



Proceedings of the
**21st Annual Conference of
the European Association
for Machine Translation**

28–30 May 2018
Universitat d'Alacant
Alacant, Spain

Edited by

Juan Antonio Pérez-Ortiz
Felipe Sánchez-Martínez
Miquel Esplà-Gomis
Maja Popović
Celia Rico
André Martins
Joachim Van den Bogaert
Mikel L. Forcada

Organised by



Universitat d'Alacant
Universidad de Alicante

transducens
research group



The papers published in this proceedings are —unless indicated otherwise— covered by the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 International (CC-BY-ND 3.0). You may copy, distribute, and transmit the work, provided that you attribute it (authorship, proceedings, publisher) in the manner specified by the author(s) or licensor(s), and that you do not use it for commercial purposes. The full text of the licence may be found at <https://creativecommons.org/licenses/by-nc-nd/3.0/deed.en>.

© 2018 The authors

ISBN: 978-84-09-01901-4

news.bridge – Automated Transcription and Translation for News

Peggy van der Kreeft
Deutsche Welle
Projects & Development
Bonn, Germany

peggy.van-der-kreeft@dw.com

Renars Liepins
LETA
Latvian News Agency
Riga, Latvia

renars.liepins@leta.lv

Abstract

news.bridge provides a platform for multilingual video processing, including automated transcription and translation, subtitling, voice-over, and summarization, with post-editing facility of videos in a broad range of languages. The platform is currently in beta testing at Deutsche Welle for republishing of videos in other languages.

1 The Project

news.bridge is an 18-month research project funded by Google Digital News Initiative (DNI). It is currently running (January 2018 – July 2019) and is a follow-up of a 6-month prototype DNI project, which proved the viability and potential of the concept. Therefore, news.bridge was started to turn this into a deployable and possibly commercially exploited platform. There has been major interest from the broadcasting and wider media world to participate in beta testing the platform, which shows high potential. news.bridge is a small consortium of four members. (1) German broadcaster Deutsche Welle is coordinator and user partner; (2) LETA, the Latvian News Agency, is platform developer; (3) Le Mans University is technology provider, in particular for transcription, translation and punctuation; and (4) Priberam, a Portuguese spinoff, focuses on summarization.

Visit our website (<http://newsbridge.eu>) for more details and follow us on Twitter (@newsbridge_htl).

2 The Platform

The news.bridge platform is a modular, dockerized system, developed by LETA, which currently runs remotely, but can also be installed locally. It ingests content from the Deutsche Welle repository, processes (transcribes, translates) it on demand, after selecting specific items. As it is meant for use in the production department, it has a post-editing user interface, allowing corrections to be made to the translation and transcription. To widen the number of languages covered, it includes internal, customized translation services from universities as well as off-the-shelf services, such as Google Translate, IBM Watson transcription and voice-over, via API. This combination leads to a tool that covers an extremely wide range of language combinations. Over 100 languages are included.

3 Multilingual News Production

Deutsche Welle, as coordinator and user partner, is currently enhancing and testing the tool for production. It has run a few user evaluation workshops and more are planned. It envisages different use cases. The tool is used, for instance, for transcription of interviews in any of the languages covered by the system. It is being tested for reprocessing of DW videos (with existing transcripts) in several languages and is being evaluated for gain in time and effort. The use of existing transcripts is a major factor, as the error rate of transcription (e.g. names, missing punctuation) is much higher than that for translation. Overall, initial evaluation results indicate that journalists welcome a tool that can help them produce videos cross-lingually. A high-quality transcript in one language is a solid basis for subtitling and voice-over into different languages. Finding the best tools for each language pair is part of the benchmarking effort.