



SUMAT: An Online Service for Subtitling by Machine Translation

Georgakopoulou P.¹ yota.georgakopoulou@bydeluxe.com, Bywood L.² lindsay@vsi.tv,
Etchegoyhen T.³ tetchegoyhen@vicomtech.org, Fishel M.⁴ fishel@cl.uzh.ch, Jiang J.⁵
jie.jiang@capita-ti.com, van Loenhout G.⁶ gerard@ondertiteling.nl, del Pozo A.³
adelpozo@vicomtech.org, Spiliotopoulos D.⁷ d.spiliotopoulos@atc.gr, Sepesy Maucec M.⁸
mirjam.sepesy@uni-mb.si, Turner A.⁹ Anja.Turner@titelbild.de

¹Deluxe Digital Studios, ²Voice and Script International, ³Vicomtech-IK4,
⁴TextShuttle, ⁵Capita TI, ⁶Invision Ordertiteling, ⁷Athens Technology Center,
⁸University of Maribor, ⁹Titelbild Subtitling and Translation

www.sumat-project.eu

Description

Subtitling plays a key role in ensuring audiovisual content is accessible across languages. Owing to the increasing use and transmission of digital multimedia and multilingual content through the web, DVDs, and the cinema, as well as current European and national policies promoting the subtitling of content broadcast by public television networks, the demand for subtitle production is rising continuously. With content volumes increasing, relying on human translation alone is not a sustainable model. There is a need for new technology to enable the industry to grow along with the market.

The SUMAT project aims to explore the integration of machine translation into the subtitle translation workflow thus providing a much-needed tool to support the translation processes of both freelance subtitlers and subtitling companies. We are developing an online subtitle translation service catering for nine European languages in 14 language pairs, which will be available in 2014. The bi-directional language pairs are English-Dutch, English-French, English-German, English-Portuguese, English-Spanish, English-Swedish, and Serbian-Slovenian, thus covering languages spoken by approximately two thirds of the EU 27 population. The project is a collaboration between four subtitling companies (Deluxe Digital Studios, Voice & Script International, InVision Ordertiteling and Titelbild Subtitling and Translation) and five technical partners (Athens Technology Center, Capita Translation and Interpreting, TextShuttle, University of Maribor and Vicomtech-IK4).

The participating subtitling companies have contributed a large number of subtitle corpora as training material to build state-of-the-art statistical machine translation systems. Several experiments have been performed to increase the scope and quality of the baseline translation systems, ranging from the incorporation of linguistic information to the creation of translation models that combine additional data from various domains.

A large-scale evaluation by professional translators is now under way in all language pairs. The first round of the full-scale evaluation, which focuses on measuring MT quality both objectively and subjectively, is taking place between April and September 2013. In this phase, MT output is post-edited to professional standards, each machine translated subtitle is assigned a quality rank, and systematic errors are collected in order to improve the systems. The final MT systems will then be assessed for productivity gain in a second evaluation round between October 2013 and March 2014. An overview of the evaluation results to date will be offered by our SUMAT representatives in the poster area of the conference.

SUMAT is funded through the EU ICT Policy Support Programme (2011-2014) under grant agreement N 270919.