

**Report on the R&D situation for automated translation, speech
recognition and speech synthesis in Europe (Summary)**

January 2016

NICT Europe Center

General summary

Part 1- The EU support policy for the R&D of automated translation, speech recognition and speech synthesis

The EU's R&D support policy for automated translation and speech technologies in FP7 and Horizon 2020

- For some particular languages (English or French), automated translation technologies are developed and the quality of translation has been improved. But for the other languages, this progress is at a standstill. Horizon 2020 aims to improve the quality of translations into all EU languages.
- But, though some people consider language barrier is an obstacle for European digital single market, the EU's R&D support Policy for language technologies has changed in Horizon 2020.
- The workframe budget for the R&D for automated translations has barely remained in the ICT Work Programme 2014-2015 of Horizon 2020 (the research topic called " ICT-17 : Cracking the language barriers amounts to €15 million), but language technologies together with transcription and automated translation applications will be supported in the research topic called "Media and content convergence" as technologies enabling access to media and content in the Work Programme 2015-2016. So, Horizon 2020 doesn't focus on language technologies any more. Some researchers have expressed their dissatisfaction with this lack of financial support for language technologies.

Machine translation service of the EU : MT@EU

- MT@EC is the online machine translation service developed under ISA programme (the Interoperability Solutions for Public Administrations programme). This is a service for officials in the EU institutions and EU member states' public administrations.
- MT@EC provides the most precise of EU documents.
- MT@EC translates literally from and into all EU official languages. When more precision is required, translation should be improved by human translators.
- Before the introduction of MT@EC, another automated translation service, called ECMT, had been used (until December 2010). While ECMT is built on rule-based machine translation technology, MT@EC is built on statistical machine translation technology.
- The EU's Connecting European Facility programme (CET) supports the R&D of machine translation platform for public administrations in Europe, called CEFAT (CEF Automated Translation). CEFAT will be built on MT@EC for translating the EU 24 official languages, Norwegian and Icelandic.

Part 2 The R&D situation regarding automated translation, speech recognition and speech synthesis technologies in Europe

UK: Statistical Machine Translation Group at the University of Edinburgh

- The Statistical Machine Translation Group at the University of Edinburgh is famous for the development of a machine translation system called MOSES.
- This research group participates in many FP7 and Horizon 2020 research projects.
- Examples of FP7 projects include: Accept, Casmacat, EU-Bridge, Matecat, MosesCore, EuroMatrixPlus, LetsMT .

- Examples of Horizon 2020: QT21, MMT, HimL, TraMOOC, Cracker

Ireland: ADAPT (ex CNGL)

- Like the United Kingdom, Ireland is an English-speaking country. As corporate tax is low, many foreign companies have set their European headquarters in Ireland. As a result, the localisation activities, including automated translation, are very active. In order to help with this social background, the Centre for Next Generation Localisation (CNGL) was established in 2007.
- This research center has been renamed ADAPT since January 2015,
- In ADAPT, the researchers from four universities (Trinity College Dublin, Dublin City University, University College Dublin and Dublin Institute of Technology) and the researchers from their industrial partners work together.
- In addition to automated translation, ADAPT also develops financial services, E-commerce, digital media, entertainment games, digital culture and human sciences, E-learning, etc.
- KantantMT.com and ICONIC are ADAPT's spin-off companies developing automated translation technologies.

Germany: Interactive Systems Lab at Karlsruhe Institute of Technology (KIT)

- Interactive Systems Lab at Karlsruhe Institute of Technology researches and develops speech and text automated translation, face tracking, attention tracking, multi-modal tracking, and neural network.
- In 2012, this laboratory delivered an automatic simultaneous speech translation system for university lectures.
- As regards text machine translation technology, this institute research focuses on European languages. And it has for several years participated in WMT, an International workshop on statistical machine translation.
- Concerning the development of text automated translation technology, it has participated into Franco-German Quaero project, FP7's EU-Bridge project, and Horizon 2020's QT21 project.

Germany: German Research Center for Artificial Intelligence (DFKI)

- The German Research Center for Artificial Intelligence conducts the research and development for language technology, including machine translation, in two groups : one's name is Text analytics (in Berlin) and another is Multilingual technologies (in Saarbrücken).
- The language technology lab at DFKI is a key research centre in Europe. It coordinates META-NET, which enhances an alliance of European research centers for language technology. It is also the coordinator of the CRACKER project funded through Horizon 2020, which strengthens the alliance of European research centers for machine translation.

France: Informatics Laboratory at the University of Maine (LIUM)

- The Informatics Laboratory at the University of Maine, established in Le Mans, France, studies speech language translation, robust speech recognition, speaker diarization, speaker recognition and speaker language understanding. It is involved in research on the applicability of deep neural networks into all these research fields.
- Deeplingo and Voxolab are the LIUM's spin-off companies. Deeplingo offers customized machine translation models for companies. Voxolab offers customized speaker identification application and customized transcription application for companies.

- LIUM has just launched the M2CR project (2015-2017) through the EU's CHIST-ERA. This project aims to establish a unified architecture, based on deep neural networks, for speech understanding and translation and image understanding. LIUM also participated into FP7's MateCat project (2011-2014) for the development of the Computer-assisted translation (CAT).