



SUMAT: An Online Service for SUBtitling by MACHine Translation

CIP-ICT-PSP-270919



D9.3

Scientific and Industrial Activities Report (Y3)

Editor:	Anthony Walker
Company:	CAPITA-TI
Status-Version:	V.4
Date:	26/05/2014

Project Number:	CIP-ICT-PSP-270919
Project Title:	SUMAT

Title of Deliverable:	Scientific and Industrial Activities Report
Date of Delivery to the EC:	26/05/2014

Workpackage responsible for the Deliverable:	WP9
Editor:	Anthony Walker
Contributor(s):	Arantza del Pozo, Gerard van Loenhout, Mirjam Sepesy Maucec, Lindsay Bywood, Yota Georgakopoulou, Thierry Etchegoyhen
Reviewer(s):	Arantza del Pozo, Lindsay Bywood, Yota Georgakopoulou, Thierry Etchegoyhen
Approved by:	Arantza del Pozo

Abstract:	This report lists the scientific and industrial dissemination activities carried out by the partners during the final year of the project.
Keyword List:	Publication, dissemination

Document Description

Document Revision History

<i>Version</i>	<i>Date</i>	<i>Modifications Introduced</i>	
		<i>Modification Reason</i>	<i>Modified by</i>
V.1	07/04/2014	Draft Version Year 3	Anthony Walker
V.2	10/04/2014	Revised version	Arantza del Pozo
V.3	15/04/2014	Integration of contributions	Arantza del Pozo
V.4	26/05/2014	Extended version following final review recommendations	Yota Georgakopoulou, Lindsay Bywood, Thierry Etchegoyhen

Table of Contents

Part I - Executive Summary.....	6
Part II – Full description of deliverable content.....	6
1. Introduction	6
2. Dissemination strategy	6
3. Dissemination activities performed	8
4. Publications.....	14
5. Interviews	14
6. Pending dissemination activities and publications	15

List of Tables

Table 1. Dissemination activities performed 13

Part I - Executive Summary

This document provides the list of dissemination activities that took place during the final year of the project.

Part II – Full description of deliverable content

1. Introduction

This deliverable concerns the activities performed during the third year of the project (April 2013-March 2014) with regard to dissemination and other actions promoting awareness of the project.

2. Dissemination strategy

Following the EU review meeting in May, the recommendations made by the reviewers were taken on board and applied through Y3 of the project.

It was decided that a new dissemination strategy was required, that would involve the creation of a new website, a new leaflet with greater information on the project objectives and better material for use at exhibitions.

Accordingly, the marketing collateral material employed for display at events and leaflets to be handed out at presentations, exhibitions, and for further dissemination was revamped, ensuring that the presentation of SUMAT is consistent across media. The purpose of all the new collateral is to inform and add more detail about the progress and development of the SUMAT technology and to drive people towards the new website where an online demo of the service is available. More information regarding the new dissemination materials employed throughout the last year of the project can be found in D9.8_v3.

Particular emphasis has been placed on two dissemination aspects specific to the goals of the project and its future exploitation.

The first concerned the need to reduce residual negative perceptions of machine translation technology within the community of professional translators, since this technology is still perceived as a threat by a significant portion of the community. To this effect, a workshop was organised at one of the main industrial events, as described below, giving professional

users a much-needed platform for hands-on experience and constructive exchanges. Additionally, several Q&A sessions and interviews have been planned and executed, to further reach the community and stimulate wide-ranging constructive discussions on the use of MT for subtitling. From the feedback gathered so far, these targeted activities have been fruitful and well-received by the community.

The second targeted effort focused on the current and future need for subtitling professionals trained in post-editing. As the outcome of the project will involve actual exploitation of the technology by subtitling companies, it is crucial that professional translators are trained in the specifics of post-editing as well. To this effect, university seminars have been organised, as described below, where the findings and experience gathered throughout the SUMAT project have been used to offer training-oriented presentations targeting professionals in the domain. This dissemination activity has generated much interest and will continue beyond the completion of the project.

Targeted audiences have thus included the following communities identified as potential interested parties:

- The **scientific community**, through publication of the scientific results achieved within the project at two relevant conferences held in 2013 on machine translation and language technology: **MT Summit XIV** and the **6th Language and Technology Conference**.
- The **industrial community**, through participation in two of the main industrial events attended by subtitling stakeholders: **Media for All 5** – *“which brings together professionals, scholars, practitioners and other interested parties to explore audiovisual translation in theory and practice”* and **MIPCOM 2013** – *“the world’s entertainment content market”*. In addition, SUMAT also participated in three other smaller conferences targeting the subtitling field: **Subtitling: A Collective Approach** – *“which aims to bring together professional translators, fansubbers, broadcasters, researchers, consumers, software programmers, and students of translation with the aim of opening conversation and exchange between all those involved in the subtitling process in some way or other”*, **Translating and the Computer 35** – *“a meeting point for translators, researchers and students in the translation and localization field”* and **Subtitling and intercultural communication. European languages and beyond** – *“which aims to explore key concerns associated with subtitling and intercultural communication with a particular focus on European languages to promote the cross-fertilization of practices, ideas and theoretical approaches”*.
- The **subtitling academic community** was also addressed separately through targeted dissemination activities, which were prompted by the success of our workshop and

presentation at the Media for All 5 conference in September 2014. This event was attended by representatives of leading European universities in the field of audiovisual translation, who share our view that the post-editing role will soon become key for subtitling companies. In order to address the shortage of skills in the post-editing area, various universities have asked for our help to start training their students in the task of post-editing subtitles and relevant presentations and seminars have been arranged.

- The **wider public**, through the project website and LinkedIn group.

3. Dissemination activities performed

Table 1 summarizes the dissemination activities performed during the second year of the project, while the following paragraphs describe each of the dissemination events and SUMAT's involvement in more detail.

Subtitling: a collective approach, University of Nottingham / July 2013

At this conference, SUMAT gave a presentation entitled “Embracing the Threat: machine translation as a solution”, describing the project approach and preliminary end-user evaluation results. A demo of the system was also presented.

MT Summit XVI / September 2013

At MT Summit XVI, SUMAT presented two posters. One was entitled “SMT Experiments for Commercial Translation of Subtitles”, describing the lessons learnt from the various experiments carried out to improve system performance through linguistic information on the one hand, and out-domain data exploitation via mixture modeling on the other hand. The other presentation described the overall SUMAT approach and was entitled “SUMAT: An Online Service for Subtitling by Machine Translation”. Both posters were attended by an extended audience that showed interest in the project, the data used to build the systems and our results to date. The new leaflets were also handed out among the poster visitors.

Media for All 5 / September 2013

Media for All 5 has probably been the most relevant event in which SUMAT participated this year.

A 3-hour workshop entitled “SUMAT: Machine Translation Systems for Subtitles: A Hands-On Workshop” took place on September 25th. In the workshop, we first presented the scope and methodology of the project, including the approach used to collect and prepare aligned corpora, the statistical translation techniques that were developed and tested, and the

evaluation metrics for the SMT systems that were built. In a second part, attendees were given the opportunity to test the approach in real time, by going through the process of translating subtitles using the Live Demo. This interactive session was formed of small groups of users, with at least one project member attending to each group. As a result, we were able to informally gather feedback on the system's usability and usefulness, and give attendees the possibility to have in-depth discussions on the project approach with members of the consortium.

In addition, a presentation entitled "More subtitles, more languages: results of an extended evaluation of machine translation systems" was given at the conference, where we presented results of the systems' quality evaluation by end-users.

As in Languages & the Media last year, SUMAT generated strong feelings, both positive and negative, at Media for All 5. Many subtitle translators still fear the changes that introducing machine translation can bring to the industry. However, more translators were positive about post-editing as compared to Languages & the Media the year before. In addition, industry representatives and academics from the audiovisual translation field showed a lot of interest in the approach, and many of the university seminars mentioned below were arranged at the conference.

MIPCOM 2013 / October 2013

MIPCOM is a big industrial fair for the entertainment content market attended mostly by content producers, broadcasters and subtitling companies. SUMAT had a poster at the Dutch Public Broadcasters stand, where we handed out leaflets and answered questions.

Translating and the Computer 35 / November 2013

DDS and VSI were invited to give a SUMAT presentation at TC35. They entitled it "New tools for subtitle translation", and presented the SUMAT approach and Round 1 results.

6th Language and Technology conference / December 2013

Maribor University gave a presentation entitled "Improving Statistical Machine Translation with Additional Language Models" in the MT session of the conference, describing the n-best list rescoring approach employed for the Serbian-Slovenian language pair within the project, which has achieved good results.

Subtitling and intercultural communication. European languages and beyond / February 2014

In this conference, SUMAT representatives gave an overview of the project approach entitled "Machine translation and subtitling: an overview".

Seminars in Universities / ongoing

As the project progressed and the subtitling community learned more about the SUMAT approach, a number of universities contacted the consortium and expressed their interest in knowing more about the subtitle post-editing process. Scholars in charge of training professional subtitlers have acknowledged that subtitle post-editing is going to be a significant trend in the field of audiovisual translation and are looking for ways to start including it in their academic curricula. After the success of the hands-on workshop organised at Media for All 5, a series of seminars on post-editing machine-translated subtitles have been arranged in various universities across Europe.

The seminars are in the form of a two-part presentation. Part 1 focuses on post-editing from a high-level point of view, describing this new proposed workflow, the rationale behind it, the evaluation of the results of the SUMAT MT systems and the main project findings, supported by quotes from the SUMAT post-editors themselves. Part 2 focuses on the details of the job of the post-editor, comparing and contrasting it with the 'translation' and 'review' tasks subtitle translators are typically asked to work on. This part presents the error categories typically present in MT output, with concrete examples from the SUMAT evaluation, the solutions that could be applied to minimise such errors, as well as a first attempt to define the profile of subtitle post-editor. For seminars held in London (Roehampton and UCL so far), the two-part presentation was followed by a post-editors' panel, where three of the DDS post-editors answer free questions from the audience on their post-editing experience.

Seminars were carried out during the life cycle of the project at Antwerp University in February and Roehampton University in March. More seminars have already been carried out after the project end at Zurich University in April, University City London (UCL) and University of Vienna in May, and more are being planned for the coming months. A repeat of both the UCL and Zurich seminars for next year's students have already been confirmed, and a short summer course on post-editing is planned to take place at Salento University, provided there is sufficient student uptake.

Looking further into the future, Antwerp University have expressed their intention to start teaching post-editing to their students as a separate module. To this effect, a meeting was held in February between Yota Georgakopoulou and staff from the Applied Linguistics/Translation & Interpreting department of the university, to brainstorm on a possible syllabus for such a module and the tools that would be needed to teach it.

The seminars have been very well received and constitute a strong dissemination activity that bridges the interests of the academic and industrial subtitling communities.

Year	Date / Venue	Type	Website	Dissemination purpose (what was disseminated)	Dissemination activity (presentation, poster, demo)	Type of audience	Countries addressed	Partners responsible/involved
2013	12-13 July, Nottingham, UK	Subtitling: A Collective Approach, University of Nottingham, Centre for Translation and Comparative Cultural Studies	http://www.nottingham.ac.uk/ctccs/events/conferences-and-public-lectures/ctccssubtitlingconf12-13july2013.aspx	Integration of machine translation in the subtitling workflow	Presentation	Scientific	European	VSI
2013	September 2-6, Nice, France	MT Summit XIV	http://www.mtsummit2013.info/	SUMAT approach and evaluation results	Two Posters	Scientific	World-wide	Vicomtech, Textshuttle, Deluxe, TitelBild, inVision, University of Maribor

Year	Date / Venue	Type	Website	Dissemination purpose (what was disseminated)	Dissemination activity (presentation, poster, demo)	Type of audience	Countries addressed	Partners responsible/involved
2013	25-27 September, Dubrovnik, Croatia	Media for All 5	http://mediaforall5.dhap.hr/	SUMAT approach	Workshop and presentation	Scientific and Industrial	World-wide	All partners
2013	October, Cannes, France	MIPCOM	http://www.mipcom.com/	SUMAT approach	Poster and leaflets	Industrial	World-wide	inVision
2013	28-19 November, London, UK	Translating and the Computer 35	http://www.mt-archive.info/10/Aslib-2013-TOC.htm	SUMAT approach	Presentation	Scientific and industrial	World-wide	VSI and DDS
2013	7-9 December, Poznan, Poland	6th Language and Technology Conference	http://www.ltc.amu.edu.pl/	MT of subtitles	Presentation	Scientific	Europe	Maribor
2014	17 February, Antwerp, Belgium	Antwerp University	n/a	Post-editing seminar	Presentation	Scientific	Belgium	VSI, DDS

Year	Date / Venue	Type	Website	Dissemination purpose (what was disseminated)	Dissemination activity (presentation, poster, demo)	Type of audience	Countries addressed	Partners responsible/involved
2014	27 February, Siena, Italy	Subtitling and intercultural communication. European languages and beyond.	https://www.iatis.org/index.php/community-resources/calls-for-papers/item/783-subtitling-and-intercultural-communication-european-languages-and-beyond	SUMAT approach	Presentation	Scientific	European	VSI, DDS
2014	13 March, Roehampton, UK	Roehampton University	n/a	Post-editing seminar	Presentation	Scientific	UK	VSI, DDS

Table 1. Dissemination activities performed

4. Publications

The following journal publications have been completed:

L. Bywood, M. Volk, M. Fishel, & P. Georgakopoulou. "*Parallel Subtitle Corpora and their Applications in Machine Translation and Translatology*", *Perspectives: Studies in Translatology*. Special Issue: Corpus linguistics and AVT: in search of an integrated approach, Volume 21, Issue 4, pp. 595-610, 2013

P. Georgakopoulou "*SUMAT: sottotitolazione assistita dalla traduzione automatica*", In Eugeni, C. e L. Zambelli (a cura di) *Respeaking*. Specializzazione on-line. Numero monografico n.1 www.accademia-aliprandi.it pp. 120-123, 2013

P. Georgakopoulou and L. Bywood "*Machine translation in subtitling and the rising profile of the post-editor*", *Multilingual Computing* January/February 2014, pp. 24-28 <http://dig.multilingual.com/20140102/#?page=24>

The following conference proceedings have been published:

T. Etchegoyhen, M. Fishel, J. Jiang, M. Sepesy Maucec, "*SMT Experiments for Commercial Translation of Subtitles*", *Proceedings of MT Summit 2013*, pp. 369-370, User Track Poster Session, Nice, France

P. Georgakopoulou, L. Bywood, T. Etchegoyhen, M. Fishel, J. Jiang, G. van Loenhout, A. del Pozo, D. Spiliotopoulos, M Sepesy Maucec, A. Turner, "*SUMAT: An Online Service for Subtitling by Machine Translation*", *Proceedings of MT Summit 2013*, pp. 439, European Projects Poster Session, Nice, France

L. Bywood, T. Etchegoyhen, M. Fishel, P. Georgakopoulou, M. Volk, "*More subtitles, more languages: results of an extended evaluation of machine translation systems*", *Proceedings of Media for All 5*, September 2013

5. Interviews

Three interviews have also taken place on the subject of post-editing and machine translation in the subtitling domain:

- September 2013: A Q&A session with Lindsay Bywood and Yota Georgakopoulou was organised as promotional material to support the Media 4 All 5 conference, providing details on SUMAT and on the content of the workshop organized by the consortium..

http://mediaforall5.dhap.hr/Content/documents/5thMedia4All_Sumat_Interview.pdf

- November 2013: Following the Media 4 All 5 conference, Yota Georgakopoulou was interviewed by Dawn Jones, the *I Heart Subtitles* blogger, on the conference itself, SUMAT and post-editing subtitles in particular. The I Heart Subtitles blog is one of the most reputable blogs in the subtitling industry.
<http://iheartsubtitles.wordpress.com/2013/11/08/machine-translation-subtitles-qa-with-yota-georgakopoulou/>
- January/February 2014: Yota Georgakopoulou was interviewed by Lindsay Bywood for the Journal of Specialised Translation on her career, her current job, and more specifically on post-editing and machine translation. JoSTrans is a highly-regarded, peer-reviewed, international journal that serves as a forum for translators and researchers in specialized translation. The 25-minute interview can be seen here: http://www.jostrans.org/issue21/int_georgakopoulou.php

6. Pending dissemination activities and publications

Although the project is now officially over, several dissemination activities and publications are still underway and the consortium has agreed to carry them out within the next few months:

Subtitling: A Collective Approach

The following publication has been submitted to the conference organisers, to be published as a book:

L. Bywood, P. Georgakopoulou, T. Etchegoyhen, *“Embracing the threat: machine translation as a solution for subtitling”*, Subtitling: A Collective Approach, Editors Maike Oergel and Pierre-Alexis Mevel, Publisher: Bloomsbury, 2014

LREC 2014

LREC has become the major event for language resources and evaluation of human language technologies. The aim of the conference is to provide an overview of the state-of-the-art, explore new R&D directions and emerging trends, exchange information regarding language resources and their applications, evaluation methodologies and tools, ongoing and planned activities, industrial uses and needs, requirements coming from the e-society, both with respect to policy issues and to technological and organisational ones.

The following publication has been accepted for presentation at LREC 2014, to be held on 26-31 May in Reykjavik, Iceland:

Thierry Etchegoyhen, Lindsay Bywood, Mark Fishel, Panayota Georgakopoulou, Jie Jiang, Gerard van Loenhout, Arantza del Pozo, Mirjam Sepesy Maucec, Anja Turner and Martin Volk “*Machine Translation for Subtitling: A Large-Scale Evaluation*”, LREC 2014

FIT 2014

The FIT World Congress brings together all those involved in the language sector: translators, interpreters and terminologists, students and teachers, newcomers and experienced professionals, product and service providers. The 2014 edition will take place on 4-6 August in Berlin.

Lindsay Bywood and Yota Georgakopoulou will give a presentation entitled “*The changing role of the subtitler in the 21st century*”, showing a diachronic view of the development of the subtitler's role in context and making some tentative predictions about the future.

COLING 2014

The COLING conference, organised under the auspices of the International Committee on Computational Linguistics, is held every two years and covers a broad spectrum of technical areas related to natural language and computation.

A paper entitled “*A Mixed Approach for Quality Estimation and Filtering of Machine Translated Subtitles*” has been submitted to COLING 2014, to be held in Dublin on 23-29 August. Notification of acceptance/rejection to authors is expected by May 23rd.

MT Journal

Springer's Machine Translation Journal covers all branches of computational linguistics and language engineering, wherever they incorporate a multilingual aspect. It features papers that cover the theoretical, descriptive or computational aspects of many topics, including machine translation and machine-aided translation.

The consortium is currently preparing a paper entitled “*Statistical Translation of Subtitles*” to be submitted to the MT Journal.

Languages & The Media 2014

The 10th Languages & The Media conference will bring together researchers, language practitioners, translators, interpreters, software developers and all those who produce, market, or distribute audiovisual materials for information, entertainment or educational purposes to discuss if the increasing use of smart technologies in the field is leading to smart translations.

Yota Georgakopoulou (DDS), Lindsay Bywood (VSI) and Elitza Bantcheva (DDS) have been invited to organize a workshop on post-editing machine-translated subtitles as part of the pre-conference workshop programme.

More Seminars in Universities

Additional seminars on post-editing machine-translated subtitles have already taken place at *Zurich University* (April), *UCL* (May), and *Vienna* (May).

The consortium has agreed to keep the SUMAT website and Live Demo online at least until the end of the year, so that they can be used in the planned dissemination events described above. Following the final project review held in May 2014, discussions are taking place within the consortium to extend the availability of the project's website beyond the end of 2014.

7. Conclusions

This document comprises the report of the dissemination activities performed during the third and final year of the SUMAT project. As described, the dissemination strategy objectives have involved making the website and marketing materials a more consistent source of information on the progress of the project. Additionally, special emphasis has been placed on further reaching the subtitling community, through a dedicated workshop, Q&A sessions and interviews in key media arenas.

Six articles and papers have been published in journals and conference proceedings and further events and publications, either informative or focusing on project results, are still planned, while targeted dissemination has taken place and will continue to be delivered at universities that teach subtitling, so as to address the shortage of skills in the post-editing market.

Overall, the dissemination activities performed in the final year of the project have increased the project's impact within the targeted scientific, industrial and academic communities, generating positive feedback and increased interest. These activities have thus helped pave the way for better exploitation of the technology in the field of professional subtitling.