# EU★BRIDGE

**Bridges Across the Language Divide**
**www.eu-bridge.eu**

## EU-BRIDGE – the Goal

*Bridges Across the Language Divide* (EU-BRIDGE) aims to develop speech and machine translation capabilities that exceed the state-of-the-art in new and more challenging business cases. It will advance spoken language technologies to permit transmission of human information content from one language to another in situations that cannot be handled by automatic techniques to date.

Another key challenge is to reduce the cost of data collection and the dependence on large, manually collected and annotated data for the development of the core recognition and translation technologies. The team of partners proposes to build one of the largest combined language capabilities in speech recognition and translation.

EU-BRIDGE seeks to achieve rapid technology transition and market insertion by building distributed services instead of transferring software and by making deployment part of the project.

## EU-BRIDGE – Objectives

Effective, innovative and independent alternatives must be provided for Europe to serve its translation, communication, content processing, and publishing needs. The joint expertise of the consortium will drive the technological development forward, ensure best performance of core capabilities and ensure that insertion in actual business operations is achieved. At the core of EU-BRIDGE's objectives is the development of a speech translation service infrastructure upon which several use cases will be built.

## EU-BRIDGE – Research

In order to build the service infrastructure for speech translation, engines from the academic partners will be integrated into a service framework that offers an easy-to-use infrastructure for the use case developers. Therefore, the academic partners perform research at their laboratories on the following:

### Performance:
We will advance spoken language technologies so they process and transmit human information content from one language to another in situations that can so far not be handled by automatic techniques. This includes specialized but varied topics (lectures, seminars, presentations), as well as highly disfluent, conversational, accented, and noisy speech (meetings, telephone calls). We will perform research in the areas of robustness, rapid adaptation in speech and translation, semantic modelling, and content summarization. We will also develop personalization schemes that adapt systems to individual users and groups of users for more specific high performance operation that will address business needs better than a web-based one-size-fits-all solution.

### Language Portability for Europe:
EU-BRIDGE will provide speech and translation capability for languages of main interest to Europe. Building on key efforts and expertise, it will build one of the largest combined repertoires of languages available both in speech recognition and translation and will stretch to do so robustly for all communication channels. We will include core European languages and under-resourced European languages and will reach out to languages of the BRIC economies. Focused research efforts to improve portability itself play an important role. These efforts will lower the cost of moving capabilities effectively from one language to another.

### Reduction of the Dependency on and Cost of Data:
Since data is the "crude oil" of information processing, solutions must make production cheaper and reduce our dependency on it. By making speech and machine translation components adaptive and language and style independent, and by streamlining the process, EU-BRIDGE will significantly reduce data needs. By involving the users themselves in correcting and building the systems

implicitly, i.e. by crowd sourcing, the cost of data acquisition will be reduced, thus, building and improving the systems. By taking better advantage of available but not well-prepared data, the cost of data preparation will be reduced and the effective usable data increased. This includes comparable data, monolingual data, spoken and textual data, noisy data, and automatic methods for judging the quality of data.

## Rapid Technology Transition and Market Insertion:

The program will strive to transition research development to commercial deployment more rapidly. This will be done by building distributed services instead of transferring software and by making deployment part of the project. The systems will be applied to real-world data, and we will carry out pilot experiments and field studies around four business opportunities.





## EU-BRIDGE – Evaluation and Dissemination

The consortium will monitor the progress of our activities in this project and thoroughly evaluate the success. EU-BRIDGE will evaluate the improvements from the core research activities on in-house test sets using automatic measures. The test data for the in-house tests will be closely related to the use cases developed in this project and will be collected with the help of the partners heading the use cases. The engines developed for the integration into the speech translation service infrastructure will be evaluated on an international stage against worldwide participants in external evaluation campaigns.

Finally, all activities, results, and progress will be disseminated through various channels not only to the scientific community, but also to the general public, decision makers, and most importantly to the European industry that would benefit from the availability of the service infrastructure which is intended to live beyond the life-time of the project. The infrastructure is intended to be used in many more applications than the four use cases of the project, which will be the pioneers inspiring other users and prove the feasibility of the concept.

## EU-BRIDGE – Four Use Cases

In order to show the benefits from the speech translations services developed in EU-BRIDGE, to prove the feasibility of industry being able to build applications that use the services, and in order to stimulate and foster the development of further applications by other industry partners, the following four use cases will be developed:

### Captioning Translation Use Case for TV Broadcasts:

In this use case a service will be developed that will improve the work progress in captioning TV shows and translating these captions into multiple languages. This use case will demonstrate that by using machine translation technology it is possible to reduce the costs for producing subtitles for media content in many languages. Thus, it will be possible to subtitle and translate more media content and make it available to Europe's citizens.

### University Lecture Translations:

Using the services developed by EU-BRIDGE, a system for simultaneously translating university lectures will be advanced and will be installed in lecture halls at KIT. This use case will demonstrate how speech translation technologies can be applied in situations in which the use of human simultaneous translators would be too expensive and in this case making university lectures available to a much wider audience, e.g., international students.

### European Parliament Translations:

EU-BRIDGE will explore and identify opportunities for applying speech translation technology in the translation process of the European Parliament in cooperation with the Directorate Generals for Interpretation and Conferences – Support to Multilingualism. Possible application scenarios are for example, those situations in which human translators are not available.

### Unified Communication Translation:

In this use case, we will introduce speech translation technology as a web service within a unified communication platform. Depending on the devices used and the mode in which they are being used, translation will be done from the speech and will be translated into another language and output as audio (in particular for mobile use), or possibly in text form, e.g. as subtitles. This use case will demonstrate how the user experience when using different devices in a communication setting can be improved by using speech translation technology, and can help to bridge the language divide in day-to-day situations for Europe's citizens and business.

**Project Coordinator**

Karlsruhe Institute of Technology
Prof. Alexander Waibel
Adenauerring 2
76131 Karlsruhe, Germany
phone: +49 721 608 44730
e-mail: alexander.waibel@kit.edu

**Contact**

Karlsruhe Institute of Technology
Press and Communication
Margit Rödder
Adenauerring 2
76131 Karlsruhe, Germany
phone: +49 721 608 48676
e-mail: margit.roedder@kit.edu

**www.eu-bridge.eu**

## EU-BRIDGE – Project Partners

EU-BRIDGE brings together academics, engineering, and business experts in order to create competitive solutions to existing translation, communication, content process-ing, and publishing needs. Ten partners from university/research and industrial sectors form the EU-BRIDGE consortium:

- **Karlsruhe Institute of Technology**, Germany, Adenauerring 2, 76131 Karlsruhe
  Contact: Alexander Waibel, Web: www.kit.edu
- **Fondazione Bruno Kessler**, Italy, Via Santa Croce 77, 38122 Trento
  Contact: Marcello Federico, Web: www.fbk.eu
- **Polish Japanese Institute of Information Technology**, Poland, ul. Koszykowa 86, 02-008 Warszawa, Contact: Krzysztof Marasek, Web: www.pjwstk.edu.pl/en/
- **RWTH Aachen University**, Germany, Templergraben 55, 52056 Aachen
  Contact: Hermann Ney, Web:  www.rwth-aachen.de
- **The University of Edinburgh**, United Kingdom, South Bridge, Edinburgh EH8 9YL
  Contact: Philipp Koehn, Web: www.ed.ac.uk
- **The Hong Kong University of Science and Technology**, China, Clear Water Bay, Kowloon, 999077, Contact: Dekai Wu, Web: www.ust.hk/eng/index.htm
- **Red Bee Media Limited**, United Kingdom, 58 Uxbridge Road, W5 2ST London
  Contact: Nicola Greaves, Web: www.redbeemedia.com
- **Pervoice SpA**, Italy, Viale Verona 190/1, 38123 Trento
  Contact: Dario Franceschini, Web: www.pervoice.it/en/
- **Accipio Projects GmbH**, Germany, Roermonder Str. 216, 52072 Aachen
  Contact: Volker Steinbiss, Web: www.accipio-projects.eu
- **Andrexen,** France , 28 Rue de Châteaudun, 75009 Paris
  Contact: Grégoire Boutonnet, Web: www.andrexen.com