referrals in Y3 were as expected: translectures.eu (direct), polimedia.upv.es, Facebook, LinkedIn and Twitter, with other referrals including the VideoLectures.NET blog, the OCWC2014 conference website, and the OpenCast, Pascal and K4A websites. This indicates that people are finding out about TransLectures as a direct result of, for instance, our role in the OCWC2014 conference and our case studies.

Our outreach potential on Facebook is rather low, with a total of 115 likes (up from 90 in August 2013), though this likely reflects the fact that our main stakeholders are either computer scientists, large academic institutions (both as collaborators and potential users) or companies that generate or process multimedia files on a massive scale, and as such there are perhaps too many people or management levels separating TransLectures from those doing the clicking (community managers, interns, etc.).

Our 86 followers on Twitter, meanwhile, nicely illustrate our stakeholder landscape, with developers; researchers in education, lecture capture, computer science, and language tech; subtitling and technology (video, streaming) companies; and related EU initiatives making up the bulk of our followers.
5.3 Subjective feedback

- M21: Feedback from first external users - poliMedia, UPVLC

Prior to project start, poliMedia had been adopted by other Spanish and Latin American universities. Taking advantage of this, the UPVLC contacted some of the closer-to-home universities to explore the possibility of integrating transLectures technologies into their video repositories, or at least, at this early stage, trying our subtitles on a few of their videos. We received the following subjective feedback:

“This is a truly marvellous tool, and one that will be very successful I’m sure. We’re sending you more videos [for transcription] now!”

Ana Castellano, Head of Multimedia Projects at the Faculty of Medicine, UAB, for whom we transcribed a video recorded under studio conditions. Estimated WER: 5%.

“I’m sure [as you say] the transcription can be improved, but the levels of accuracy achieved for this video already make it simply wonderful for use in searches... Top job!”

Vicente Goyanes, Head of IT-Media Technical Assistance, UVIGO, for whom we transcribed a Matterhorn-style video (live classroom recording). Estimated WER: 30%.

We then transcribed one of UVIGO’s “learning pills”, more in line with poliMedia recording conditions, prompting the following feedback:

“I really am very impressed by the transcription of the video with good audio I sent you. It’s terrific!”

(WER not calculated, but likely in line with the first UAB video).

- M12, M24: Feedback from Meena Hwang, OCWC Director of Outreach and transLectures advisory board member
From day one Hwang has been very enthusiastic about the project, but when, in the year 2 project meeting, she announced that she was “very excited” about transLectures in so far as it is open source and so not going to die a death in the laboratory at project end, plus the basis for myriad derivative products, and said that the only question really was “just how far we wanted to take it”.

- M35: Feedback from UC3M (see Section 4.2)

Once the transLectures-TLP was up-and-running at this university, we received the following email from Francisco Cruz, Head of Multimedia and Teaching Innovation:

“We’d be interested in having a direct way to confirm the transcription for a whole video without having to check it line by line. Considering how well the system is working, sometimes we’d like to be able to confirm a whole transcription.”

He is referring to the process of intelligent post-editing whereby the prosumer (in this case, lecturer) is directed to possible problem areas in the transcription or translation, in order to supervise (correct or confirm) the automatic subtitle. Here they have asked for a “validate all” button in the transcription to simply confirm the entire transcription without checking it.

6 Conclusions

transLectures was conceived at a time of rapid expansion in the use of online video material in higher education, both within and in parallel to existing education paradigms. Its potential for allowing universal access to quality education was clear. There was only one problem: usable subtitles were costly and therefore something of a rarity, which put a cap on this potential.

transLectures’ solution was simple: intelligent automation. To avoid the pitfalls of generic automated subtitling services, the models used would be custom adapted for specific use on higher education video lectures. They would also be further adapted over time and use to a particular speaker or a particular subject area, as well as include a human-machine interaction element to further refine output to the satisfaction of lecturers and students alike in a way that was cost-effective and sustainable.

Three years on, more and more universities are incorporating video content into their curricula in one form or another. Massive open online courses (MOOCs) are fast becoming mainstream, and second-generation products are already beginning to emerge to address issues such as student retention rates and accessibility. And say what you will about Youtube’s subtitles, it has at least served to whet the appetite of the consumer for this kind of service.

The timing and success of this project meant that dissemination had a relatively easy job. However, we have not rested on our proverbial laurels. The work carried out in WP7 during this reporting period and throughout project lifetime has been strong. Activities have been proactive and varied. Through event attendance and organisation, we have reached out to the education sector, in particular the open education and edutech communities. We have appealed to policy-makers at EU and international institutional level. We have not only disseminated transLectures as a means of generating multilingual subtitles, but as the basic building block for myriad derivative features. We have also made forays into the accessibility and language learning sectors. Above all, we have zoomed in on the end user communities in the form of various evaluation phases and the release of a comprehensive set of open source tools, not to mention our online trial service.

Between the drive of JSI-K4A and the practical groundwork of the UPVLC, we are certain that transLectures will not end on 31 October 2014. Indeed, the know-how and experience
gained in this project will be carried through and built on in future research projects by all project partners. Formal project funding comes to an end, but alternative sources of income to sustain further research are already being exploited at the UPVLC, not to mention that the technology is also set to reach the market via EML’s existing channels, and Xerox will spend the next year exploring business models.

Lecturers want it, online video repositories need it in order to remain competitive, and students around the globe stand to benefit from it. This, combined with the results we have achieved, make us optimistic about the future of transLectures.
References


A Appendix

A.1 Events timeline M1-M36

- 16-18.4.12: Attended Cambridge 2012 - Cambridge (UK) (UPVLC)
- 23/25.5.12: Attended LREC 2012 - Istanbul (Turkey) (UPVLC, Deluxe)
- 28/30.5.12: Attended EAMT 2012 - Trento (Italy) (UPVLC)
- 20/21.6.12: Attended META-FORUM 2012 - Brussels (Belgium) (Deluxe)
- 07.11.12: Organised Co-creation of Emerging Trends workshop - Ljubljana (Slovenia) (JSI-K4A)
- 13.11.12: Attended REC:all pre-conference workshop - Leuven (Belgium) (UPVLC)
- 14/15.11.12: Attended Media and Learning 2012 conference - Brussels (Belgium) (UPVLC)
- 21/23.11.12: Attended IberSpeech 2012 conference - Madrid (Spain) (UPVLC)
- 21/23.11.12: Attended Languages and the Media 2012 conference - Berlin (Germany) (JSI-K4A)
- 28/31.11.12: Attended Online Educa Berlin conference - Berlin (Germany) (UPVLC)
- 08.01.13: Delivered REC:all webinar (UPVLC)
- 01.13: Article in the January edition of the Media and Learning newsletter (UPVLC)
- 14.01.13: Filed for patent with US Ser. No.: 13/740,508 (Xerox)
- 07.02.13: Opened a transLectures channel [20] on VideoLectures.NET (JSI-K4A, UPVLC)
- 19/27.02.13: Press coverage of VideoLectures.NET and its upcoming WSIS+ award - Slovene (JSI-K4A)
- 25/27.02.13: VL.NET received award at World Summit on Information Society 10 Year Review Meeting (WSIS+10) - Paris (France) (JSI-K4A)
- 11.03.13: VideoLectures.NET in the Digital Agenda [2] (JSI-K4A)
- 12.03.13: Attended 4th International Symposium on Live Subtitling conference - Barcelona (Spain) (Deluxe)
- 20.03.13: Featured in the PASCAL2 brochure [1]
- 01.04.03: Press coverage of IASD challenge
- 11.04.13: Organised EUCOGIII/PASCAL2 IASD challenge - Palma de Mallorca (Spain) (UPVLC)
- 09.4.2013: Meeting with Irina Bokova, Director-General of UNESCO - Brdo pri Kranju (Slovenia) (JSI-K4A)
- 03.05.13: Released first version of the transLectures-UPV toolkit, TLK (UPVLC)
• 08/10.05.13: Attended OCWC Global 2013 conference - Bali (Indonesia) (JSI-K4A)
• 26/31.05.13: Attended ICASSP 2013 conference - Vancouver (Canada) (RTWH, UPVLC)
• 25.05.13: Produced a new demo video of the transLectures interactive player for ESWC2013 (UPVLC)
• 28.05.13: Attended ESWC2013 conference - Montpellier (France) (UPVLC)
• 13.06.13: Attended UCL Digital Pedagogies: E-Learning & Digital Humanities Unconference - London (UK) (JSI-K4A)
• 24/26.06.13: Attended Opencast 2013 Unconference - Osnabrück (Germany) (JSI-K4A, UPVLC)
• 06.13: Ljubljana confirmed as the venue for OCWC2014 (JSI-K4A)
• 02.07.13: Uploaded new videos to the transLectures channel on VideoLectures.NET. (UPVLC)
• 19.07.13: Matterhorn/transLectures “success story” published on Opencast (JSI-K4A, UPVLC)
• 07.13: First feedback from external users (UPVLC)
• 2/6.09.13: Attended the MT Summit 2013 - Nice (France) (UPVLC)
• 19/20.09.13: Attended META FORUM 2013 - Berlin (Germany) (EML)
• 26-27.09.13: Attended the MicroLearning Conference 7.0 - Krems/Götweig (Austria) (JSI-K4A)
• 30.09.13: Attended Science Dissemination and On-line Certification for All - Trieste (Italy) (UPVLC)
• 1.10.13: Attended Neth-ER seminar: Reflections on “Opening Up Education” - Brussels (Belgium) (JSI-K4A)
• 2.10.13: EADTU-EU Special Summit on Higher Education - Brussels (Belgium) (JSI-K4A)
• 13/16.10.13: Attended IEEE SMC 2013 - Manchester (UK) (UPVLC)
• 11/12.11.13: Organised the Internet of Education 2013 conference - Ljubljana (Slovenia) (JSI-K4A, UPVLC)
• 5/6.12.13: Attended IWSLT2013 - Heidelberg (Germany) (RWTH, XEROX)
• 4.12.13: Pre-conference workshop at Online Educa Berlin - Berlin (Germany) (JSI-K4A)
• 11.12.13: Attended REC:all workshop - Leuven (Belgium) (UPVLC)
• 1.14: Yota Georgakopoulou interviewed for JoSTrans (Deluxe)
• 1.14: Yota Georgakopoulou co-authored article on MT for Multilingual magazine (Deluxe)
• 7.1.14: Released the transLectures-UPV toolkit (TLK) v.1.1.0. (UPVLC)
• 6/8.1.14: JSI-K4A held the 1st Winter School on Multimedia Processing and Applications - Dublin (Ireland) (MediaMixer, JSI-K4A)
• 2.14: Resuscitated the transLectures LinkedIn account, plus opened a transLectures Slideshare account. (UPVLC)

• 13.3.14: Try our Tools (ToT) service made available to the general public. (UPVLC)

• 11.3.14 Multiple presentations by K4A at the European Commission - Brussels (Belgium) (JSI-K4A)

• 11.3.14: Spring workshop on Mining and Learning (SMiLe) workshop - Ostende (Belgium) (K4A)

• 3.4.14: K4A presentation at the Ministry of Education, Science and Sport, International Cooperation and European Affairs Service, EU Coordination Unit - Brussels (Belgium) (K4A)

• 4.4.14: New landing page for the website, with a more current and mobile-friendly design. (UPVLC)

• 22.04.14: OCWC 2014 transLectures workshop (JSI-K4A, RWTH, UPVLC)

• 23/25.04.14: transLectures at OCWC global conference 2014 (JSI-K4A, UPVLC)

• 6.14: Formal agreement drawn up between UPVLC and UC3M (UPVLC)

• 18/19.6.14: ClipFlair - Barcelona (Spain) (Deluxe)

• 27.6.14: Attended International Conference on Computer Systems and Technologies (Comp-SysTech’14) - Ruse (Bulgaria) (JSI-K4A)

• 7/9.7.14: Attended the 6th Annual International Conference on Education and New Learning Technologies, EDULEARN14 - Barcelona (Spain) (JSI-K4A)

• 15.7.14: Released the transLectures Platform “TLP” version 1.0.0. (UPVLC)

• 29.7.14: Released the transLectures-UPV toolkit “TLK” version 1.3.0. (UPVLC)

• 1.9.14: transLectures mention in the EMMA newsletter, sent to the EMMA mailing list. See Appendix A.3

• 25.9.14: UPVLC (not just transLectures) meeting with Johannes Heinlein, Vice-President of Strategic Partnerships at edX.

• 28.9.14: Visit from Helene Gram, pedagogical designer at the University of Stavanger (UPVLC)

• 30.9.14: Organised UNESCO Chair Workshop on Open Technologies for Open Educational Resources and Open Learning - UNESCO Headquarters, Paris (France) (JSI-K4A)

• 10.14: transLectures disseminated via poliFormaT, the platform via which teachers and students at the UPV access their poliMedia videos.

• 10.14: transLectures plugin for PuMuKIT developed for use at UNED (UPVLC)

• 21.10.14: Released the transLectures-UPV Toolkit (TLK) v.1.3.1. (UPVLC)

• 22-26.10.14: Eleventh Biennial Conference of the Association for Machine Translation in the Americas (AMTA) - Vancouver, Canada. (Xerox)

• 25/29.10.14: Conference on Empirical Methods in Natural Language Processing - Doha, Qatar (RTWH)

• 29.10.14: Released transLectures-UPV Platform (TLP) v. 1.0.1

• 30.10.14: Released transLectures Matterhorn (TLM) plugin (UPVLC)
A.2 Publications M1-M36

2012

- **Integrating a State-of-the-Art ASR System into the Opencast Matterhorn Platform.**

- **Character-Based Handwritten Text Recognition of Multilingual Document.**

- **transLectures.**

- **Explicit length modelling for statistical machine translation.**

- **Does more data always yield better translations?**

2013

- **Comparison of Feedforward and Recurrent Neural Network Language Models.**

- **Language model adaptation for video lectures transcription.**

- **A System Architecture to Support Cost-Effective Transcription and Translation of Large Video Lecture Repositories.**
  Authors: J.A. Silvestre-Cerdà, A. Pérez, M. Jiménez, C. Turró, A. Juan, & J. Civera. Accepted at IEEE SMC 2013. (UPVLC)

- **The RWTH Aachen Machine Translation Systems for IWSLT 2013.**
- **Assessing Quick Update Methods of Statistical Translation Models.**

- **Cross-Entropy vs. Squared Error Training: a Theoretical and Experimental Comparison.**

- **Development of the RWTH Transcription System for Slovenian.**

2014

- **Evaluating intelligent interfaces for post-editing automatic transcriptions of online video lectures.**

- **Evaluación del proceso de revisión de transcripciones automáticas para videos Polimedia.**

- **Comparison of Data Selection Techniques for the Translation of Video Lectures.**

- **Translation Modelling with Bidirectional Recurrent Neural Networks.**

- **Language model adaptation for lecture transcription by document retrieval**

- **The transLectures-UPV toolkit**

- **Using Automatic Speech Transcriptions in Lecture Recommendation Systems**
• Statistical text-to-speech synthesis of Spanish subtitles

• What can VideoLectures.NET and transLectures do for Opening Higher Education to the Multicultural World

Incrementally updating the SMT reordering model

A.3 Dissemination material
transLectures is a dynamic European project that is set to bring quality automated subtitles to the education sector.

- Make your lectures accessible to the deaf and hard-of-hearing with our verbatim subtitles
- Watch your videos in noisy and noise-restricted environments
- Translate your subtitles to open up your videos to a global audience
- Watch the original and translated subtitles simultaneously to improve language learning

PLUS! Unlimited access to tL player, our subtitle editing suite, to make your subtitles word perfect in a fraction of the time.

Get your videos transLectured now!

www.translectures.eu
## transLectures workshop at OCWC2014

Grand Hotel, Ljubljana, 22 April 2014

<table>
<thead>
<tr>
<th>8:45 - 9:00</th>
<th>Welcome</th>
</tr>
</thead>
</table>
| 9:00 - 10:00 | Invited speaker: Title to be confirmed  
Presenting two key open source technologies that have enabled this university to build a highly scalable lecture recording system. | Stuart Phillipson  
Media technologies coordinator at the University of Manchester |
| 10:00 - 10:30 | Coffee break |
| 10:30 - 11:15 | transLectures: quality subtitles for video lectures  
An overview of this EU-funded project to develop low-cost, accurate and multilingual subtitles for educational video, and what it means for you. | Alfons Juan  
Project coordinator and tenure lecturer at the Universitat Politècnica de València (UPV) |
| 11:15 - 12:00 | Domain adaptation for subtitles you can take seriously  
Custom adapting transcription/translation models for specific use with educational videos is one of transLectures’ main lines of research, and is yielding great results. | Muhammed Ali Tahir  
Researcher at Rheinisch-Westfälische Technische Hochschule Aachen |
| 12:00 - 13:30 | Lunch |
| 13:30 - 14:15 | VideoLectures.NET and transLectures  
The largest of the project’s case studies and an award-winning video lecture repository, Rihtar talks integration, user evaluations, and impact. | Matjaz Rihtar  
Jožef Stefan Institute |
| 14:15 - 15:00 | Integrating transLectures into poliMedia  
poliMedia is a lecture capture and distribution system developed at the UPV; it is also one of the transLectures case studies. Turró introduces the tL player. | Carlos Turró  
Head of media services at the UPV |
| 15:00 - 15:30 | Questions and closing comments |
| 15:30 | End |

Register  www.translectures.eu
transLectures Try our Tools (ToT) service

transLectures is proud to announce that, after two and a half years of dedicated state-of-the-art research, we now have a working transcription and translation system available on our website. Intended as a 'try before you buy' function to allow universities and other MOOC providers to see the kind of quality they can expect from our service, it is open to anyone who needs subtitles for their educational videos.

The result of a joint project between three European universities, three commercial enterprises from various fields and a non-profit foundation, the transLectures “Try Our Tools” service (TOT) currently supports transcription and translation of videos in English, Spanish and Catalan (more languages coming soon!).

The process is very simple: 1. Sign up for a transLectures account. 2. Upload your media files. 3. Do a bit of “intelligent post-editing” using our bespoke post-editing suite, tl Player, which will highlight in red text what it considers problem areas. Et voilà! Your subtitles are available to download.

Visit our website to find out more: translectures.eu

www.translectures.eu
A.3.4 EMMA newsletter

EMMA: Providing multilingual access to European MOOCs

The European Multiple MOOC Aggregator: EMMA for short, is a 30 month project showcasing the diversity of MOOCs in Europe and piloting an innovative approach to their use. EMMA’s special features and tools will facilitate learners to follow whole MOOCs or to construct their own personalized learning path with units from several MOOCs and from different countries, as their building blocks. A transcription and translation tool will help students to access courses in multiple languages as the project goes on. Analytics and tracking data allow for monitoring and upgrading of the service in countries, including experts in technology, analytics. As the project evolves, other non MOOCs' available through the EMMA platform will be opened at the beginning of October. 

Inspire

Automatic Transcription and Translation of MOOC content

by Dr. Jorge Civera, Associate Professor of computer science at UPV, Spain (partner in EMMA)

One of the most appealing services of the EMMA platform is the possibility of having your course materials automatically translated into other European languages opening up your student community. In the case of audiovisual materials, such as videos, subtitles are automatically generated in multiple languages. At the current state of the project, transcription systems that generate subtitles from videos in English, Spanish, Italian, Dutch and Portuguese have been developed. Translation systems have also been deployed for the following translation pairs: from Italian, Spanish, Dutch and Portuguese into English, from Spanish into English and Catalan. The translation systems will be employed to translate both, subtitles and textual content of MOOCs in the original language. The Automatic Speech Recognition (ASR) technology behind the EMMA transcription systems have been developed at the European project transLectures (translectures.eu) coordinated by the Universitat Politècnica de València. This technology is based on state-of-the-art ASR techniques combining Deep Neural Network for acousting modelling and n-gram models for language modelling. The translation systems are grounded on the publicly available MOSES translation toolkit that implements a state-of-the-art phrase-based translation system. In addition, lecturers and students may have the possibility of reviewing automatic transcriptions and translations of their MOOC courses to improve not only current transcriptions and translations, but also future ones taking advantage of transcription and translation system updates from these reviewed materials. As the project progresses, MOOC courses from other European universities will be uploaded involving new languages, such as French and Estonian, being automatically transcribed and translated into English. Moreover, an English into Italian translation system will be put into place to translate Spanish MOOC courses into Italian (and vice versa) passing over English as a pivoting, quality assurance language. To find out more and to see the transLectures tools in action check: www.translectures.eu