EU-BRIDGE FINAL REPORT

Authors

Sebastian Stüker, Marcello Federico, Philipp Koehn, Hermann Ney,
Margit Rödder, Matt Simpson, Volker Steinbiss,
Alessandro Tescari

Scientific Coordinator

Alexander Waibel

The work leading to these results has received funding from the European Union under grant agreement n°287658.

31.01.2015, Karlsruhe, Germany

1. Executive Summary

As we enter into the 2nd decade of the 21st century, Europe is facing larger and more critical language challenges than ever before. After several decades of political integration the language barrier within the multilingual Europe remains as the last barrier against true integration.

Since bridging the language divide in Europe by means of human translators or by support in language learning is impossible, we must turn to automatic methods in the form of machine and speech translation.

In order to build the necessary technology and make it available to application developers and users, EU-BRIDGE has pooled academic, engineering and business expertise in Europe. Through this joint expertise EU-BRIDGE has achieved three goals: 1) to enhance the core technologies of speech translation systems, 2) to offer them in en easy-to-use fashion to application developers, and 3) to insert the technology directly into the market by implementing four use cases.

The goal of making language processing and, especially, speech translation technology available to application developers, EU-BRIDGE has implemented an easy-to-use service architecture which runs speech and language processing services in the cloud. On the side of the core technology developers the service architecture allows the flexible integration of engines, e.g., for automatic speech recognition and machine translation, from arbitrary entities. On the side of the application developers the service architecture allows access via the Internet through a simple-to-use API.

To achieve this goal EU-BRIDGE has conducted research along two lines: 1) to improve machine translation and speech recognition to a level to be useful in the planned use cases, 2) to discover and exploit cheap data sources and to make use of their data for training, in order to lower the time and costs for training new systems.

In order to demonstrate the effectiveness of the service architecture the consortium has successfully implemented four use cases:

1. Captioning of broadcast news: In this use case we have developed systems for automatically captioning broadcast news such as BBC weather forecasts and Skynews shows under the leadership of the consortium member Red Bee Media.

2. Simultaneous translation of lectures: In this use case we have deployed a system at KIT that simultaneously translates lectures given in German into English.

3. Support of interpreters at the European Parliament: The consortium has developed a tool to support the interpreters of the European Parliament in their preparations for upcoming meetings.

4. Simultaneous translation of webinars: Consortium member Andrexen has integrated simultaneous speech translation into its webinar application, in order to translate webinars from English into French.

All use cases have been successfully field tested. For the captioning use case the system was evaluated by the Red Bee Media using the same procedure and metrics as being used for evaluating their business captioning service. For the other use cases we conducted user studies under real-world conditions. All user studies showed that the translation services offer an added value to the users making it easier for them to navigate through a multilingual Europe.

Project Context and Objectives

The challenge of multilingualism in Europe

As we enter into the 2nd decade of the 21st century, Europe is facing larger and more critical language challenges than ever before. After several decades of political integration and considerable effort in language training and language related services, the European dream of true multilingualism still appears elusive. Three new factors now contribute to its complexities:

- European Expansion
  With continuing expansion of the EU, the number of languages and language pairs has not only increased considerably, but many of the new member states (for historic reasons) do not share common secondary languages (e.g., English, French, Russian) thus making effective communication more complex.

- Emerging Economies
  As the world is changing, emerging economies outside the European framework and its linguistic cousins, have asserted a dominant role on the global stage beyond Europe's borders. For Europe's export led economies, effective communication and interaction with several of these economies (foremost, BRIC nations, Brazil, Russia, India, China) outside its borders have become a necessity as well.

- Exploding Communication
  A growing offering of media and communication channels continues to broaden the way in which we absorb and disseminate information. Aside from TV, Radio and Internet, mobile devices, social media, blogs, podcasts, tweets, feeds, now demand distribution in multiple languages.

The production of multilingual content now far outpaces our ability to translate it by human effort and we must turn to automatic methods to cope. Luckily, the same changes that exacerbated the problem, now also contribute to better and more effective solutions. The growing availability of large volumes of data make automatic learning approaches for better transcription and translation feasible and indeed quite successful. And now also, the adoption of cloud computing, the rapid spread of mobile phones (now almost 5 times more in numbers globally than PCs!), and the emergence of app stores, further encourage and enable the use of speech and natural language technologies on a broad scale. This has generated a sudden and intense interest in speech and translation worldwide by strong players. Just in 2009/2010 alone, voice dictation (Dragon), speech search (Google), video conferencing (Apple), speech translation (Jibbigo), text translation (Google), all arrived on mobile phones. And Google, Apple, Microsoft, Cisco and Baidu all have been expanding their research teams in speech recognition and machine translation aggressively.

This appears to be great news as it suggests that years of research on speech and translation will finally break down communication barriers for everyone. But for Europe, the job of multilingualism is not yet done. For a united Europe with many regions, languages and information needs, 1.) broad coverage of and between many of its languages must be ensured in a balanced way, 2.) privacy, security and confidentiality of information must be ensured for its citizens and corporations, 3.) independent and adaptable technology services must be available to provide customized solutions to European governments, businesses and individuals. These desires cannot only be satisfied by companies supporting information search (like Google or Baidu), since their business models are based on advertising, and thus rely on processing and analysing all data passed through their services to derive an increasingly good model of their users for better matched advertising. While this provides users with better information access at a great price or for free, it cannot guarantee full privacy and independence of its content processing. But just as Cable TV, which commands fees, complemented and in many ways replaced free advertisement supported broadcast TV, there is a market for "premium" multilingual content processing services that are secure, private, independent, commercial free, of better quality, and network independent. Among them likely are multilingual content processing services that support publication of high quality content, creation/maintenance of confidential records (healthcare, personnel, corporate…), telecommunication, enterprise solutions behind firewalls, and services requiring disconnected operation (police, emergency response, military).

Supporting multilingualism through speech translation services

Effective, innovative and independent alternatives must therefore be provided for Europe to serve its translation, communication, content processing and publishing needs. We believe that Europe has the people and resources to create competitive offerings of such alternative, innovative services. We intended to mobilize such resources by pooling academic, engineering and business expertise in Europe. The joint expertise was supposed to drive the technological development forward, ensure best performance of core capabilities and ensure that insertion in actual business operations is achieved. The services should not be a one-size fits all solution for generic multilingual needs, but to be explored within specific markets and businesses. As a technology driver, the advancing systems were to be developed and tested first and throughout the project in several use cases, e.g., a commercial media transcription and translation service with the objective of lowering cost and increasing volume for European media publishing businesses. The project's scientific and technological agenda was guided by these objectives and includes four important elements:

- Develop better state-of-the art speech and MT capabilities in view of new and more challenging business use cases
- Improve language portability and apply the technology to languages of interest to Europe
- Reduce the dependency on data
- Explore/direct/facilitate rapid market insertion and deployment:

Performance: The project was expected to advance spoken language technologies so they process and transmit human information content from one language to another, in situations that could until then not be handled by automatic techniques. This includes specialized but varied topics (lectures, seminars, presentations). The projected thus intended to do research in the areas of robustness, rapid adaptation in speech and translation and semantic modelling. Another important objective was also to develop personalization schemes that adapt systems to individual users and groups of users for more specific and targeted high performance operation that address business needs better than a web-based one-size fits all.

Language Portability for Europe: The objective was here to provide speech and translation capability for languages of main interest to Europe. Building on key efforts and prior projects such as Euromatrix, Gale, TC-STAR and Quaero the project's team of partners was uniquely positioned and motivated to build one of the largest combined repertoires of languages available both in speech recognition and translation. The objective was to include core European languages, under-resourced European languages, and reach out to languages of the BRIC economies. The objective was to not will achieve this not just by a gargantuan engineering exercise, but by focused research efforts to improve portability itself. These efforts will lower the cost of moving capabilities effectively from one language to another.

Reduce the dependency and cost of data: Data is the "crude oil" of information processing and solutions must make production cheaper and reduce our dependencies on it. First by making speech and MT components adaptive and language and style independent and by streamlining the process, the project set out to 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 and thus building and improving the systems was to be reduced. By taking better advantage of available but not well prepared data, the cost of data preparation was scheduled to be reduced and the effective usable data to be increased. This included comparable data, mono-lingual data, spoken and textual data, noisy data, and automatic methods for judging the quality of the data.

Rapid technology transition and market insertion: The program targeted to transition research, development into commercial deployment more rapidly. This was planned to be done by building distributed services instead of transferring software, and by making deployment part of the project. The systems were planned to be applied to actual media captioning and translation services and the project's prototypes were supposed to be evaluated in terms of performance, but also effectiveness in commercial trials. For this pilot experiments around additional business opportunities were planned.

To achieve these objectives the consortium includes leading research laboratories in Europe, known for their state-of-the art speech recognition and translation technology and expertise with many of the languages of interest in Europe. It also include SMEs who have pioneered mobile speech translation apps and software and net based speech products and are keen and able to build a workable service infrastructure. One of Europe's premier media captioning and translation service, Red Bee Media, set out to provide access to massive amounts of media data and provide a demanding application environment for the consortium to assess and optimize

real live production workflow through improvements in speech and translation technology. Effective insertion in different markets were schedules to be attempted with additional partners and use cases.

By achieving the objectives described the project set out to make cross-cultural understanding and communication in yet unseen applications possible and language barriers transparent. Ultimately the project's overall objective is to contribute to forming an integrated European community without language barriers.

2. Main S&T results/foregrounds

The main scientific and technological results and exploitable foreground of EU-BRIDGE fall into four categories.

1. The project has driven progress in speech and language processing technologies, mainly automatic speech recognition, machine translation and speech translation.

2. The consortium members took the results of the research activities and transformed them into usable engines that perform such tasks as automatic speech recognition, machine translation, language identification or the punctuation of speech recognition output.

3. The project has implemented a service infrastructure into which the project's speech and language processing technology can be plugged and offered to application developers. On both side of the service infrastructure, the side of the technology developers and the side of the application developers, easy-to-use APIs were implemented that make it possible to easily plug-in and out new technology as it develops, and to easily access the language technology services by application developers without the need of having deep inside into the way this technology works and needs to be handled.

4. The project has realised four use cases that are concrete applications that make use of natural language processing technologies. These use cases were realised with the help of the service infrastructure, thus proving the effectiveness of the infrastructure realised.
In the following we will describe each of the results in more detail.

2.1. Scientific progress in speech and language technology

The scientific work within EU-BRIDGE was organised in two work packages:

- Work Package 1: Transcription, Translation and Production of Multimedia Content
- Work Package 2: Reducing Data Dependency and Costs

2.1.1. Results of WP1: Transcription, Translation and Production of Multimedia Content

Within Work Package 1 (WP1), the EU-BRIDGE project partners conduct research in the area of spoken language translation in order to advance the state-of-the-art and enable the project to provide engines, such as automatic speech recognition (ASR) and machine translation (MT) systems, that are of sufficient quality to provide the services envisioned by the project.
The most prominent areas of interest for WP1 are the following six

1. **Robust audio processing:** In order for speech recognition systems to perform well in real-life tasks they need to be robust to varying acoustic conditions, such as back-ground noises, reverberation etc. EU-BRIDGE has therefore researched methods to achieve better

robustness.

2. **ASR and adaptation:** Automatic speech recognition systems still perform best, when they are optimally adapted to the spefic tasks that they are supposed to be applied to, e.g., recognition of a certain lecture or recognition of a specific TV show. The adaptation of an ASR system to a task has traditionally been very labor intensive, requiring work by skilled personnel which is expensive and hard to find. EU-BRIDGE has therefore developed methods to semi-automatically adapt ASR systems to new domains with little adaptation data required and with minimal human intervention. Thus the adaptation process has become much more affordable and therefore applicable in real-world applications.

3. **MT and adaptation:** With respect to domain specificness MT systems have the same limitations as ASR systems. They only work optimal when exactly tailored to the specific task that they are supposed to work in. The translation of lectures needs a different system than the translation of webinars. Thus for MT, EU-BRIDGE also has researched new methods for adapting MT systems to new domains with little human intervention and with as little as possible data needs.

4. **Named entities for ASR and MT:** Named entities, such as person names, names of cities and places, are a great challenge for ASR and MT systems, often limiting their perceived performance. The vocabulary of an ASR system is usually fixed before being deployed. Words that are not in the vocabulary of the ASR system cannot be recognized. This often affects named entities, because of the sheer number of possible items and the poor predictability of their occurence in a task. For MT systems the same holds true as for ASR systems. Only named entities in their models can be properly translated. While passing unknown words through in translation by just replicating the original word can somewhat alleviate the problem, it often destroys language model context and very often leads to imperfect or even not understandable translations. EU-BRIDGE has therefore researched methods of how to be better recognize and translated named entities in order to improve the performance of the systems as perceived by the user.

5. **Morphosyntactic models for MT:** Morphology in languages very often poses a challenge to machine translation. Translating morphologically rich languages can often lead to explosions in the size of vocabulary and translation models that are necessary for good translations, leading to the problem of obtaining sufficient amounts of training material for estimating these models. Further the translation output needs to adhere to a potentially complex morphology with complex agreements between morphs in the target language. Therefore EU-BRIDGE has researched morphosyntactic models that produce valid output also for languages with a complex morphology, such as, e.g., Polish.

6. **MT and semantic context:** Modern statistical machine translation systems traditionally occur semantic information relying soley on statistics of co-occurences of phrases in training data. However, semantic context is important in translation in order to resolve semantic ambiguities. Also, semantic analysis of the result of a translation system is important in

order to correctly asses the quality of a translation. EU-BRIDGE has therefore performed work in integrating semantic knowledge into the translation and evaluation process, e.g., via such measures as MEANT, HMEANT or xMEANT.

2.1.2. Results of WP2: Reducing Data Dependency and Costs

Within Work Package 2 (WP1), the EU-BRIDGE project partners conduct research in finding and exploiting cheap data sources, such as the data from the TED website or TV broadcasts. In order to utilize the data obtained from these cheap data sources the consortium examined different techniques:

1. **Data selection for acoustic model training:** From the cheaply annotated data harvested, methods were researched to train acoustic models for speech recognition systems on them. This was mainly done by identyfing portions of the data for which reliable transcriptions were already available or could be automatically created.

2. **Comparable data for machine translation:** Often multilingual data is only available in comparable for instead of sentence parallels form as required for training machine translation systems. We thus researched methods for extracting parallel data from comparable corpora.

3. **Bootstrap methods for underressourced languages:** For many languages only insufficient training data resources are available. The consortium thus examined ways of applying training data across languages in order to boost the performance of ASR systems for under ressourced languages.

4. C**rowd sourcing techniques:** The traditional way of annotating training data by experts is expenisve. Thus within EU-BRIDGE we researched methods for obtaining annotations in a crowd-sourcing manner, perferably by volunteers, e.g., students listening to a lecture.

2.1.3. Evaluation of the scientific activities

The methods researched within these two work packages, as well as all the other work within EU-BRIDGE was evaluated using scientific methods that are common and best practice within this research community. Besides assuring quality the evaluations performed also drive progress in this area through the concept of Coopetition. Coopetition" is an artificial word composed of "cooperation" and "competition". Within this framework research groups compete on a common task in order to implement the best system. After system performances have been measured in a competitive evaluation, the groups then come together in a joint workshop and publish the design, research and implementation of their systems. From this, all groups can identify deficiencies in their systems and can identify new research directions.

Evaluations to drive the scientific progress were done at two levels:

1. Internal evaluation campaigns were regularly conducted. Only members of the EU-BRIDGE consortium participated in these campaigns. These campaigns were conducted on tasks that were of relevance to the project, i.e. were part of one of the four use cases to be realized within EU-BRIDGE, and had the goal to let consortium members develop the best engines for these tasks.

2. External evaluation campaigns: Consortium members regularly participated in external evaluation campaigns, most prominently the Workshop on Machine Translation (WMT) and the International Workshop on Spoken Language Translation (IWSLT).
EU-BRIDGE consortium members are part of the organising committee of both external evaluation campaigns–WMT and IWSLT–and were thus able to steer the workshops in such a way as to optimally support the goals of EU-BRIDGE.

2.2. MCloud: A service infrastructure for applications that use speech and language technology

The service architecture MCloud is a platform that abstracts the integration of client applications and the transcription/translation service providers. Based on a lightweight library, the MCloud lets application developers easily create complex transcription and translation workflows without knowing anything of the underlying transcription and translation engines.
The Service Architecture decouples clients and service providers by providing a simple, XML based protocol and a reference implementation library, available for the major platforms, to connect both end-user application and service engines to it.
Creating transcription and translation workflows requires the audio to be processed in specific sequence by multiple engines. E.g., to get translated and punctuated text in Italian out of an English speech one needs to invoke the following engines:

- English speech to English text transcription engine;
- English phrase segmentation engine;
- English to Italian translation engine;
- Italian punctuation engine.

The Service Architecture simplifies the creation of this workflow by providing automatic workflow creation given the input and output language pairs (called fingerprints).
The Service Architecture provides APIs for both, batch and real-time processing, supporting all of the transcription and translation needs. To simplify the batch processes integration, it also provides a set of web based REST APIs.

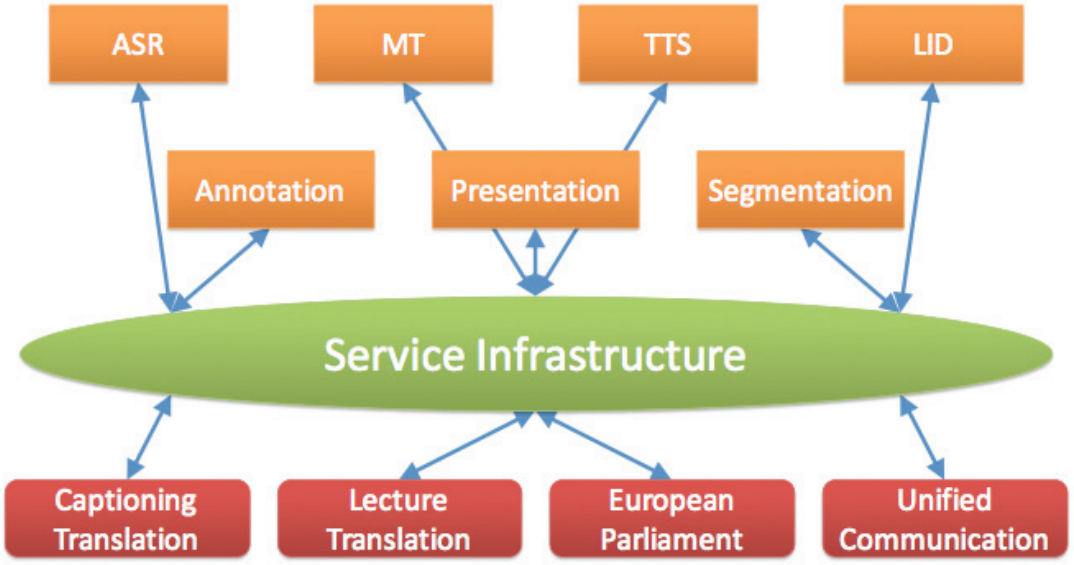


Figure 1 Overview of the role of the MCloud service infrastructure

2.3. Four use cases: TV captioning, simultaneous lecture translation, support for interpreters at the European Parliament, simultaneous translation of webinars

2.3.1. TV captioning

The goal of the use case is how automatic speech recognition techniques can be of use in a Broadcast TV environment in the preparation and transmission of a subtitling service.
Originally, Red Bee together with the partners had planned to field test the Weatherview system which had been chosen because of its commercial viability: Several regional weather forecasts have to be served with a seemingly limited style of speech and limited variation (a forecast does typically not change much within a few hours). Although this expectation turned out to be incorrect – the style of language proved to be much richer than anticipated – the transcription accuracy turned out to be better and more usable than the commercial had expected. Shortly before the field test, we thus decided to add a second use case, notably the transcription of Sky News broadcast, which originally had not been foreseen for the field test but was supported by the project's technology partners on short notice. The language is somewhat richer, but the main difference to us is in the business case: While the Weatherview system is used in an offline setting, supporting our personnel in preparation of the subtitles (business case: efficiency gain, cost savings), the Sky News system is intended to be used in either an offline/prepared setting or a live setting, the latter as a fall-back should the human service fail for any reason (business case: loss cut, fall-back system – Red Bee can be financially penalized for any missing or wrong subtitles).

The main goals of the use case were thus:

- To utilise the MCloud infrastructure in order to provide access to Automatic Speech Re-

cognition (ASR) workers provided by the other project partners.

- To assess the outputs of ASR workers against the business requirements of broadcast captioning.

- To provide feedback about the results of testing to the project and partners to inform iterative improvements and further developments.

- To provide a usable technology demonstrator which could, using a simple workflow, realistically illustrate the creation and use of automatically created captions internally (within Red Bee) and to the project.

- To compare and contrast automatically generated captions with those created within Red Bee Media by traditional methods in order to articulate any differences in a simple way to partners and the project.

- To assess the suitability of the Weatherview ASR text produced for use in a real broadcast environment, e.g. to assist with the existing production processes for the preparation of file-based subtitles.

- To assess the suitability of the Sky News ASR text and latency for use in an emergency situation where it may be used to cover any failures in the standard live, human-originated live subtitling processes, or the offline preparation of pre-cut news packages.

For testing and experimenting purposes Red Bee developed the Clip Captioner application was It generates subtitles for a media clip chosen by the user. It can show a transcript of the speech content of the clip, present the text as subtitles overlaid on the video and save the timed text files in a number of different, common formats.

Drop-lists allow selection of an ASR worker from the published list of available, active ASR workers which produce a text output.

A second, optional list gives the ability to add a secondary worker to provide punctuation or translation of the output of the first worker. In this case the second list is populated by the available workers which can accommodate the output of the primary worker as their input.

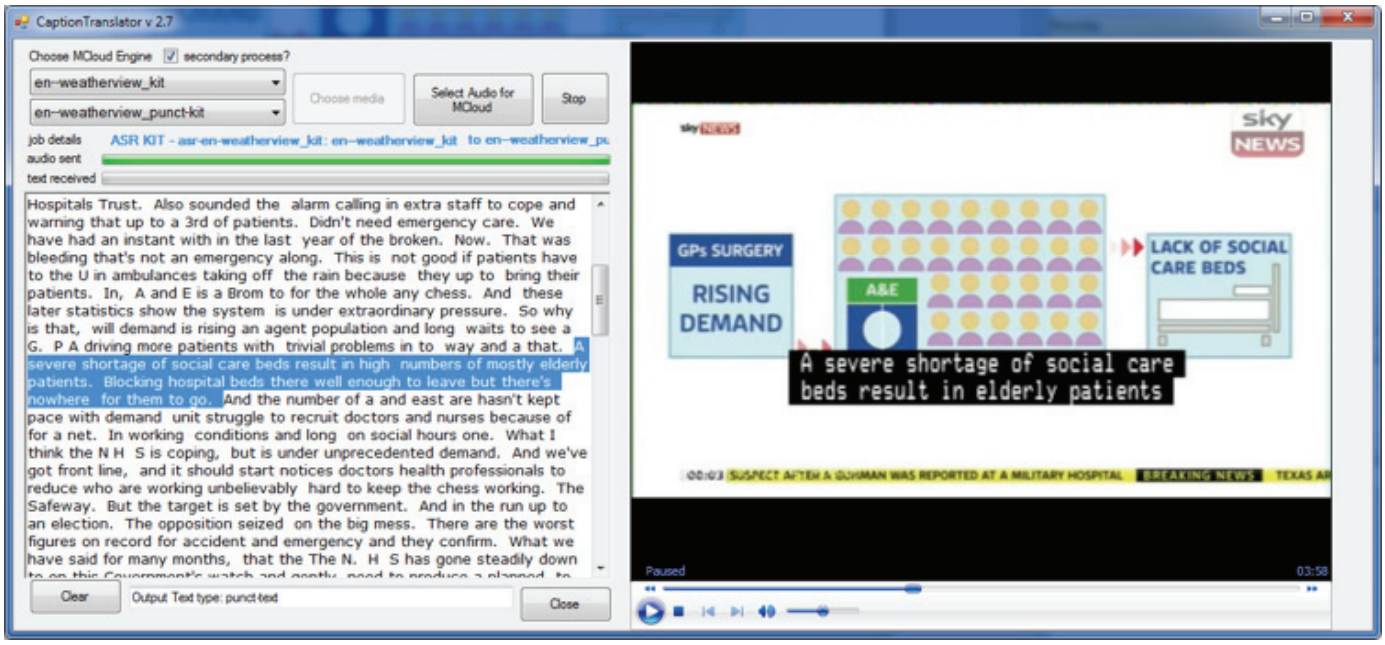


Figure 2 Screenshot of the Clip Captioner

Field test

Field testing was undertaken by two members of Red Bee Media staff. One focused on testing the deployment of MCloud and Red Bee Media's ability to access MCloud. The other focused on the accuracy of the resulting captions. The deployment tests were run on nominated days. Partners engaging in the field tests were notified of these days in advance and asked to ensure their workers are available for testing. Tests were run on a single day and repeated multiple times over the course of several weeks. The intention was to use the results to address potential deployment issues and to track accuracy improvements as engines continued to be developed during field-testing.

Red Bee Media used an internal mark scheme to determine the accuracy of captions. This scheme was applied to captions produced by the automatic service in order to assess their accuracy. Sky News content was marked from the top of the hour to 10:00 minutes on audio file. For Weatherview, only actual weather content marked, so continuity announcements either side ignored.

As a result of the field test the infrastructure worked well on the whole. The interface and client tools developed were greatly improved by feedback and co-operation between the various partners; it is true to say that the success of the field testing process was down to the unified interfaces between all parties and that any future large consortium project should consider such an approach to avoid unnecessary effort being wasted on establishing interfaces between engines and engine users.

In the field tests the accuracy of the engines was as good as 95% for the Weatherview shows and 90% for the Skynews tests.

2.3.2. Simultaneous translation of university lectures

Academic lectures offer valuable content but often do not reach their full potential audience due to the language barrier. Human translations of lectures by interpreters are too expensive to be widely used. Speech translation technology can be an affordable alternative in this case. At the Karlsruhe Institute of Technology (KIT) most lectures are held in German. This is often a significant obstacle for students from abroad wishing to study at KIT, as they need to learn German first. In order to be able to truly follow the often complex academic lectures, the level of proficiency in German that the foreign students need to reach is quite high.

This use case therefore pursues the goal of aiding those students by bringing simultaneous speech translation technology into KIT's lecture halls. For this, the lecture given by a speaker is captured by a microphone, automatically transcribed by automatic speech recognition, and translated by a machine translation system. The output of the translation system is then displayed to the user as a text via a web page. In that way students can access the translation via laptops or smartphones while participating in the lecture. The output in form of text allows the listeners to follow the interaction in the lecture room and support their understanding by simultaneously following the automatic translation, possibly skipping back and forth in time. Optionally the translation result can also be output in the form of synthesized speech through the web interface.

The lecture translation system is now routinely running in the KIT's main lecture hall (Audimax) and two lecture halls of the faculty of computer science. A permanent installation captures the audio from the lecturer's standard microphone and forwards it to the EU-BRIDGE service infrastructure, which is running in a server room at KIT.

The actual use of a system depends on the consent of the lecturer due to data privacy laws. Therefore not all lectures taking place in these lecture halls can be serviced. During the last two terms at KIT (summer term 2014 and winter term 2014/2015) the following lectures were automatically interpreted:

Summer term 2014:

- Product development - methods of product development
- Production management and marketing / production operations management \\ \hline
- Higher mathematics
- Power-train systems technology A - automotive systems
- Basics of computer science
- Algorithms I
- Computer organisation
- Cognitive systems

Winter term 2014/2015:

- Tutorial mechanical design

- Programming
- Measurements and control systems
- Accounting
- Higher mathematics I
- Higher mathematics III
- Finance and accounting
- Automatic visual examination and image processing
- Automatic speech recognition
- Concepts and application of work flow systems
- Mechano-informatics and robotics

System architecture

Servers at the Institute for Anthropomatics and Robotics at KIT are running workers for speech recognition. These servers receive the audio stream from the lecture hall. The processed output is then handed over to another worker which translates the text from German into English. The routing of the data streams is conducted by a third component, the Mediator. It routes the different data streams and manages the communication. The processed output (the transcription and translation) is then displayed via a web site.

By using the EU-BRIDGE service architecture it is possible to translate multiple lectures in parallel. The three lecture halls are serviced using two servers that provide the necessary computing power. This approach also offers redundancy which is required in order to minimize service disruptions during periods of maintenance.

The recording client that performs this task is installed on a small server that is integrated into the media hardware in the lecture hall.

The recording client is also connected to the projection hardware so that slides that are being projected by the lecturer are captured as well.

On the website of the lecture translator (http://www.lecture-translator.kit.edu), students can watch a schedule of upcoming lectures and can log in to a lecture as soon as it has begun. The access to a lecture is protected by an optional password, in agreement with the lecturer. Once logged in, the student can follow the live transcription of the German speech as well as the live translation into one of several target languages.

At the same time the lecture is also recorded and stored in a background archive. This archive can be accessed by the students via a platform-independent web interface. Lectures can be browsed by title and date. The archive contains the automatic transcriptions of the source language and parallel translations to one or more target languages. Playback of the recorded audio is accompanied by time synchronous parallel text output. Transcriptions as well as translations are downloadable in the universal Subrip text format for subtitle files. The archived data can also be utilised by the lecture translation framework for unsupervised adaptation of the system.

After performing experiments in WP2 on the topic of unsupervised acoustic model adaptation and training on the lecture task, KIT has integrated the research results into the lecture translation system. For this KIT has built a system module that handles several sets of acoustic models for different speakers. As soon as new data for an already known or previously unknown speaker becomes available, an acoustic model adaptation training pipeline is triggered fully automatically. The lecture translator speech recognition components are updated automatically as well, as soon as new models for a respective speaker become available. Previous model iterations are stored and indexed in an archive, enabling the option of a roll-back to past states, if necessary. In this way it becomes possible to build up a database of speaker dependent models, one for every lecturer, over time in a cheap way while leading to optimal performance for the ASR component.

In addition to acoustic model adaptation, the translation system also has an integrated support for dictionary and language model adaptation for specific lectures.

Based on textual material provided by the lecturer, such as slides and papers, relevant previously unknown terms are automatically extracted and added to the recognition system, so that specialised terms that often appear in university lectures can be recognised.

The system also incorporates several techniques researched in WP1 such as class-based language models and multiple reordering models to improve the translation quality.

Advertising the system

At the beginning of each term, the system was presented to the students. KIT staff went to one of the first lectures of each course to give a short talk about the background of the system, to explain how to login and use it, and to show screen shots. In the summer term, KIT additionally distributed fact sheets with the most relevant information.

For the winter term KIT increased the communication and dissemination activities. It designed business cards, posters and fliers to improve visibility. KIT also created a new landing page with an easier to read web address. Professors were also asked to publish the information concerning the service with other information about their lectures online. Posters were put up throughout the university, especially in public places like the university restaurant or in front of lecture halls. Fliers and business cards were printed and handed to the international office. Business cards were also distributed during the presentation of the lecture translator as a reminder of the web page and the service. Moreover, information was spread via social media and various mailing lists. Figure 3 shows the advertisement card distributed for the lecture translator while shows the poster which we used for advertisement.

Figure 3 Advertisement card for the lecture translation system

Figure 4 Lecture translator advertisement poster

User interface

The system's interface to the students offers a modern look and is optimized for the display on various screen sizes, ranging from small screens, like smartphones and tablet PCs to bigger screens used on laptops. As shown in Figure 5 and Figure 6, it is possible to switch between splitting the output horizontally and vertically. The individual paragraphs are being marked by an alternating shading. This way, the parts of the transcription and translation belonging together can easily be identified.

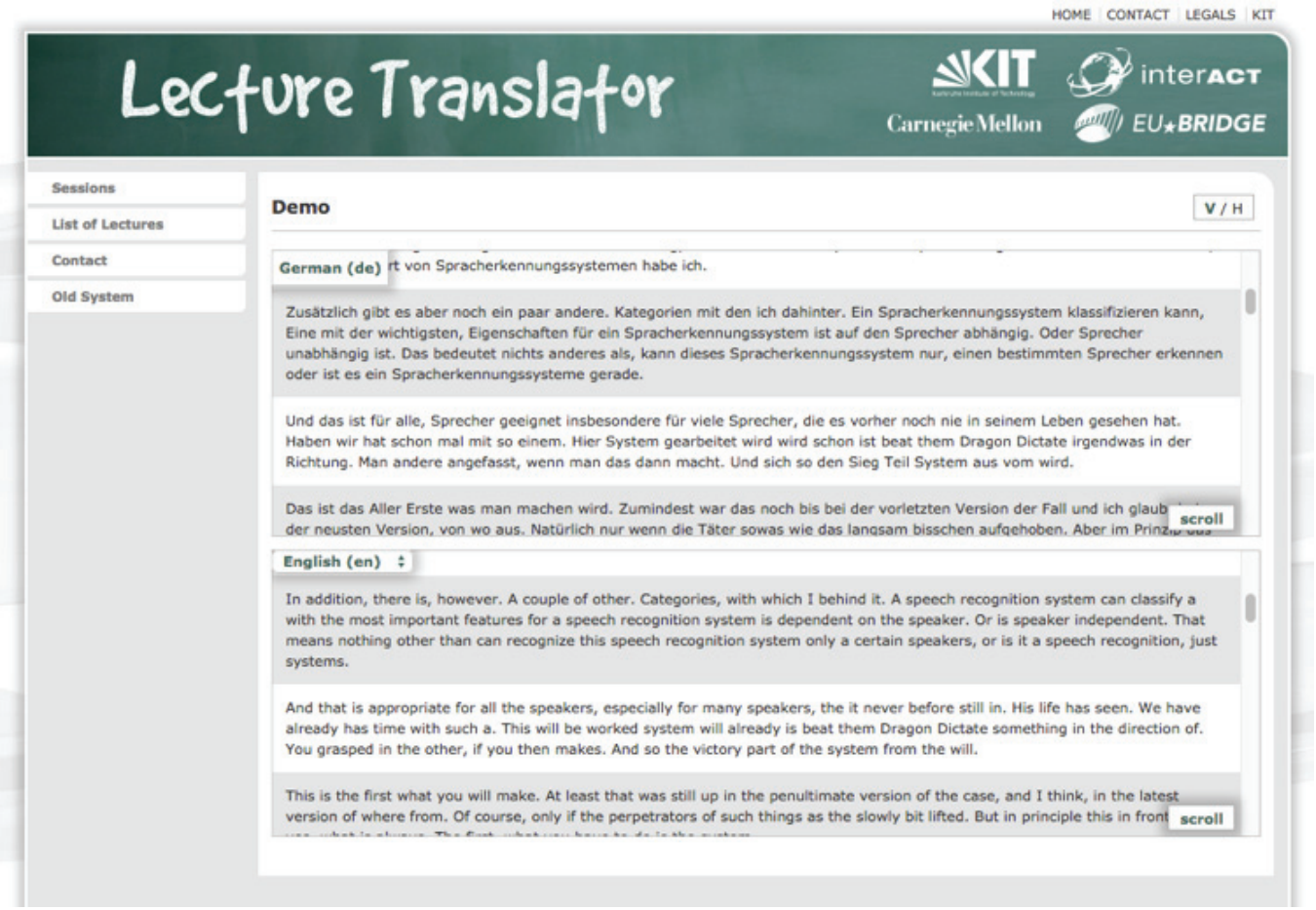


Figure 5 Lecture translator website with horizontal layout for the translation result

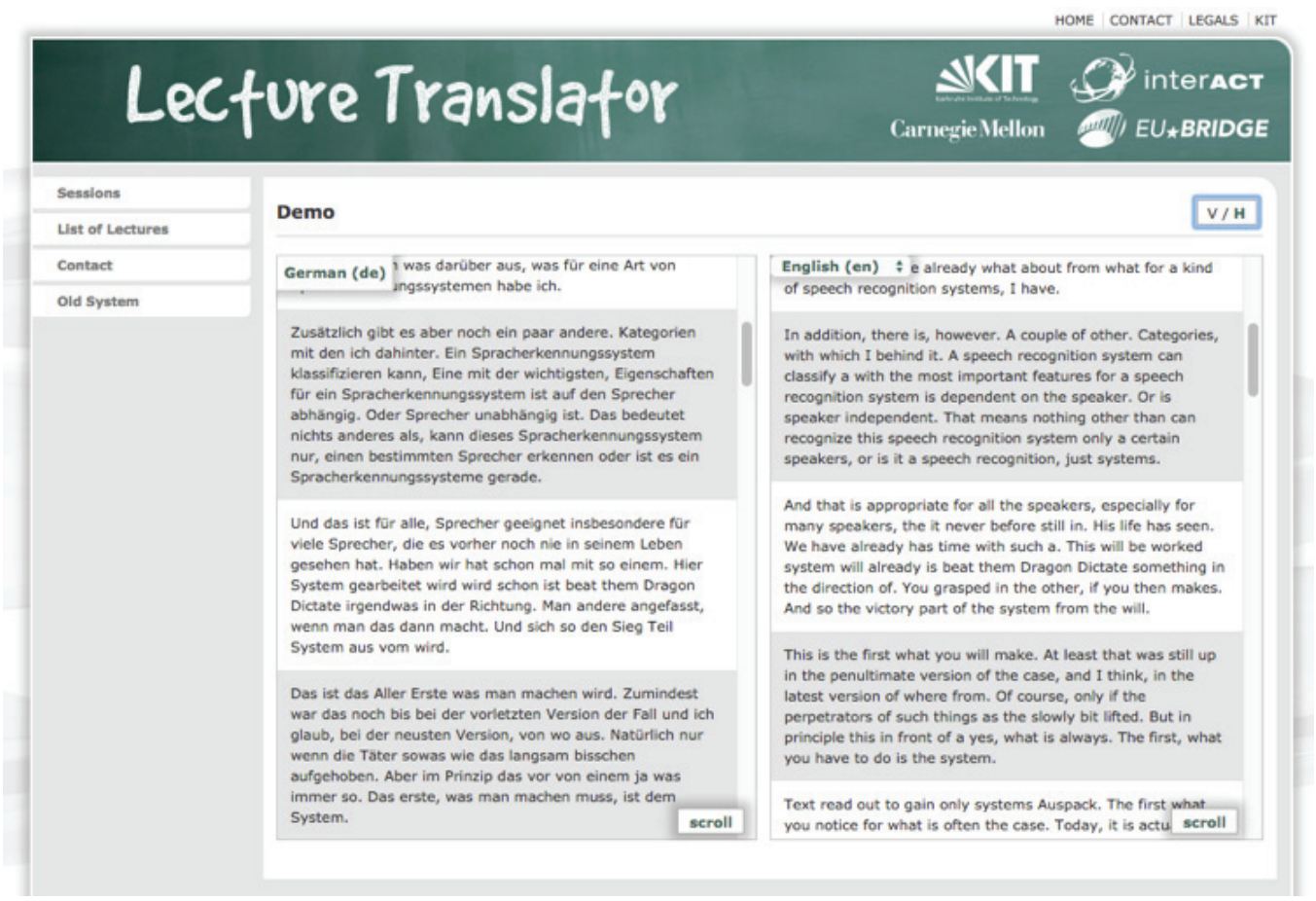


Figure 6 Lecture translator website with vertical layout for translation result

The access to the transcriptions and translations are password-protected in order to prevent unauthorized users from accessing them. The login credentials are distributed at the lecturers' discretion. Figure 7 shows the web page that gives an overview of the upcoming sessions as well as the field for entering the credentials of the ongoing lectures.

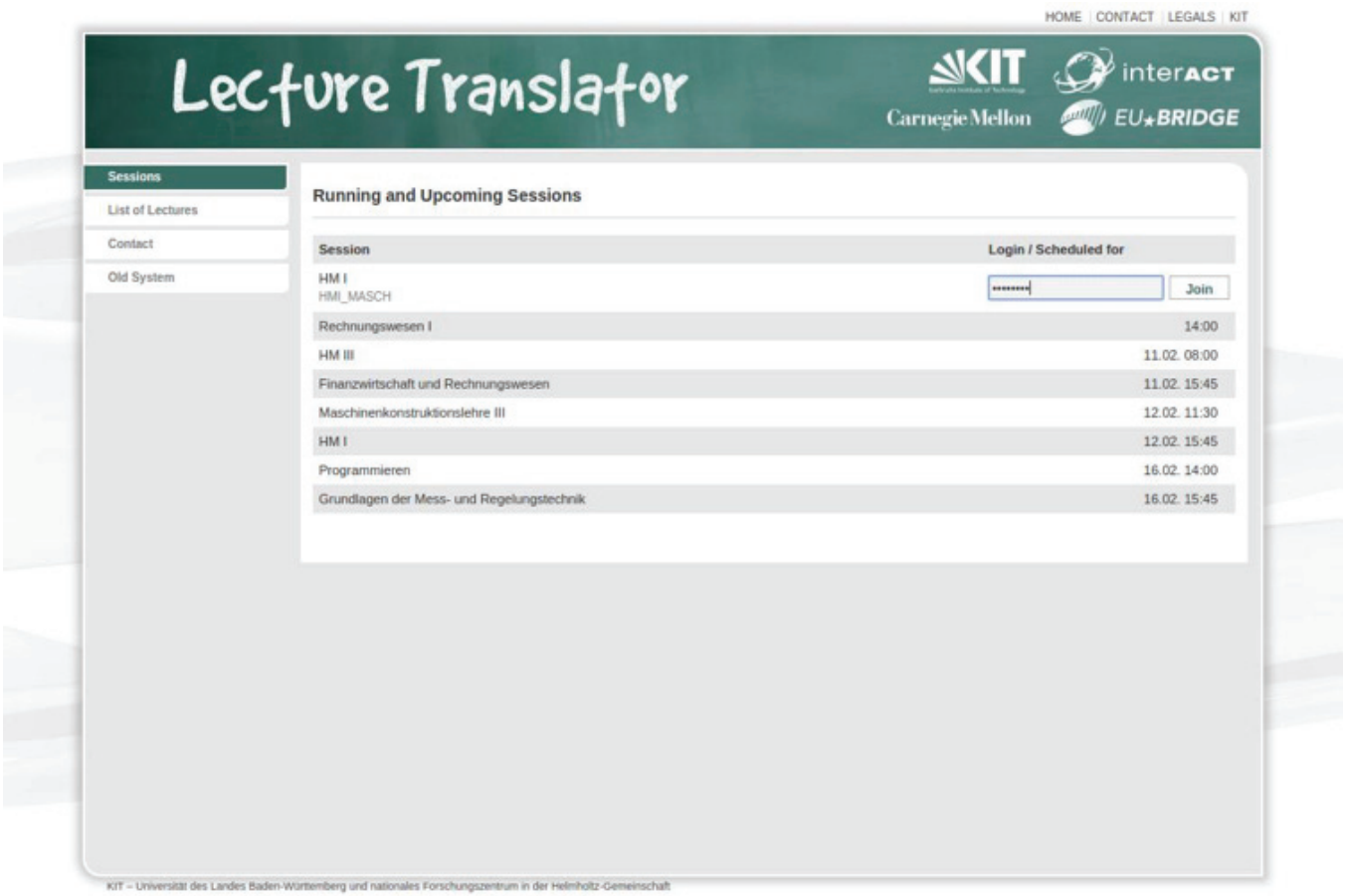


Figure 7 Lecture translator overview of upcoming sessions

User study

In the summer term 2014 as well as in the winter term 2014/2015 a full-scale user study and a field test of the system were conducted. The most important part of this test was a questionnaire which we asked users of the system to fill out at the end of each term.

The questionnaire contained questions concerning the background of the users, a system evaluation, an evaluation of the components „speech to text transcription“ and „machine translation“ and a possibility to express ideas and identify problems.

All questions where a rating was involved provided a scale ranging from one (worst option) to five (best option) and an additional field n/a, for those cases where the question could not be answered or an answer could not be given. An excerpt of the questionnaire is shown in Figure.

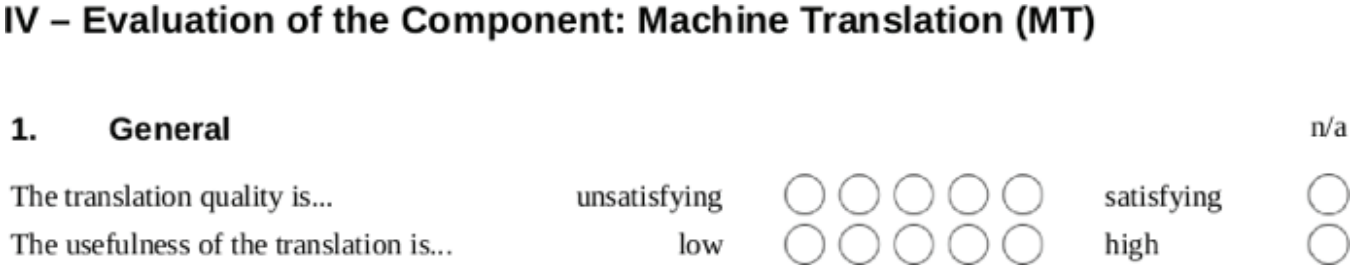


Figure 8 Excerpt from the questionnaire used to evaluate the lecture translation system

Overall, 22 students from five lectures (Computer organisation, Cognitive systems, Programming, Finance and accounting, Accounting) answered the questionnaire.

The results for the section overall system evaluation are shown in Table 1.

The general impression was rather positive, with 3.21 points on a scale from one to five. German students rated the system slightly better than foreign students. It was also considered rather useful, with 3.23 of 5 points.

When asked in more detail about the perceived usefulness, especially foreign students thought that it improved their understanding of the lectures and said they would find it useful in other lectures, too. However, they were not so sure about their performance and whether the LT made it easier to follow the lectures. Some students explained the latter phenomena by saying they sometimes considered it difficult to switch between the lecturer, the slides, and the LT-screen.

The ease of use was also rated positively, with 3.27 points. The layout of the user interface was considered very clear and got the highest marks from both groups.

Table 1 Lecture translator overall evaluation result

Criteria	Scale	Result
The service is...	terrible – wonderful	3.23
The system is...	not useful – useful	3.23
Using the LT makes it easier to follow the lecture.	disagree – agree	2.81
I would find the LT useful in other lectures.	disagree – agree	3.94
I enjoy using the LT.	disagree – agree	3.19
The features provided are sufficient.	disagree – agree	2.86
The layout of the user interface is clear.	disagree – agree	4.27

The questionnaire also asked for suggestions and things to improve. Ten students responded by putting in one or more comments. Suggestions and things to improve listed by the students were:

- A translation service for the slides or the script.
- A log or an archive of lectures to be downloaded.
- Improvement of algorithms and quality.
- Concentration on Chinese, as the Chinese are supposed to be the largest group of foreign students.
- Reduce the time lag of the LT and provide a more stable speed.
- Reduce inaccurate translations, especially of technical terms.
- Reduce difficulties with abbreviations and technical terms.

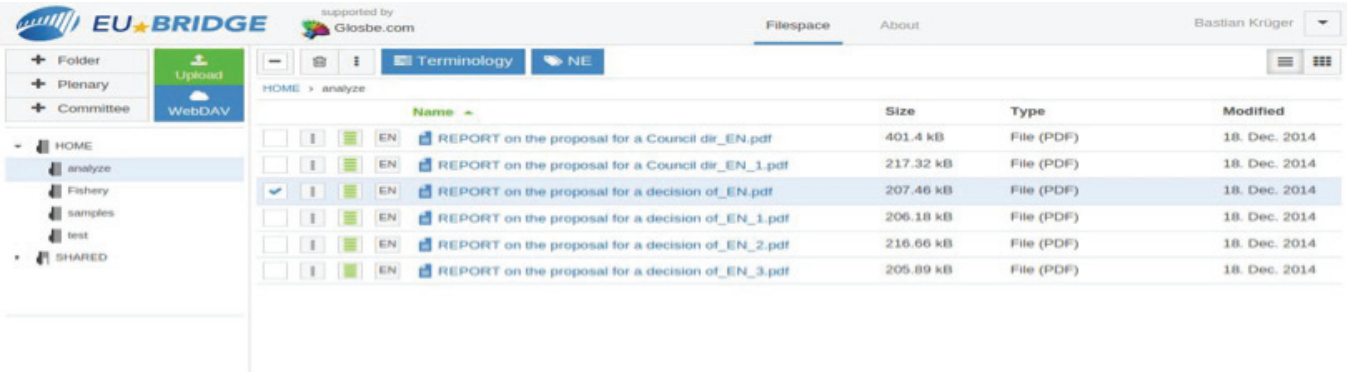
There were also comments that showed that students found the project interesting, that they were interested in its future development and above all its further improvements.

2.3.3. Support tool for interpreters of the European Parliament

In Discussions with the interpreters at the European Parliament, EU-BRIDGE identified a clear opportunity to support them by providing a tool that supports them when preparing for sessions. This led to the development of the interpreter support tool (IST).

The interpreter support tool is a Python-based, platform independent web application that can be accessed from within (almost) any web browser on (almost) any operating system. It was successfully tested on Windows, Linux, Mac OS X and iOS, using the web browsers Firefox, Chrome, Internet Explorer (>=9) and Safari.

This tool developed by KIT provides a modern, easy-to-use user interface, which is optimised for desktop computers and tablets. The user interface was improved successively and adapted to the needs of the interpreters by frequently collecting feedback from a small group of volunteers.



The interpreter support tool provides each user with their own account, making it easy to provide individual customisations and personal storage.

This tool consists of several, logically separated, work areas:

1) A fully functional file space: here the user can upload documents or import documents from upcoming plenary or committee meetings. The files for plenary and committee meetings are fetched automatically from the website of the European Parliament on a daily basis.

2) The terminology extraction section: here the user can create a terminology list out of several documents, translate the terms into several target languages, edit the terminology list to fit their needs and save or download the result. The extracted terms can also be shown within the context of the original document text.

3) The named entity section: here the user can extract named entities from one document and view the results in form of a list or highlighted inside the document text. Named entity categories can be individually selected.

Terminology support

As part of their preparation, interpreters go through the documents provided for the new session and look for terminology which is particularly hard to interpret. Based on the same (PDF) documents, the interpreter support tool allows interpreters to retrieve terminology lists and get the translations from English into the other 24 other European languages with the online resources IATE and Glosbe. In addition, five languages (German, French, Spanish, Polish and Finnish) are also supported by the IST to be translated from English with the alignment extraction on Translation Memory data. This system offers the interpreters the ability to perform terminology and named entity extraction on the EP plenary sessions, committee meetings and manually uploaded files, all within a unified user interface. The interpreters can see the term translations into more than one language at a time and add new term translations and save them.

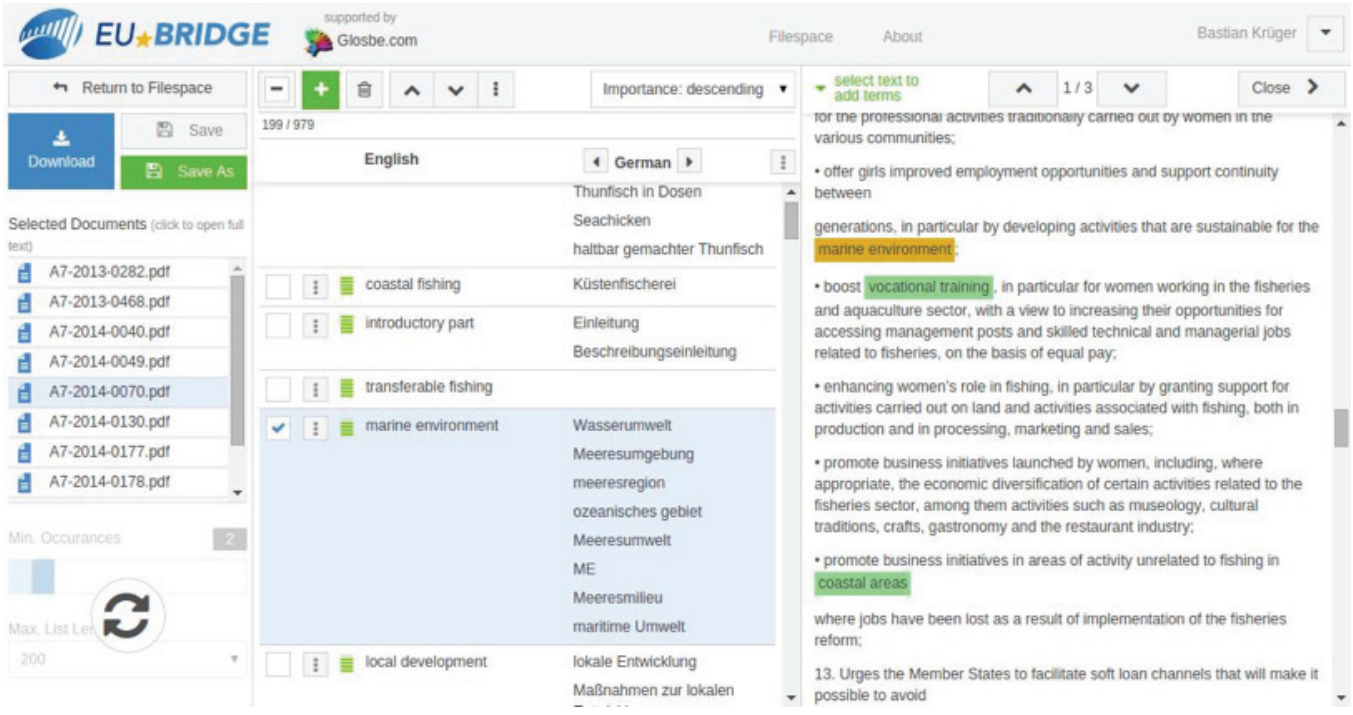


Figure 9 Interpreter Support Tool: Terminology list (with context)

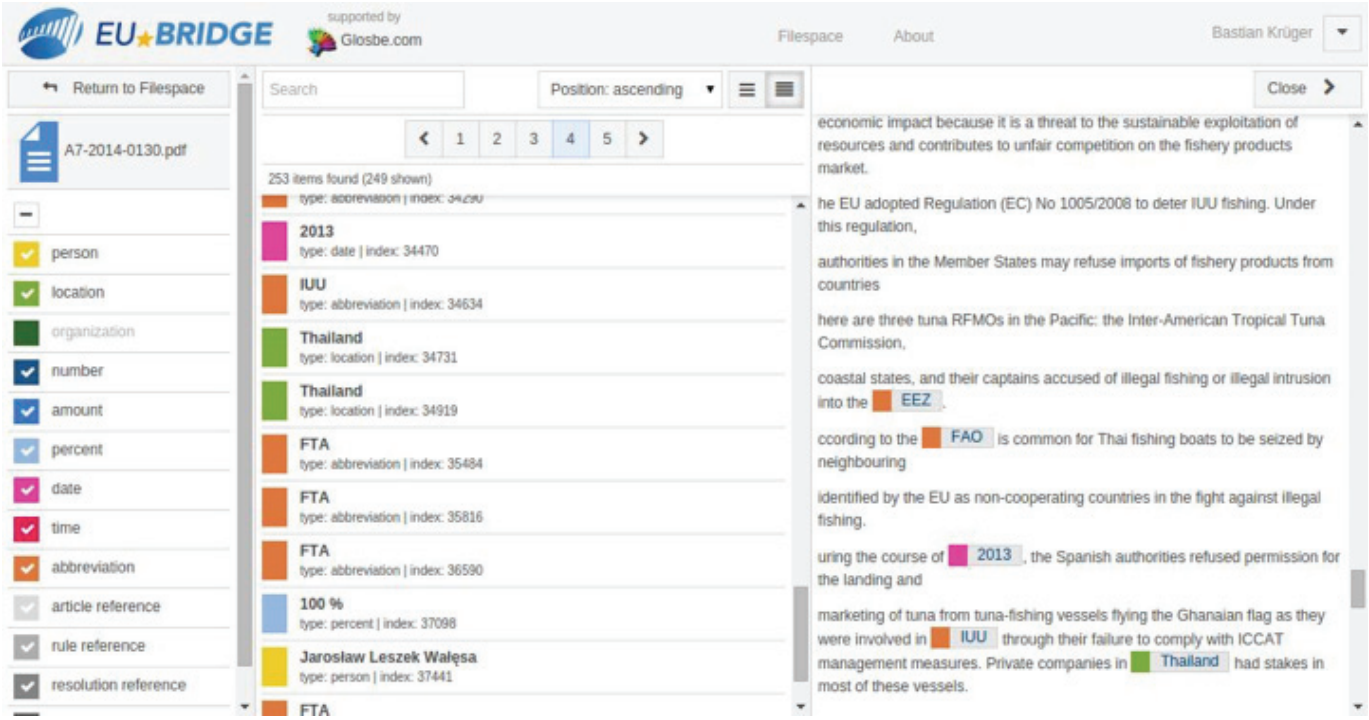


Figure 10 Interpreter Support Tool: Named entities

Named entity tagging

Inspired by the interpreters' statement that they find it difficult to remember and (therefore) accurately interpret numbers, names, etc., we have developed technology which provides helpful additional information, in particular, by highlighting and/or cross-referencing named entities. The web-based interpreter support tool includes a named entity recogniser, which can identify 13 types of NEs in a text and highlight them in different colours. The 13 types include 8 common NE types, such as location names and monetary amounts, and 5 types specific to the European Parliament, e.g., article references and resolution numbers.

Field test

Two rounds of field testing have been carried out in order to examine the usefulness of the tool to interpreters. Six interpreters from the parliament participated in the first round, and 18 in the second round.

The interpreters used the tool in their real preparation for a work assignment using EP documents.

76% of the volunteers used the tool from time to time, while 29% just used it for this field test.

The interpreters were asked both to provide feedback in the form of free-text and a fill-in questionnaire after each test round.

The questionnaire included 13 sections which cover the general impression on the tool as well as the opinions regarding its helpfulness, the interface, the quality of service, etc.

Figure 11 shows the overall satisfaction of the interpreters with the support tool. In the second round of the test more than 60% of the interpreters were satisfied or very satisfied with the final tool. The figure also shows the progress made in the tool between the first and second round of field tests.

From all the feedback that we were able to gather from the interpreters during the field test we can conclude that the system is cracking an important barrier by offering a solution good enough to be appreciated and to be found helpful by the target users.

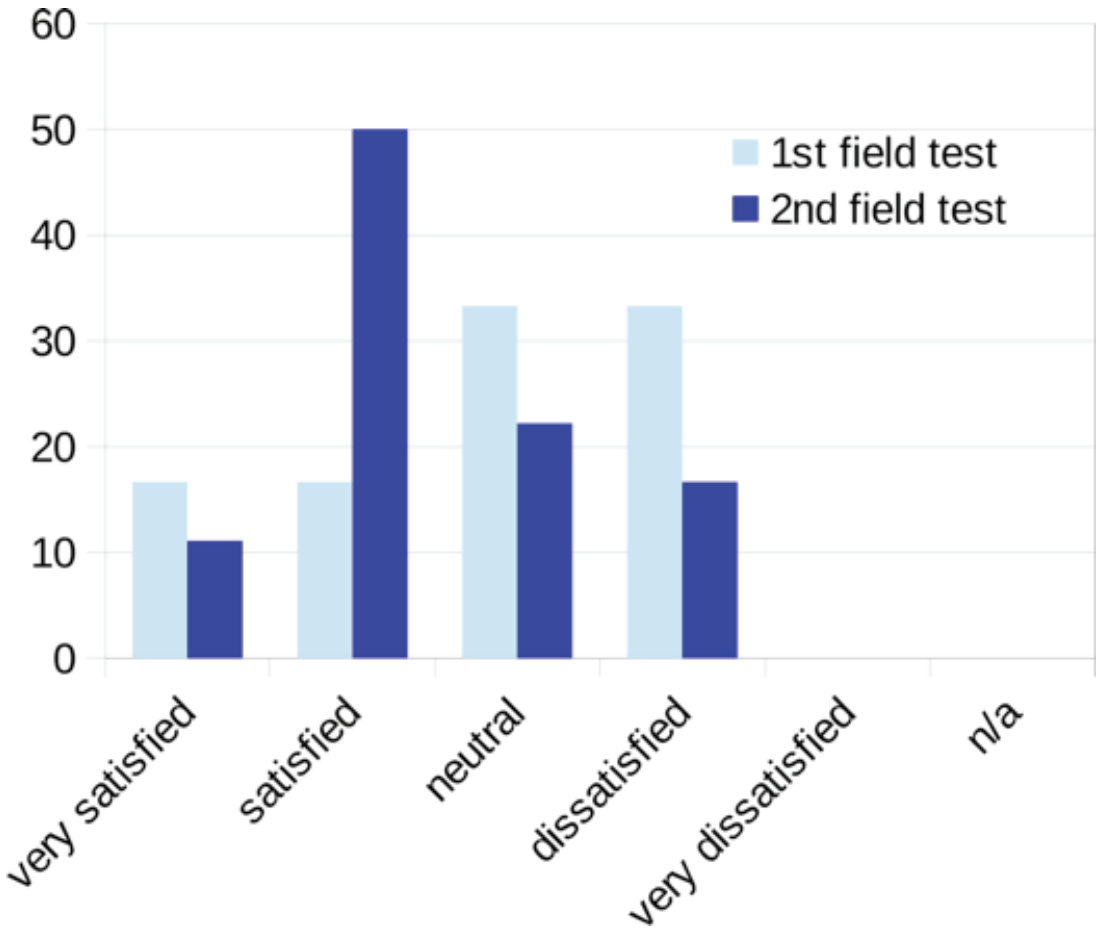


Figure 11 Overall satisfaction of the interpreters with the support tool

2.3.4. Simultaneous translation of webinars

The goal of this use case was to provide simultaneous translations of webinars within the platform Serenty developed by Andrexen. The system runs as an "OTT" (Over The Top) service

combined with a legacy WebEx / Adobe-Connect Webinar service. It allows the participant to enjoy a multilingual experience on top of the typical webinar functionality legacy webinar systems offer nowadays on the market at a time where these legacy systems do not provide any extra service to help people understand foreign-language webinars in their own languages.

System architecture

User access to the Serenty system is possible via PSTN (Public Switched Telephone Network, i.e. using legacy land lines) or Web interfaces (using a WebRTC compatible browser). These interfaces allow the capture of the participants' audio stream, whether they are using a phone or a PC with a microphone. These participants are the "Host", the "Host acting through a 3rd party system", and the "Attendees" (regular webinar participants).

The PSTN access allows the system to manage attendees and hosts using a regular phone line. The host of a webinar can speak directly into the phone or it can be an attendee using the phone access line of a "3rd party system" who listens to the webinar (hearing the original audio stream).

The Serenty WEB interface also called "The Serenty Client" is a multi-modal interface that presents webinar content (audio, text, slides, videos…) to the webinar attendees.

- First the host can use the audio interface of the client to speak. Audio capturing is performed directly from the browser, the user requiring only a browser and a microphone.

- The host can also use the WEB interface, which allows displaying other media output like slides and videos. Moreover, the WEB interface allows to display the ASR/MT text output resulted from the processing of the incoming audio streams, whatever it is origin, PSTN or WEB.

- The PSTN landline access is still usable and optional.

The „UC Server" is the server side of the Serenty system. It receives the audio, distributes the streamed audio to all the attendees, as well as to the ASR/MT systems (within the EU-BRIDGE cloud) in order to perform automatic transcription and translation.

The server in turn receives a text stream from the ASR/MT systems that is then distributed (streamed) to each corresponding attendee who is using the WEB interface.

An attendee using the PSTN access will have access to the original audio stream (like in a telephone conferencing system) using his/her landline. The same attendee can open a WebRTC browser in order to experience the multi-modality of the webinar, while cutting down the audio stream of its browser if his/her data connectivity is slow.

Figure 12 describes the way the audio stream-processing component of the "UC Server" is desi-

gned to function. Each webinar has an assigned conversation room or a "Webinar room" which is a virtual webinar space that connects the host to all other participants. This room performs the audio mix and data distribution and acts like a conferencing room.

The room can also be instructed to allow people to connect via regular phone lines (PSTN access) or to connect itself to an external line. The room would dial into the phone line access of a 3rd party webinar system and act as a listener. At this moment, the 3rd party webinar system becomes the room's "Host".

The "UC Server" specialized component acts as a virtual audio-silent participant. It will join the room, "listen" to what is being spoken and then stream that audio data to the ASR/MT workers. The component in turn receives the text transcription and translation of the spoken audio, which will then be posted as text messages to all participants in the "Webinar room" logged in using the Serenty "Web interface".

Using the components such as a "Virtual participant" allows flexibility in Unified Communication environment. Multiple such virtual participants could join and leave rooms as needed.

The Serenty prototype allows a webinar owner to start a webinar and for attendees to listen to the audio stream of the webinar. Meanwhile, the attendees receive the live transcription and Machine Translation (MT) of the audio stream using real-time ASR and MT engines.

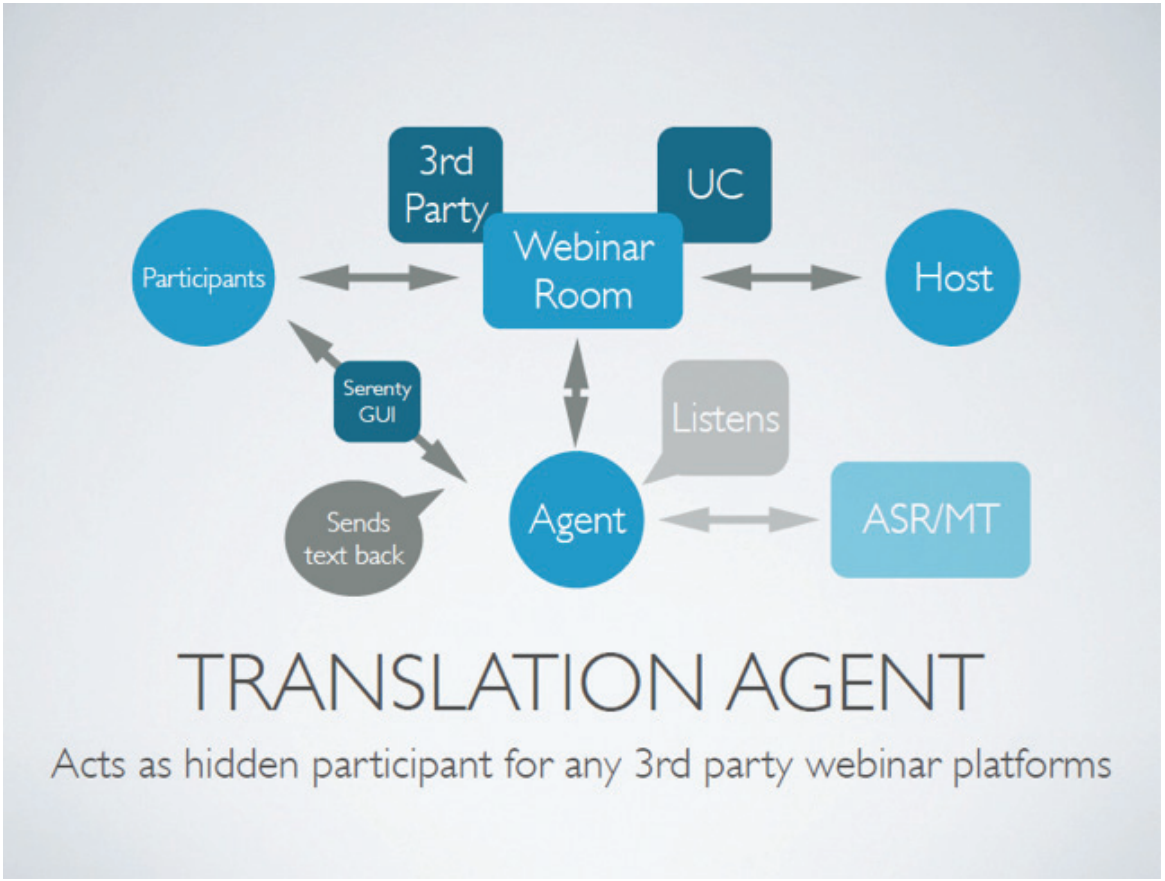


Figure 12 Reference architecture for the "OTT" ASR and MT support

Work flow management

The work flow related to the preparation and the streaming of a webinar has to be supported as much as possible to make it easy for the webinar presenter to comply with the needs related to the new functionality provided by Serenty.

First, the presenter has to be able to manage their own presentation and to be able to upload his/her own slides/content.

Second, the transcription and the translation services provided by Serenty can make use of this uploaded content in order to adapt the underlying technologies to the specific "domain" presented during the webinar.

In order to manage a webinar, the presenter has to upload its presentation to Serenty.

Figure 13 and Figure 14 illustrate the design of the document-uploader which has been developed for Serenty in order to illustrate this process. Typically, when Serenty goes into the market, this process will be supported by a 3rd party building on Serenty (e.g., Adobe).

The webinar host can select image files to be uploaded to the webinar room and then re-order the files as desired in preparation for the webinar (Figure 2 3).

Once the webinar has started, the host is shown two controls represented as arrows to select the next or previous slide (Figure 14). Once the command is issued, all the webinar participants connected to the system and using the web interface see the matching slide. The participants cannot control the slide currently being displayed.

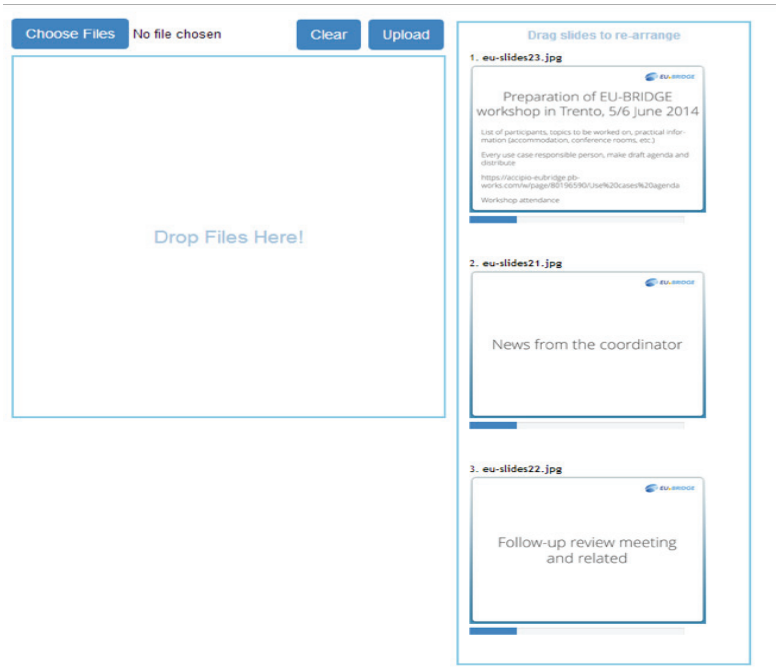


Figure 13 Document Uploader: User Interface for slide-upload and slide-ordering

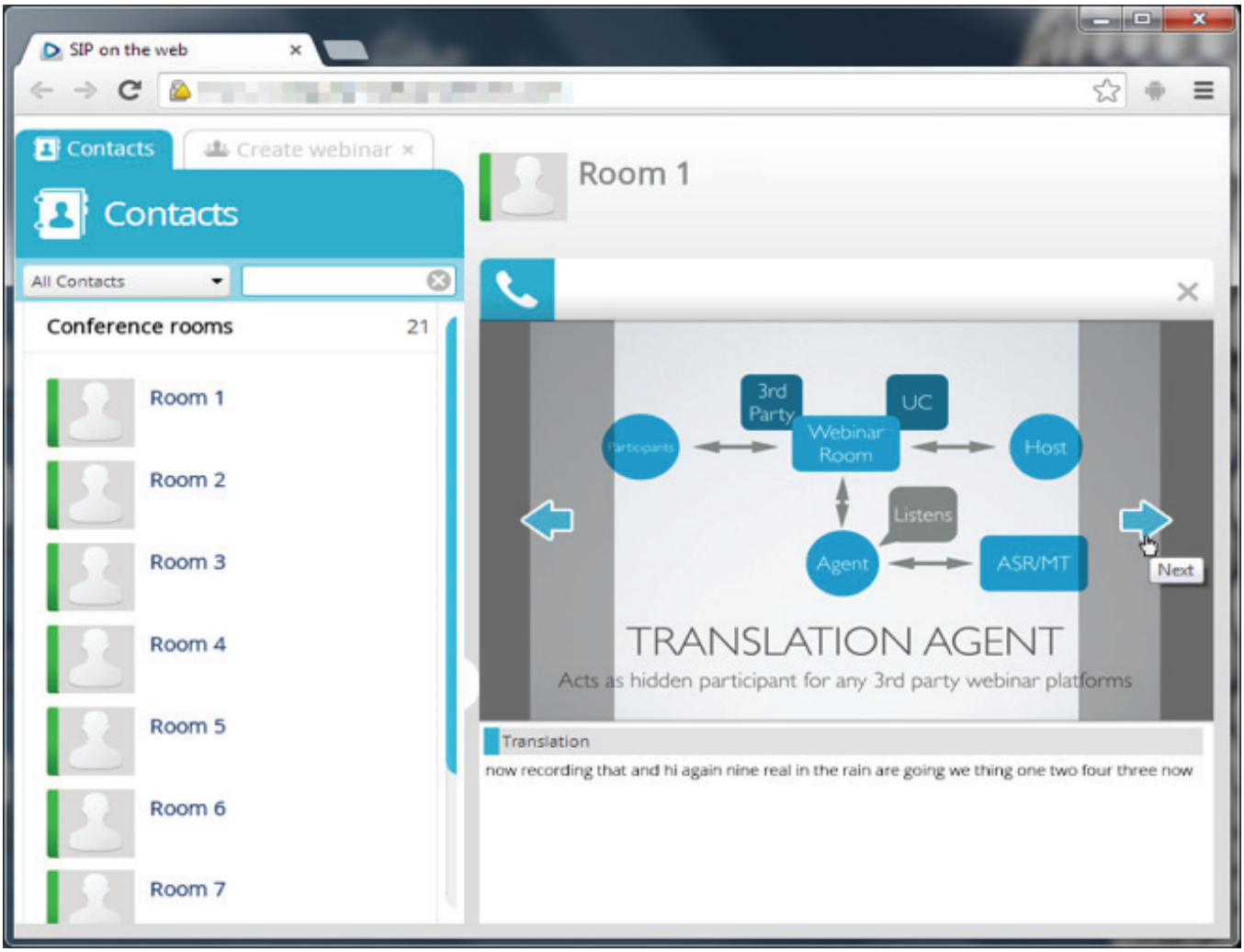


Figure 14 The host can use controls to move the set of slides forward or backwards. The selected slide is being displayed in each attendee's interface.

On-demand worker adaptation

Webinars often employ a specialized terminology, where the speaker will frequently use terms that are uncommon and usually not observed in other contexts. Examples are company names, business terms, and personal names. Correctly recognizing such terms is especially important, because failing to do so may cause the complete content to be misunderstood. Because it is impossible to build a speech recognizer that contains all possible specialized terms in its vocabulary in advance, we created adapted speech recognition systems for each webinar by exploiting slides, additional text materials, and weblinks (e.g. to the corporate website) that are provided by the speaker in advance. The adaptation module automatically extracts missing terminology from these text sources, including a full crawl of the corporate website, if available. It uses this data to adapt the speech recognizer's dictionary (see Figure 1-5).

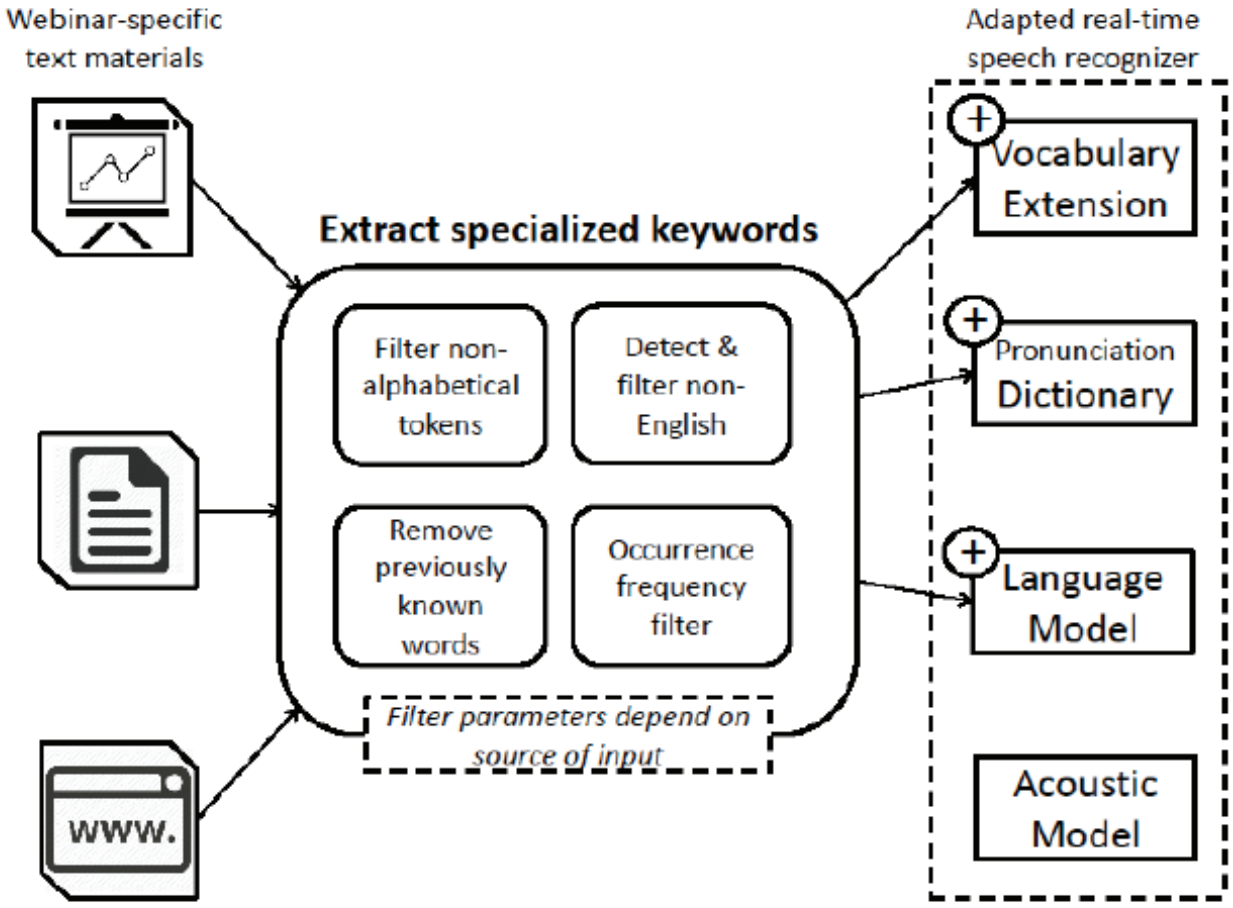


Figure 15 Creation of adapted worker process

The adaptation framework takes a list of documents in PDF or text format, as well as a list of web-links as an input, and from there on proceeds in a fully automated manner. As a first step, the given web-pages are downloaded, including any additional documents which are found by following links. Text is extracted from the web pages and PDF documents. As a second step, it is important to pre-process and filter the obtained text material. Text extraction from PDF and web documents always produces some noise that must be filtered out. Web pages in a language different from the webinar source language (here: English) might be found as a link and must be discarded. We established a set of filtering rules that uses information such as the number of times a word has been observed in the adaptation text material, predicted language, source of the data, capitalization, etc. The filtering is rather strict, because adding nonsense terms to the speech recognizer's dictionary can harm its performance.

From the filtered text, for all words that were previously unknown, pronunciations are automatically derived using a grapheme-to-phoneme converter. In some cases, multiple candidate pronunciations may be derived, for example if it is unclear whether to pronounce an acronym as an ordinary word or as a sequence of letters. Finally, the terms are mapped to appropriate pre-selected known words for which the language model can estimate accurate probabilities (unlike the newly added terms, for which no or only insufficient language model context is known). Words were pre-selected that appear fairly commonly, accounting for the fact that they have a

high chance of actually appearing in the webinar, and can occur with a variety of part-of-speech roles, since the nature and usage of the added word is unknown.

Field Test

The field test for the webinar use case was run in Paris with 10 French students listening to 12 webinars presented in English by 5 Americans and 4 non-native English speaking Europeans totalling 4 hours and 39 minutes of speech. The webinars were automatically translated into French. The subjects of the webinar where mainly business and marketing related, whereas one covered a physics theme and another one an IT theme. However, these two technical webinars were done at a very high level.

Each webinar was linked to a dedicated worker that used a language model adapted to its content (slides, website). The platform of Andrexen presented the video or slide show of each webinar together with the English transcription and the French translation of the input signal. The transcription and the translation where presented to the test users in real time with exactly the delays given by the system, typically one to two seconds for the real-time decoding and a "sentence" (as defined by the transcription engine) for the translation.

The users were all French test users in the age between 18 and 25 years old, 5 females and 5 males. The webinars were in English and were automatically transcribed and translated into French. The users have been asked to qualify their proficiency of the English language (in written, speaking, reading, listening). This allowed us to build two categories of users, a group of 5 persons having a good proficiency in English (2 females and 3 males) and a group having an average proficiency in English (3 females and 2 males).

The test users have been asked to view about 3:15 hours of webinars from the total of 4:39 hours available. We have built four thematic groups: two thematic groups were containing pure business themes, named Bus1 and Bus2 and the other two thematic groups contained a mix of business and IT themes, which were called BusIT1 and BusIT2. Test users have chosen to follow one thematic group according to their interest.

Each student had to fill a form for each of the webinars they viewed as well as a form at the end of the field test in order to give an overall feedback on their experience with the unified communication platform.

In order to draw some correlation between notes given by test users to their perception on viewing each webinar with the platform and the quality of the transcription/translation provided by the system, we have grouped the 12 webinars into 4 distinct categories depending on quality. The four groups are called grade A to grade D as described.

First we are looking at how transcription has been perceived by test users relatively to the grade of the webinar. And we have split test users in 2 groups, those thinking having a good proficiency in English (good understanding), and those thinking having an average proficiency (average

understanding) in English.

One important question in the feedback form was about the help the system can provide test users with understanding better each webinar. What was not surprising is the drop in „comprehension help“ noted by test users having a good proficiency in English while going from grade B to grade C webinars (see Figure 16) , dropping the grade given on „the system helps me understand the webinar better“ by 1.5 MOS points. This means that for these test users (having a good English proficiency), there was a clear cut for webinars of grade C: for these test users, the quality of the transcription and translation was too bad to be considered helpful.

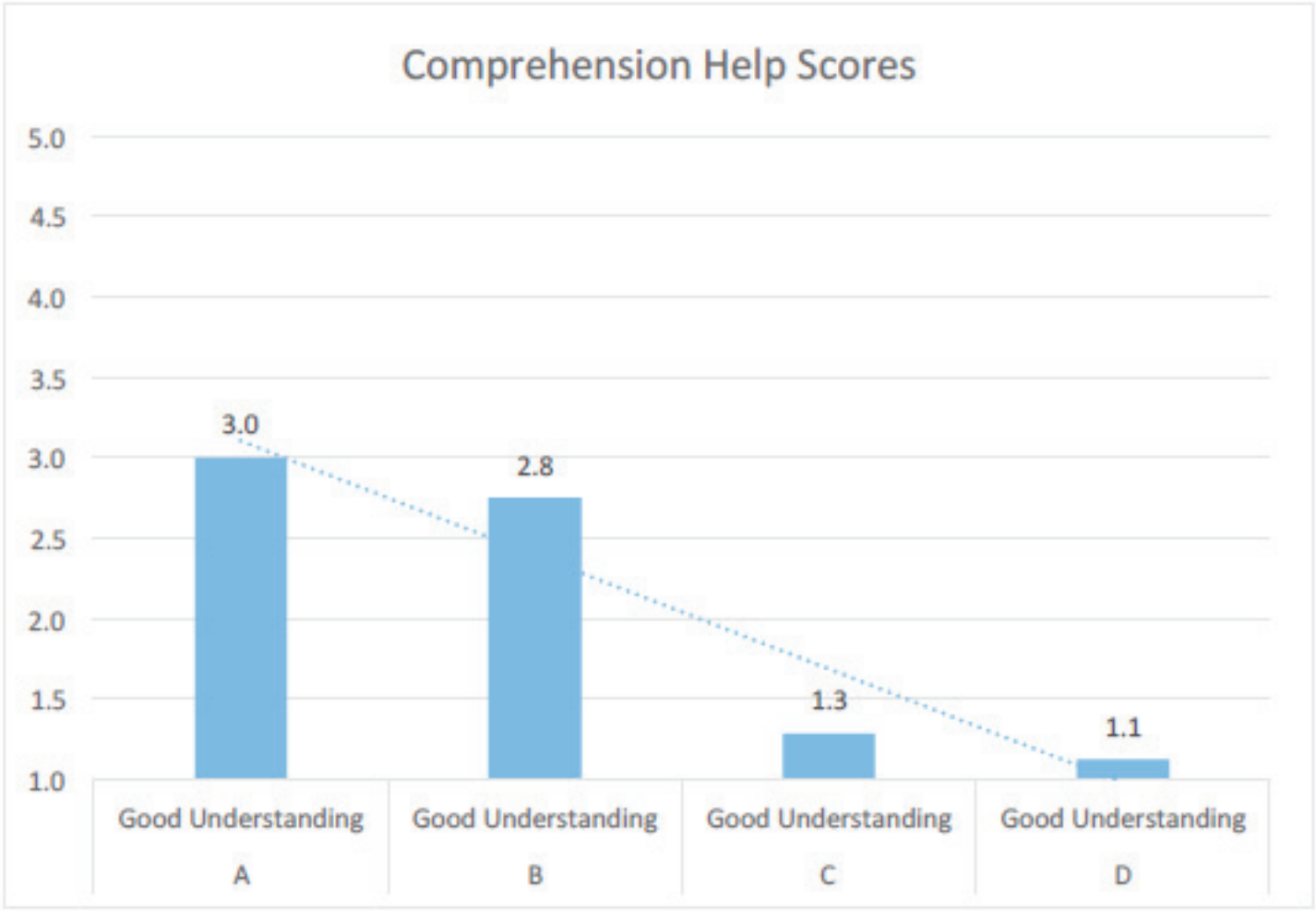


Figure 16 Comprehension help perceived by test users having a good understanding of English for A-grade to D-grade webinars

But what is very interesting is that this drop in „comprehension help“ is not observed at all with test users having only an average proficiency (average understanding) in English (as opposed to those having a good understanding of English). Even for grade D webinars (see Figure~\ ref{img:web:compreh_averageEnglish}) the translation helped.

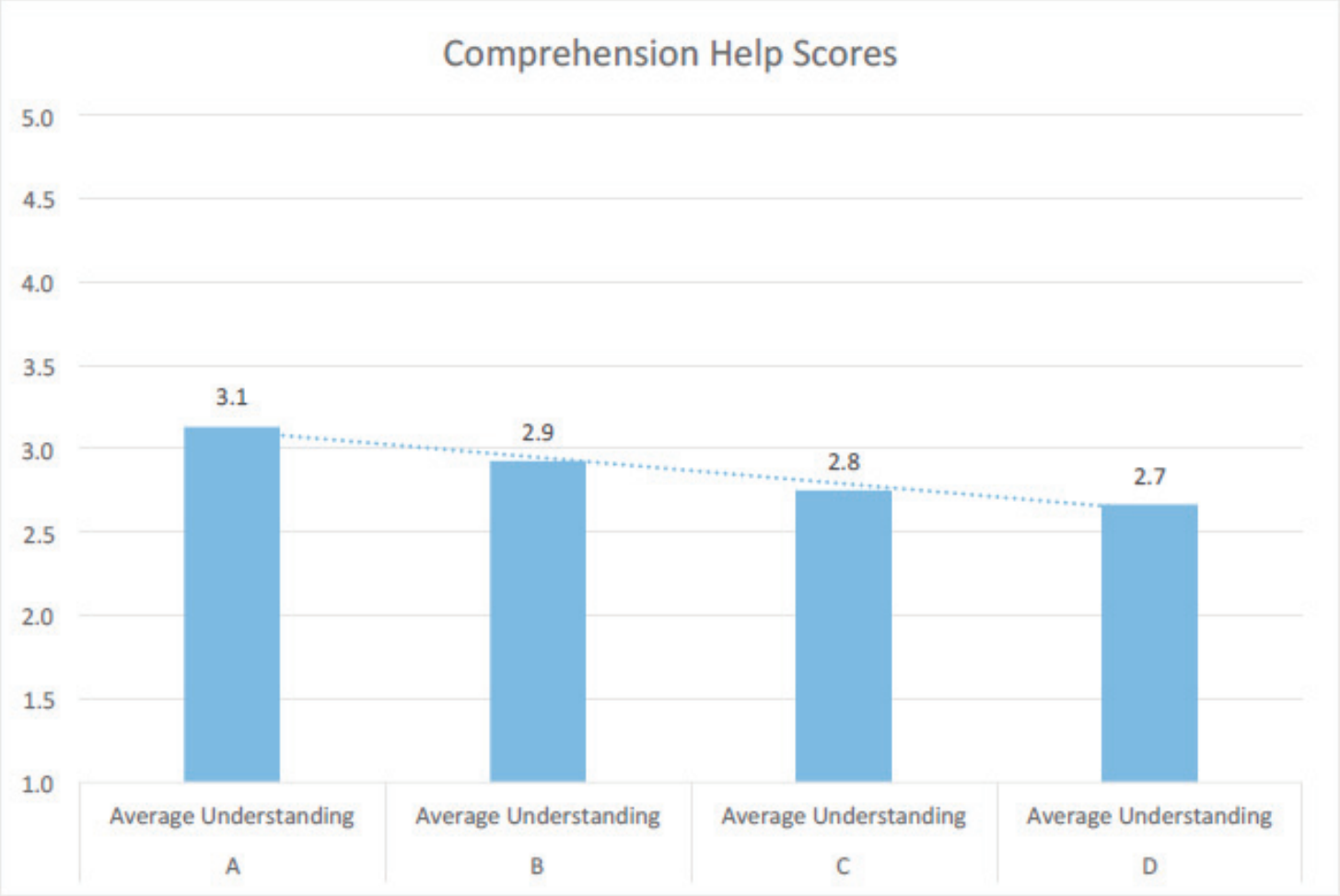


Figure 17 Comprehension help perceived by test users having a good understanding of English for A-grade to D-grade webinars

After having seen all the webinars (from grade A to grade D), the test users with average proficiency in English, consider the system somewhat useful, with an average score of 3, a score that is higher than the one given by test users with good proficiency (score 2.4).
Test users have been asked to make their own comments on their experience using the system. While quantification of comments is not possible, comments give a very good insight in the overall impression left over by a system to the testers.
Based on these comments, we are encouraged to enhance the system highlighting named entities in a sentence. We may even think about using the ASR scores and define for example two score-thresholds: a word or phrase scored above the higher threshold could be written in a larger font than the words or phrases scored between the two thresholds, and presenting in a faded grey those words or phrases scored below the lowest threshold.
Last, the time delay between sound and text appearance (especially that of translation which has an inherent longer delay than that of transcription) has been perceived by test subjects as annoying.

3. Potential impact

The project has a potential for impact in three main areas:

- Scientific sector
- Language technology industry
- Multilingual digital market

In the following we will show what this potential impact for those three sectors looks like.

3.1. Impact on the scientific sector

In the scientific sector the project helps to keep European research institutions in the top field of successful speech and machine translation research organisations worldwide.

The results of the external evaluation campaigns in the field of speech translation technologies that the EU-BRIDGE partners participated in has demonstrated that the partners from EU-BRIDGE belong to the best research groups in this field worldwide, either matching or often even surpassing the performance of other successful groups from Asia or North America.

Through the research in this project the successful scientific development as previously supported through European or national projects such as TC-STAR, EuroMatrix and Quaero was continued, giving the European research community momentum to continue their work.

As we know, speech translation technology has passed the mark of being useful in real-world applications. However it is still a very error-prone technology due to the inherently very high complexity of the problem to be solved. Thus the scientific development is far from being over and will see increasing progress over the next years.

The results of EU-BRIDGE will fundamentally support this development. The results of the research have been widely publicized in international conferences and are thus available to the whole research community. Further, the consortium is planning to organise a journal special issue on the topic of speech translation which will combine the research results from EU-BRIDGE with recent results from the whole international community.

Important technological developments over the course of the project, such as the intensified use of neural networks in the models of speech translation systems, were made available as open source toolkits, and are thus available to the research community outside of EU-BRIDGE, supporting the efforts of other European research groups as well.

The multilingual data collected in this project, such as the Euronews corpus, will support speech

translation research in Europe and will provide it with the necessary training and test data. This impact is already happening as the training data has already been released within the scope of the IWSLT evaluation campaign and is thus now available to interested researchers.

Under EU-BRIDGE the IWSLT evaluation campaign was able to establish itself further as the predominant international speech translation evaluation attracting a large number of participants from the top laboratories of the world. IWSLT also acted as valuable data source for research groups, e.g., by distributing Euronews training data and special releases of the WIT3 corpus for evaluation purposes.

3.2. Impact on language technology industry

The world wide market for speech translation technologies has been increasing and developing greatly over the last decade, with speech translation technologies now becoming able to provide useful services to customers with reasonable performance. Applications range from obvious translation scenarios for human-to-human communication in civilian and military settings, over less visible applications, such as providing translations for media content or the translation of manuals and other documentations, to such unnoticed translation needs as the localization of software. Behind these translation needs thrive large translation industries consisting of large, European and non-European companies, but also many European SMEs acting in the European market. Backed by large investments into research and development, the Google Translate service, for example, has gained high visibility and is now widely used by people to access digital content across languages. In order for the European industry to stay competitive in this market it needs to deploy state-of-the art technology, by actively incorporating the latest research results from universities and research institutes into their products, and actively developing them.

The service architecture developed within EU-BRIDGE has the potential for keeping Europe competitive . It offers an interface that allows connecting providers of basic technologies and engines such as automatic speech recognition and machine translation systems with application developers that require spoken language processing technologies. This lowers the barrier for players and developers in the field of speech translation technology when it comes to entering the market and offering their services and research results commercially. It also closes the gap between the pure technology development in speech translation and the actual applications that will incorporate these technologies. This especially important as, like described above, the applications for speech translation technology often go beyond mere dictation and translation, bur rather are hidden deeper in specific work flows.

The system architecture that makes all this possible will be marketed by PerVoice SpA. PerVoice has the necessary experience for this task and a good standing on the European language technology market, thus having a good starting point for a successful commercialisation of that architecture.

Today's best performing systems use statistical models that are trained on annotated data. Access or creation of this data has become so expensive, especially since it needs to be performed

multiple times for the many languages addressed by technology. In that sense data has become what crude oil is for modern industries. Access to and control of this resource determine the success that an industry will have. The techniques developed within EU-BRIDGE will enable the European language technology sector to process and acquire data more cheaply and make better use of it, thus giving it a competitive edge or narrow the gap to large data collecting companies such as Google.

Further EU-BRIDGE's impact on the language industry technology for EU-BRIDGE has been already demonstrated through several mergers and acquisitions of industry partners within EU-BRIDGE. Especially important for the European market are the following two acquisitions:

- Acquisition of Pervoice by the Almawave group
- Acquisition of Red Bee Media by Ericsson

Almawave is part of the AlmavivA group and its technological innovation company. Almawave is, on its own accord, one of the leading Italian players globally operating in CRM, Big Data Knowledge Management, and Customer experience sectors. Through the acquisition of PerVoice Almawave completed its portfolio of technologies and solutions integrating voice, social media channels, semantics, advanced statistical algorithms, real-time conceptual business intelligence solutions and new-generation business process management models.
Ericsson's acquisition of Red Bee Media supported its strategy to grow in the broadcast services market and to take advantage of Red Bee Media's technology and services leadership to help broadcasters and content owners address the convergence of video and mobility. This will further strengthen Ericsson's broadcast services business, which was started in 2007 and expanded in 2012 with the acquisition of Technicolor's Broadcast Services Division.

3.3. Impact on the multilingual digital market

The results of EU-BRIDGE have the potential for transferring speech translation systems into the European market, in order to increase their use and to utilize them to lower costs and improve efficiency of speech translation. In this way the project will foster the use of speech translation systems to give Europe's citizens and businesses the possibility to communicate across languages, and to access multilingual and multimedia content across the language border through those systems.

We anticipate that this will lead to:

- An improved European competitive position in a multilingual digital market through the provision of better products and services to citizens and businesses.
- Cooperation and exchanges between European and national efforts, closer dialogue and partnership between research and industry, better understanding of user requirements, thus stimulating innovation and technology uptake.

The increased use of speech translation technology by participants market is of high importance, as today's world market is a largely global market in which information is shared across languages, and goods flow across the globe. The market interaction is organized in a digital way. At the same time digital goods, such as audio/video content in many different forms has grown in importance. In order for a company to be able to stay competitive in this digital market, but also even in the non-digital market, it needs to be able to cater the different language needs of its customers, by localizing its products, but also accompanying information, such as documentation, advertisement etc. Dealing with the issues in multilingualism that arise from these requirements in a cost and time efficient way can be decisive for the success of placing one's own products on the market, either in digital or physical form.

The results of EU-BRIDGE will give the European players in this multilingual market and the multilingual European market as such a competitive edge, as it will help to overcome the language barrier that currently inhibits the flow of goods, services and information across national and thus very often language boundaries.

The results that impact the multilingual market fall into two categories: direct impact through the use cases developed within EU-BRIDGE for concrete multilingual market needs and by inspiring new applications and enhanced applications that incorporate speech translation technology.

The direct impact through the use cases varies from use case to use case:

- **Captioning and translation of TV broadcasts:** The field tests in incorporating speech transcription technology into the captioning process has created large interests in the broadcast community and has brought awareness to the possibilities when speech technology is applied in a correct way. It is thus expected that the market will soon see more automated and semi-automatic captioning technology, thus lowering the cost for captioning and in turn leading to more content being captioned. The results of the speech translation demonstrations on Skynews have also ignited interest, so that in the midterm future we expect to see more activities in multilingual subtitles, thus making more multimedia content available across languages.

- **Simultaneous translation of university lectures:** The deployment of the simultaneous lecture translator at KIT has demonstrated the possibilities of teaching across language barriers. This will increase KIT's attractiveness to foreign students, making the KIT student and researcher body more international, leading to more cultural, scientific and technological exchange in Europe. Also, several side trails have been opened through lecture translator, such as making recordings of lectures searchable (across languages), or making archives, such as the audio/video archive at the KIT library (DiVA) searchable. The potential of the future impact is underlined by pick-up through entities outside of EU-BRIDGE already initiated in the last phases of the project. E.g., a professor at the University of Kassel is using the EU-BRIDGE transcription service to make recordings of his lectures searchable, and together with the University of Heidelberg and other partners, a new research proposal

has been initiated which, if successful, will lead to a new lecture translation installation at the University of Heidelberg.

- **Interpreter support for interpreters at the European Parliament:** The successful field test of the interpreter support tool will probably lead to an established use of the tool by the interpreters, with the parliament willing to pay for the maintenance and possibly extension of the tool. This will lead to improved working conditions for the interpreters at the parliament. The tool has the potential for also being marketed to other interpreters. Further, the tool has ignited interest in extending the EU-BRIDGE technology to other application scenarios, where currently human interpreters cannot be used, e.g., for parliament member events in their respective districts.

- **Simultaneous translation of webinars:** The successful test of the translation technology for webinars will have an impact on the market for webinars. Users will now have access to contents that before was no within reach for them due to the language barrier. For those with imperfect command of the language of the webinars that they are following, their understanding of the content will improve due to the assistance from the translation system. This will work towards a truly integrated market for webinars in Europe that is not inhibited by the presence of many languages in Europe.

With respect to the indirect effects, the successful implementation of the four use cases above and the creation of the EU-BRIDGE service infrastructure for speech technologies is expected to inspire additional, new applications that have not been thought of by the consortium members before. The consortium is now frequently being contacted by entities outside of the consortium for advise and help on concrete needs for speech technologies and making them available for their applications.

Scenarios discussed so far include:

- Transcriptions and translations of large archives of lectures such as videolectures.net, the DiVA archive or individual lecture recordings

- The use of speech transcription technology for natural language programming research projects

- The use of speech transcription and translation technology for cross-lingual information retrieval

- The use of machine translation technology for translating product descriptions in an online shop

- The use of speech transcription technology for recordings of market study interviews

- The use of speech translation and transcription technology for transcribing and translating

MOOCs

In summary, the project has brought together technology developers and market players with specific application needs for speech translation, implementing four use cases that are expected to develop further after the end of the project and finding a direct entrance into the European digital market. Further, within the consortium, the successful conducting of the project has brought awareness to decision makers within the industry partners of the possibilities that lie within the utilisation of speech technology. On top of that, the awareness of these possibilities has also spread outside of the consortium to other partners, helping to transform the products on Europe's digital market into truly multilingual products that bridge across the language barrier, giving European businesses to take advantage of the multilingual nature of Europe.

Website

The website of the project can be found at http://www.eu-bridge.eu where the publications, dissemination materials and other information about the project can be found. Further, the website gives information for interested application developers and customers that want to realise their own applications with the help of the EU-BRIDGE service infrastructure or who want to buy one of the services already implemented within EU-BRIDGE.

4. Contact addresses of the consortium

Primary Partner Contact	Website
Prof. Alex Waibel Institute for Anthropomatics and Robotics Karlsruhe Institute of Technology Germany	http://isl.anthropomatik.kit.edu/english/21_74.php
Dr. Volker Steinbiss Accipio Projects GmbH Aachen, Germany	http://accipio-projects.eu/team/34-2/?lang=de
Grégoire Boutonnet Andrexen Levallois-Perret, France	http://www.andrexen.com
Marcello Federico Fondazione Bruno Kessler Trento, Italy	http://hlt.fbk.eu/people/profile/federico
Dekai Wu The Hong Kong University Science Technology Hong Kong	http://www.cs.ust.hk/~dekai/
Alessandro Tescari PerVoice SpA Trento, Italy	http://www.pervoice.com/
Krzysztof Marasek Polish Japanese Institute of Information Technology Warsaw, Poland	http://www.pja.edu.pl/en/

Juliet Gauthier Red Bee Media London, United Kingdom	http://www.redbeemedia.com/
Prof. Dr.-Ing. Hermann Ney Chair of Computer Science 6 RWTH Aachen University Germany	http://www-i6.informatik.rwth-aachen.de/web/Staff/ney/
Steve Renals The Centre for Speech Technology Research The University of Edinburgh United Kingdom	http://www.cstr.ed.ac.uk/ssi/people/srenals.html

Use and dissemination of foreground

5. Section A (public)

Use and dissemination of Foreground

In a time when a great variety of information and especially scientific information is published, a project has to ensure, that it communicates the results to the different target groups in the most appropriate way. It was EU-BRIDGE's purpose to incorporate its key messages, addressing all target audiences, exploring new instruments and dissemination channels and ensuring cost-effectiveness. Therefore we pursued an active dissemination, communication and use strategy in order to have the highest possible impact on the target groups like the scientific community, the general public, decision makers including the media, and the industry.

Therefore the EU-BRIDGE project allocated a major effort to disseminate the contributions of the program along several dimensions. These dimensions were designed to maximize the benefit of the publicly funded undertaking to the society at large (as opposed to only a narrow community, e.g. the scientific community). The project also pursued technology transfer activities along several dimensions, so as to maximize the likelihood of takeoff and adoption of the ideas and methods.

Every public relations campaign should start with a communication plan. The plan should be dynamic and change as the situation/project's development change but a well-thought-out communications plan is essential for a successful project. EU-BRIDGE's communication plan defined its target groups, the key messages, goals and objectives, channels and the evaluation.

Target Groups and Key Messages

EU-BRIDGE defined the following six target groups for its communication:

The purpose was to address all target audiences, exploring up-to-date instruments and dissemination channels and ensuring cost-effectiveness. Therefore, the communication plan pursued an active dissemination, communication and use strategy in order to have the highest possible impact on the different target groups.

After having defined our target audience we had established key messages pertaining to each of our key target groups. Of course there are some overlaps in the messages, but a key message is essentially what we want our audience to "take away" from EU-BRIDGE:

Scientific community:
Bringing the new scientific results and innovative achievements out of the laboratory and the project into the scientific community was one of major goals in the field of exploitation of project results. Presenting the EU-BRIDGE consortium a highly qualified networking partner was another main goal.

General public:
The EU-BRIDGE project considered it essential to communicate its results to and the benefit for the general public. Answering the question: "What is in it for me?" involved communicating the goals and activities of the project in lay-men's terms and to reach out to the public and to public bodies.

Decision makers and Politicians:
"EU-BRIDGE-funding is of great value": This message was communicated to decision makers and politicians with special emphasis to decision makers in the European Union.

Industry:
"EU-BRIDGE can sell": We showed the impact on industry, innovations which could be integra-

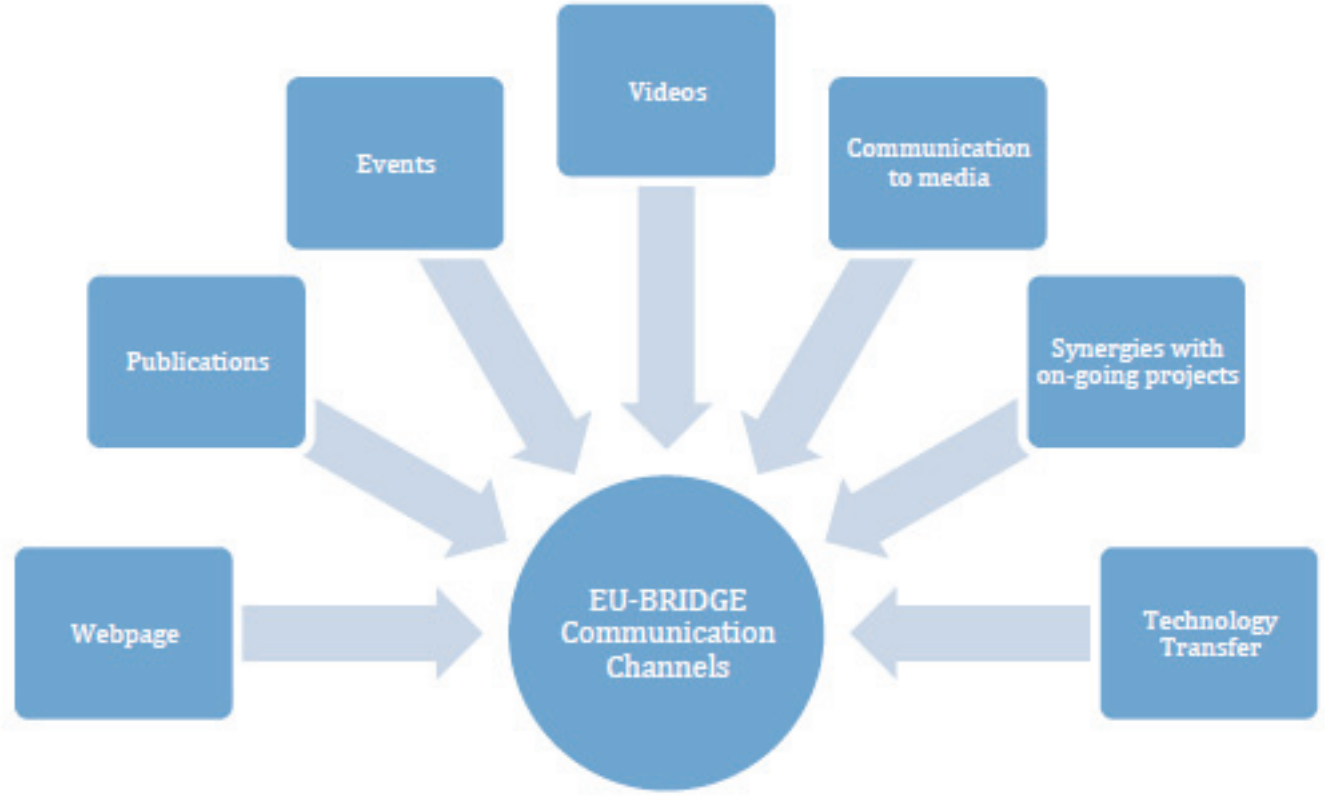
ted in industry products, exploitable knowledge and application sectors.

Media:
Media was one of our target groups but was also a vehicle to communicate EU-BRIDGE to the above mentioned groups. Target media audience covered local, national, European, and international media, scientific journals, internet press as well as television and radio.

EU-BRIDGE project partners:
Perfect project communication needs all partners on board. Beside the external communication special emphasis was put on transparent and continuous communication within the project. EU-BRIDGE ensured that partners were kept fully informed about any development within the project.

The Communication Channels

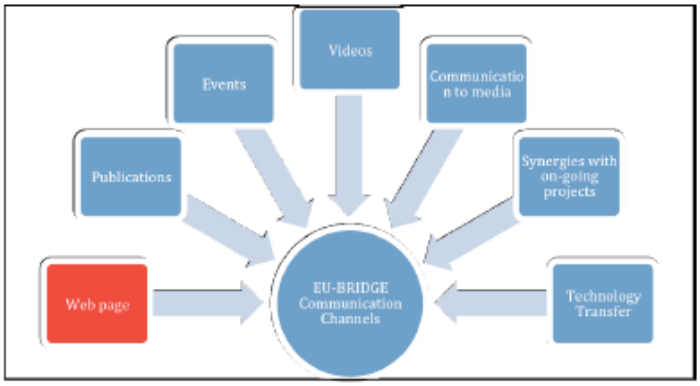
Communication can be split into two parts: the message/content and the channel it is transmitted on. The messages can be communicated over a number of channels. As the different communication channels have their special strengths and weaknesses it is important to communicate via diverse tools. For the defined target groups of EU-BRIDGE with their different key messages we had defined the following dedicated communication channels:



As every single communication channel addressed different target groups there were overlaps in the tools. As a matter of course e.g. the web-site, publications or events were dedicated to all groups. That's why we had set up e.g. on the web-site special areas dedicated to the different groups, or organise special events for special groups.

Webpage, Social Media Networks, Intranet, Mailing Lists

In today's visual world, an attractive, modern and up-to-date web page is one of the most important tools to inform about the project. Due to this fact, we had launched the project web page at the first day of the project (on February 01, 2012) under: www.eu-bridge.eu. We had also reserved the domains .de, .net, .org, and .com. The webpage is hosted at the computation centre of KIT in order to ensure persistency of the website after the end of the project.



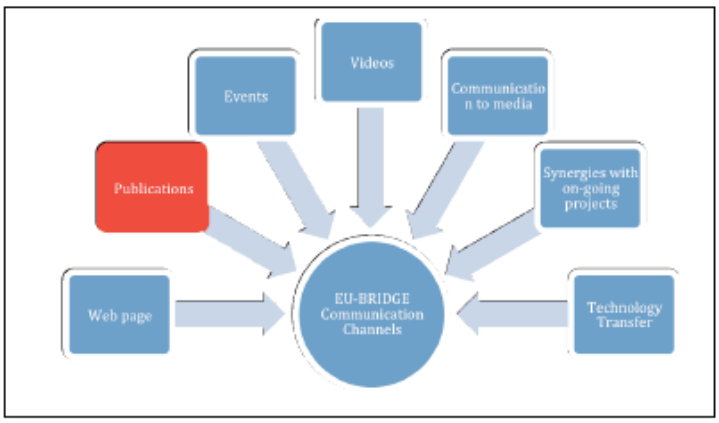
EU-BRIDGE developed a webpage that communicates the results of the project in simple and engaging terms. Via the webpage, the visitors can easily learn what "EU-BRIDGE" is about. Some parts of the webpage are of special interest to the scientific community, like publications or research and technology. The webpage offers information on: The project itself, news, partners, research, publications, demos & videos, related projects, press, contact and a link to EU-BRIDGE's intranet. The web page contains the following statement: „The work leading to these results has received funding from the European Union under grant agreement n° 287658". We have ensured web accessibility. The main language of the web page is English. In order to provide interested individuals with the opportunity to receive news, updates on the project, they could subscribe to a mailing list through which news on the project (incl. newsletter) were distributed.

EU-BRRIDGE also participated in social media networks like YouTube. We profited from already established channels like the KIT-YouTube channel on YouTube. This allowed us to use the already existing competence and to profit from the networks, which have been already established.

Transparent and continuous communication ensured that EU-BRIDGE-partners were kept fully informed about any development within the project. Day-to-day communication within the consortium was carried out mainly by e-mail and file sharing via the EU-BRIDGE intranet. In order to minimize the amount of e-mails and to bring the fitting information to the appropriate recipient, we had set up diverse mailing lists, all in all we had set up 34 different ones. The mailing lists are hosted at KIT (which ensures persistency, also after the end of the project) and are maintained by ACCIPIO projects.

Publications

Dedicated publications were published for use of the different target groups:

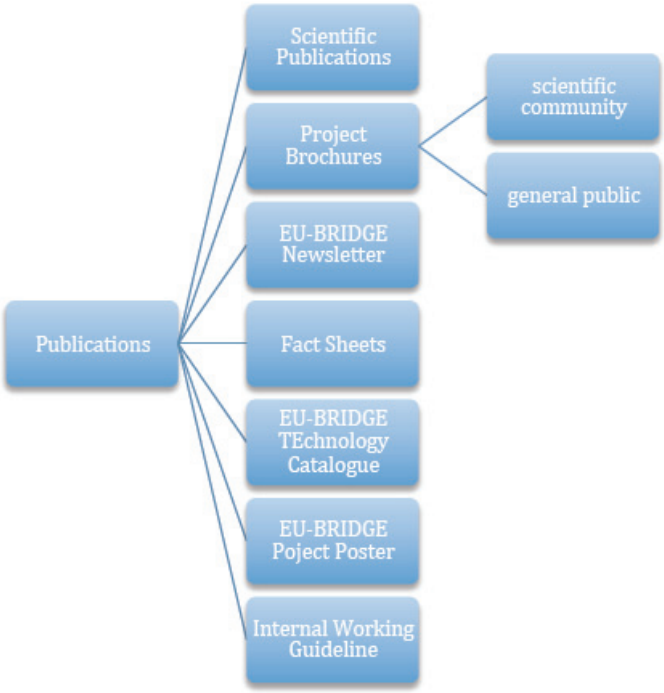


Scientific Publications

Project members produced publications at journals and conference proceedings. We provided an internal form to communicate the publications to the dissemination coordinator in order to maintain a proper list of publications and to upload them immediately to the website. The form was uploaded to the EU-BRIDGE intranet and also communicated via the "Internal Working Guideline".

Two Project Brochures to gain first interest

We compiled two different brochures: One for the scientific community and scientific media. A second one was dedicated to the general public and general media. The latter was printed in English, an online version was published in French, German and Italian. Following the progress which was made during the lifetime of the project the brochures were updated for three times. The brochures were distributed at events, sent out by mail and could have been downloaded on the website.



EU-Bridge Newsletter

The EU-Bridge newsletter informed on news and important events. The newsletter appeared two imes per year and was published on the website and was sent out by e-mail to different mailing-lists.

Fact Sheets

A brief project fact sheet is available since the beginning of the project. It will be maintained and updated until the end of the project. The project fact sheet was made available for the use of the European Commission for its own dissemination and awareness activities right at the beginning of the project and uploaded to www.eu-bridge.eu

Recto-verso Fact Sheets

This series of fact sheets reported on the developed technologies (including exploitable know-

ledge, application sectors, technical requirements, terms of availability and IPR protection). These fact sheets were the basis of the EU-BRIDGE technology catalogue (see below).

EU-BRIDGE Technology Catalogue

This catalogue described the technologies developed in the project in detail. This catalogue included information on the exploitable knowledge, application sectors, technical requirements, terms of availability and IPR protection. It served as a practical advice how to use the developed techniques. Mainly for the industry, it was distributed at the EU-BRIDGE technology day, via the partners and as download on the website.

EU-BRIDGE Project Poster

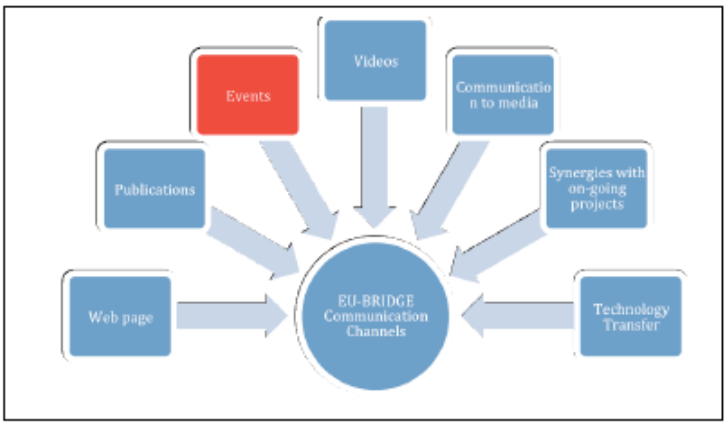
We created some eye-catching posters with general and deeper information on the project. The posters have benn made available to all partners for further use.

Internal Working Guideline

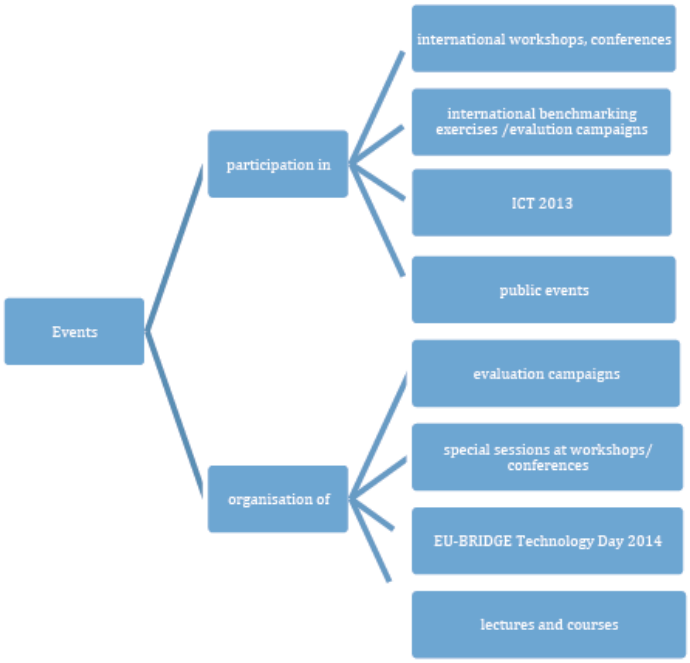
At the beginning of the project an internal working guideline was published with relevant information on the corporate design and requirements by the European Commission. It was uploaded to the intranet and was updated when necessary.

Events

Via participating in various kinds of events all target groups of EU-BRIDGE were reached: Researchers, students, the general public, politicians, industry, SMEs etc. We have choosen the best and most appropriate out of the great variety of events. EU-BRIDGE considered of importance to participate in or to organise the following events:



All EU-BRIDGE partners presented their work at international conferences and workshops. The most relevant conferences and workshops were communicated to the partners by the calendar on the EU-BRIDGE-intranet and by sending around a list with relevant and upcoming events. The events were furthermore published on www.eu-bridge.eu.
The consortium participated in international benchmarking exercises and evaluation campaigns



to establish the joint consortiums advances and innovations vis-a-vis the state of the art. Members of the consortium are the initiators or co-initiators of some of the most relevant evaluation campaigns in this domain (IWSLT, WMT). The project participated in the main evaluation campaigns like IWSLT, WMT, and NIST-openMT.

The ICT Conference is one of the major European fairs in the field of computer science and a perfect tool to get the project and its results communicated to the different target groups. The EU-BRIDGE consortium took this opportunity to showcase its capabilities to the European public, scientists and decision makers at ICT 2013 in Vilnius.

An **EU-BRIDGE-Technology Day** communicated results to commercial players, service providers and product developers – addressing among others the European Speech Translation technology market. Hands- on and life demonstrations presented the technologies developed in the project. Researchers of the project explained the technologies and their potential to the attendees. Commercial players, service providers and product developers were able to deeply discuss the developed technology, application sectors, technical requirements, terms of availability and IPR protection. The technology day took place on March 17, 2015 at CeBIT 2015.

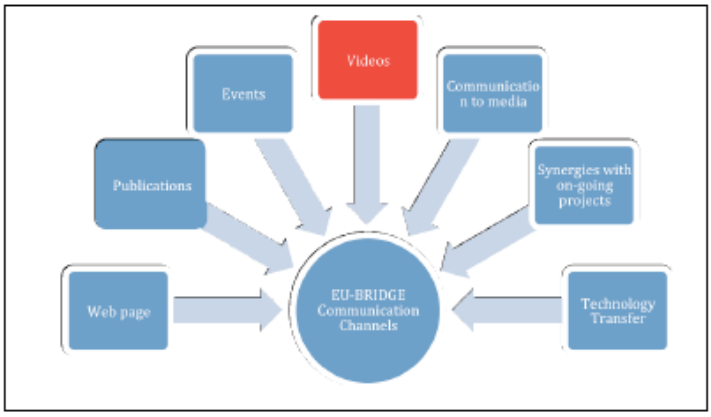
Training and education

The EU-BRIDGE project also saw its role in educating the public as well as future generations of scientists working on machine translation and multilingual technologies. At the participating universities and research institutions (KIT, RWTH, PJIIT, UEDIN, HKUST, FBK) lectures and courses were offered in machine translation as well as related sciences. The proposed technology was object of our education as well as tool for our education, i.e., we proposed to train scientists in human language technology, but also do so in a multilingual setting using organically evolving MT: we will be "eating our own dog food" as part of our educational efforts.

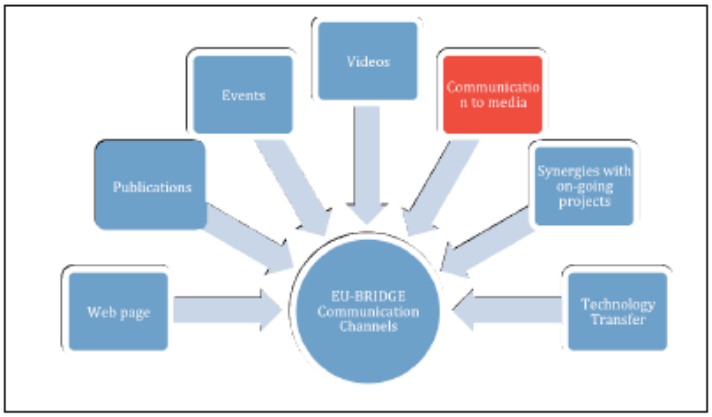
Videos

In today's visual and acoustic world it is a particularly effective mechanism to easily explain, educate and entertain a project by broadcasting an understandable video what EU-BRIDGE is about. A professionally designed video presented the idea of EU-BRIDGE and showcased its use for everybody and additionally to a broad audience.

Demo-videos of technologies developed during the project allowed the viewer to catch up the technology in an easily understandable and rapid way. Demo videos were shown on the website, on YouTube (or other internett platforms), and of course on conferences and in presentations. EU-BRIDGE's partners used the videos during presentations at events like conferences, workshops, open-door-days, etc.



Communication to the media



The media was both our target group and a channel to communicate EU-BRIDGE to the other groups – the media was of highest importance in the loop of communicating EU-BRIDGE's messages. Target media audience covered local, national, European and international media, scientific journals, online press as well as television and radio. Articles in the daily press as well as in the technical press made the project known and reached a wide spectrum of readers.

Press information were send out at the beginning of the project, during the project on breaking news and at the end of the project. As every partner has a press department we strived for a close cooperation with these departments, which gave us the possibility to use their contacts and channels to pass EU-BRIDGE's message to the media.

We invited the media to the technology day and to our participation in fairs and exhibitions in order to give a lively and hands-on insight in the project and its people.

The most suitable Channels for the specific Target Groups

After having defined our target groups and the key messages we passed on via the communication channels, we would like to summarize the most suitable channels for the specific target groups. As the groups, their messages (see under point 2. Target Groups and Key Messages) and the channels (see under point 3. Communication Channels) were explained in detail, a short

overview will be given here:

	Scientific Community	General Public	Decision Makers, Politicians	Industry	Media	EU-BRIDGE partners
Webpage,....						
Webpage	x	x	x	x	x	x
Social Media Networks		x				x
Intranet						x
Mailing-Lists						x
Publications						
Scientific publications	x			(x)	(x)	x
Project brochure for scientific community	x		(x)	x	(x)	x
Project brochure for general public		x	x	x	x	x
EU-BRIDGE newsletter	x		x	x	x	x
Fact Sheets	x		(x)	x	(x)	x
EU-BRIDGE technology catalogue	x		x	x	x	x
EU-BRIDGE project poster	x	x	x	x	x	x
EU-BRIDGE internal working guideline						x
Events						
international workshops, conferences	x					x
benchmarking exercises / evaluations	x					x
ICT 2014	x	x	x	x	x	x
public events		x	x		x	x
EU-BRIDGE technology day	x		x	x	x	x
university lectures/summer school	x					x
Videos						
project video	x	x	x	x	x	x
demo videos	x		x	x	x	x
Communication to Media						
					x	
Synergies with ongoing projects						
	x			x		
Technology Transfer						
	x			x	x	

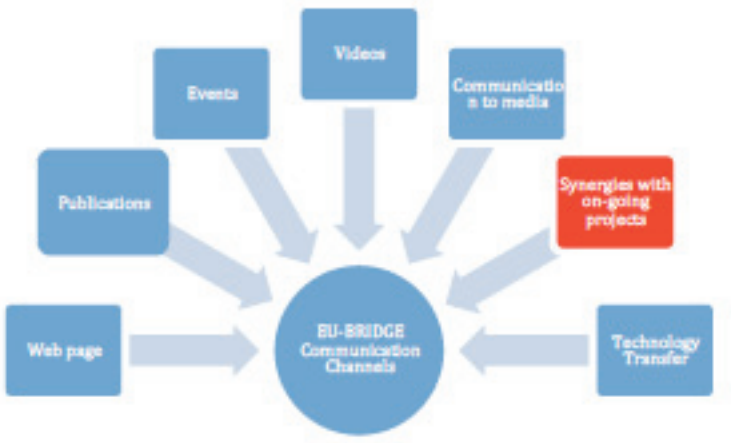
Corporate Design and Presentations

Identity development is supported by a corporate EU-BRIDGE design. We created an EU-BRIDGE logo, an EU-BRIDGE website, a template for ppt-presentations, a fact sheet with key figures of the project, a letter head, and several templates for posters. Brochures, fact sheets and the technology catalogue were all in line with the corporate design.

All templates of the Corporate Design and the Power-Point Presentation were published on EU-BRIDGE's intranet. The internal EU-BRIDGE working guideline helped EU-BRIDGE's partners to follow the corporate design of the project. Templates were adapted and updated during the project.

Synergies with on-going projects

EU-BRIDGE is cooperating with xLiMe by offering speech transcription services for German and English news broadcasts. For this xLiMe uses the EU-BRIDGE service archi- tecture via the Pervoice Mediator to access different transcription workers. xLiMe is especially inter- ested in low-latency real-time workers. Currently xLiMe is in the process of testing the workers made available by EU-BRIDGE with respect to their performance.



EU-BRIDGE is also cooperating with Albert Zündorf, professor at the University of Kassel, in the field of lecture transcription. Zündorf is recording his lectures and planning to build up a searchable archive of university lectures. For that EU-BRIDGE is offering a German transcription system that was adapted in an unsupervised way to a collection of previous recordings of Albert Zündorf.

The operators of the website sugartrends.de have contacted KIT for a service that translate product descriptions of their website from English to German and vice versa. KIT has run first preliminary trials on sample test data, to give an impression of translation quality. If the quality is sufficient, a service will be offered to sugartrends.de via the service architecture.

Technology Transfer

Technology transfer in EU-BRIDGE is done with the helpf of the service infrastructure developed by it and marketed by PEV. The partners ADX, KIT and RBM have realized use cases with the help of this infrastructure that demonstrate the usefulness of it. Especially in the case of ADX and RBM this has lead to direct technology transfer.
ADX, PEV and RBM have created business plans taken the service infrastructure and EU-BRIDGE technology into account for commercializing it. KIT is

continuing discussion with the European Parliament for bringing EU-BRIDGE technology into its operations and finding new use cases for it.

Evaluation

We considered it of high importance whether the dissemination activities were fitting, were of added value, were meeting the needs, were effective, or whether we needed to adapt our ideas and tasks. It is crucial to think about evaluation on beforehand, to develop fitting evaluation tools and to draw the consequences. That's why the consortium intended to set up a dedicated task for evaluation. EU-BRIDGE defined the following evaluation tools:

- Press and media review, in cooperation with the press departments of the partners we monitored printing and online articles.

- We provided web-statistics, including the number of hits and in-links over time. KIT's computation centre granted it's support and provided professional web statics services.

- During the participation in exhibitions, information and evaluation sheets were filled in.

- The number of scientific publications ensured the scientific presence.

Table A.1 LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS

Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers[2] (if available)	Is/Will open access[3] provided to this publication?
Improving Machine Translation via Triangulation and Transliteration	Nadir Durrani, Philipp Koehn	Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT 2014)	17th Annual Conference of the European Association for Machine Translation (EAMT 2014), June 16-18, 2014		Dubrovnik, Croatia	June 2014			yes
Dynamic Topic Adaptation for Phrase-based MT	Eva Hasler, Phil Blunsom, Philipp Koehn, Barry Haddow	Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)	14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), April 2014		Gothenburg, Sweden	April 2014			yes
Investigating the Usefulness of Generalized Word Representations in SMT	Nadir Durrani, Philipp Koehn, Helmut Schmid, Alexander Fraser	Proceedings of the 25th International Conference on Computational Linguistics (COLING)	25th International Conference on Computational Linguistics (COLING), August 23-29, 2014		Dublin, Ireland	August 2014			yes

Combining Domain and Topic Adaptation for SMT	Eva Hasler, Barry Haddow, Philipp Koehn	Proceedings of the Eleventh Conference of the Association for Machine Translation in the Americas (AMTA)	11th Conference of the Association for Machine Translation in the Americas (AMTA), October 22-26, 2014		Vancouver, Canada	Ocotber 2014			yes	
Real-Time Statistical Speech Translation	Krzysztof Wolk, Krzysztof Marasek	Intelligent Systems and Computing volume 275	Springer, ISSN 2194-5357, ISBN 978-3-319-05950-1			2014			yes	
A Sentence Meaning Based Alignment Method for Parallel Text Corpora Preparation	Krzysztof Wolk, Krzysztof Marasek	Advances in Intelligent Systems and Computing volume 275, p.107-114	Publisher: Springer, ISSN 2194-5357, ISBN 978-3-319-05950-1			2014			yes	
Spoken Language Translation for Polish	Krzysztof Wolk, Krzysztof Marasek	Proceedings of Forum Acusticum	Sep. 07-12, 2014		Krakow, Poland	Sep-14			yes	
Alignment of the Polish-English Parallel Text for a Statistical Machine Translation	Krzysztof Wolk, Krzysztof Marasek	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT)	11th International Workshop on Spoken Language Translation (IWSLT), December 4-5, 2014		Lake Tahoe, USA	Dec-14			yes	
Polish - English Statistical Machine	Krzysztof Wolk, Krzysztof Marasek	New Research in Multimedia and Internet	2014			2014			yes	

Translation of Medical Texts		Systems, Springer, ISSN: 1867-5662								
Building subject-aligned comparable corpora and mining it for truly parallel sentence pairs	Krzysztof Wolk, Krzysztof Marasek	Proceedings of the International workshop on Innovations in Information and Communication Science and Technology (IICST 2014)	September 3-5, 2014		Warsaw, Poland	Sep-14			yes	
Alignment of the Polish-English Parallel Text for a Statistical Machine Translation	Krzysztof Wolk, Krzysztof Marasek	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT)	11th International Workshop on Spoken Language Translation (IWSLT), December 4-5, 2014		Lake Tahoe, USA	Dec-14			yes	
Enhanced Bilingual Evaluation Understudy	Krzysztof Wolk, Krzysztof Marasek	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT)	11th International Workshop on Spoken Language Translation (IWSLT), December 4-5, 2014		Lake Tahoe, California, USA	Dec-14			yes	
Evaluating Improvised Hip Hop Lyrics - Challenges and Observations[1]	Karteek Addanki, Dekai Wu	Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)	9th International Conference on Language Resources and Evaluation, (LREC 2014), May 2014		Reykjavik, Iceland	May-14			yes	

Improving Egyptian-to-English SMT by mapping Egyptian into MSA	Nadir Durrani, Yaser Al-Onaizan, Abraham Ittycheriah	Proceedings of the 14th Conference on Intelligent Text Processing and Computational Linguistics (CICLING)	14th Conference on Intelligent Text Processing and Computational Linguistics (CICLING), April 2014		Kathmandu, Nepal	Apr-14			yes	
The RWTH Aachen German-English Machine Translation System for WMT 2014	Stephan Peitz, Joern Wuebker, Markus Freitag, Hermann Ney	Proceedings of the 9th Workshop on Statistical Machine Translation	9th Workshop on Statistical Machine Translation, June 2014		Baltimore, USA	Jun-14			yes	
Sesla transcriber: A speech transcription tool that adapts to your skill and time budget	Matthias Sperber, Graham Neubig, Satoshi Nakamura, Alexander Waibel	Proceedings of the Spoken Language Technology Workshop (SLT 2014)	Dec 7-10, 2014		Lake Tahoe, California, USA	Dec-14			yes	
On-the-fly user modeling for cost-sensitive correction of speech transcripts	Matthias Sperber, Graham Neubig, Satoshi Nakamura, Alexander Waibel	Proceedings of the Spoken Language Technology Workshop (SLT 2014)	Dec 7-10, 2014		Lake Tahoe, California, USA	Dec-14			yes	
The 2014 KIT IWSLT Speech-to-Text Systems for English, German and Italian	Kevin Kilgour, Michael Heck, Markus Müller, Matthias Sperber, Sebastian Stüker, Alexander Waibel	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT)	Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	

Rule-Based Preordering on Multiple Syntactic Levels in Statistical Machine Translation	Ge Wu, Yuqi Zhang, Alexander Waibel	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT)	Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
Improving In-Domain Data Selection For Small In-Domain Sets	Mohammed Mediani, Joshua Winebarger, Alexander Waibel	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT)	Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
RWTH LVCSR Systems for Quaero and EU-Bridge: German, Polish, Spanish and Portuguese	M. Ali Basha Shaik, Zoltan Tüske, M. Ali Tahir, Markus Nußbaum-Thom, Ralf Schlüter, Hermann Ney	Proceedings of 15th Annual Conference of the International Speech Communication Association	September 14-18, 2014		Singapore	Sep-14			yes	
Lattice Decoding and Rescoring with Long-Span Neural Network Language Models	Martin Sundermeyer, Zoltán Tüske, Ralf Schlüter, Hermann Ney	Proceedings of 15th Annual Conference of the International Speech Communication Association	September 14-18, 2014		Singapore	Sep-14			yes	
rwthlm - The RWTH Aachen University Neural Network Language Modeling Toolkit	Martin Sundermeyer, Ralf Schlüter, Hermann Ney	Proceedings of 15th Annual Conference of the International Speech Communication Association	September 14-18, 2014		Singapore	Sep-14			yes	

Deep neural network adaptation for children's and adults' speech recognition	Romain Serizel, Diego Giuliani	Proceedings of the Italian Conference on Computational Linguistics	December 9-10, 2014		Pisa, Italy	14-Dec			yes	
Report on the 11th IWSLT Evaluation Campaign IWSLT 2014	Mauro Cettolo, Jan Niehues, Sebastian Stüker, Luisa Bentivogli, Marcello Federico	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
FBK's Machine Translation and Speech Translation Systems for the IWSLT 2014 Evaluation Campaign	Nicola Bertoldi, Prashant Mathur, Nicholas Ruiz, Marcello Federico	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
FBK @ IWSLT 2014 - ASR track	B. Babaali, R. Serizel, S. Jalalvand, D. Falavigna, R. Gretter, D. Giuliani	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
Improving MEANT Based Semantically Tuned SMT	Meriem Beloucif, Chi-kiu Lo, Dekai Wu	Proceedings of the 11th International Workshop on Spoken Language	11th International Workshop on Spoken Language Translation		Lake Tahoe, California, USA	14-Dec			yes	

		Translation (IWSLT 2014), Lake Tahoe, USA, December 4-5, 2014	(IWSLT 2014), Dec 4-5, 2014							
Edinburgh SLT and MT System Description for the IWSLT 2014 Evaluation	Alexandra Birch, Matthias Huck, Nadir Durrani, Nikolay Bogoychev, Philipp Koehn	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
Combined Spoken Language Translation	Markus Freitag, Joern Wuebker, Stephan Peitz, Hermann Ney, Matthias Huck, Alexandra Birch, Nadir Durrani, Philipp Koehn, Mohammed Mediani, Isabel Slawik, Jan Niehues, Enuah Cho, Alex Waibel, Nicola Bertoldi, Mauro Cettolo, Marcello Federico	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
The UEDIN ASR Systems for the IWSLT 2014 Evaluation	Peter Bell, Pawel Swietojanski, Joris Driesen, Mark Sinclair, Fergus McInnes, Steve Renals	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	

The KIT Translation System for IWSLT 2014	Isabel Slawik, Mohammed Mediani, Jan Niehues, Yuqi Zhang, Eunah Cho, Teresa Herrmann, Thanh-Le Ha, Alex Waibel	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
Lexical Translation Model Using A Deep Neural Network Architecture	Thanh-Le Ha, Jan Niehues, Alex Waibel	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
Machine Translation of Multi-party Meetings: Segmentation and Disfluency Removal Strategies	Eunah Cho, Jan Niehues, Alex Waibel	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
Report on the 11th IWSLT Evaluation Campaign IWSLT 2014	Mauro Cettolo, Jan Niehues, Sebastian Stüker, Luisa Bentivogli, Marcello Federico	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	

The RWTH Aachen Machine Translation Systems for IWSLT 2014	Joern Wuebker, Stephan Peitz, Andreas Guta and Hermann Ney	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
Better Punctuation Prediction with Hierarchical Phrase-Based Translation	Stephan Peitz, Markus Freitag, Hermann Ney	Proceedings of the 11th International Workshop on Spoken Language Translation (IWSLT 2014)	11th International Workshop on Spoken Language Translation (IWSLT 2014), Dec 4-5, 2014		Lake Tahoe, California, USA	14-Dec			yes	
Better Semantic Frame Based MT Evaluation via Inversion Transduction Grammars	Dekai Wu, Chi-kiu Lo, Meriem Beloucif, Markus Saers	Proceedings of the EMNLP 8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)	8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)		Doha, Qatar	Oct-14			yes	
Ternary Segmentation for Improving Search in Top-down Induction of Segmental ITGs	Markus Saers, Dekai Wu	Proceedings of the EMNLP 8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)	8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)		Doha, Qatar	Oct-14			yes	

Transduction Recursive Auto-Associative Memory: Learning Bilingual Compositional Distributed Vector Representations of Inversion Transduction Grammars	Karteek Addanki, Dakai Wu	Proceedings of the EMNLP 8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)	8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)		Doha, Qatar	Oct-14			yes	
Translation Modeling with Bidirectional Recurrent Neural Networks	Martin Sundermeyer, Tamer Alkhouli, Joern Wuebker, and Hermann Ney	Conference on Empirical Methods in Natural Language Processing (EMNLP)	8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)		Doha, Qatar	Oct-14			yes	
Vector Space Models for Phrase-based Machine Translation	Tamer Alkhouli, Andreas Guta, and Hermann Ney	Conference on Empirical Methods in Natural Language Processing (EMNLP)	8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)		Doha, Qatar	Oct-14			yes	
Combining Techniques from different NN-based Language Models for Machine Translation	Jan Niehues, Alexander Allauzen, Francois Yvon, Alex Waibel	The eleventh biennial Conference of the Association for Machine Translation in the Americas (AMTA) 2014	11th biennial Conference of the Association for Machine Translation in the Americas (AMTA) 2014, October 22-26		Vancouver, Canada	Oct-14			yes	

Preference Grammars and Soft Syntactic Constraints for GHKM Syntax-based Statistical Machine Translation	Matthias Huck, Hieu Hoang, Philipp Koehn	Proceedings of the EMNLP 8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)	8th Workshop on Syntax Semantics and Structure in Statistical Translation (SSST-8)		Doha, Qatar	Oct-14			yes	
Direct Word Graph Rescoring Using A* Search and RNNLM	Shahab Jalalvand, Daniele Falavigna	Proceedings of 15th Annual Conference of the International Speech Communication Association	15th Annual Conference of the International Speech Communication Association, September 14-18, 2014		Singapore	Sep-14			yes	
Euronews: a multilingual benchmark for ASR and LID	Roberto Gretter	Proceedings of 15th Annual Conference of the International Speech Communication Association	15th Annual Conference of the International Speech Communication Association, September 14-18, 2014		Singapore	Sep-14			yes	
Cross-lingual adaptation with multi-task adaptive networks	Peter Bell, Joris Driesen, Steve Renals	Proceedings of 15th Annual Conference of the International Speech Communication Association	15th Annual Conference of the International Speech Communication Association, September 14-18, 2014		Singapore	Sep-14			yes	

Automated Production of True-cased Punctuated Subtitles for Weather and News Broadcasts	Joris Driesen, Alexandra Birch, Simon Grimsey, Saeid Safarfashandi, Juliet Gauthier, Matt Simpson, Steve Renals	Proceedings of 15th Annual Conference of the International Speech Communication Association	15th Annual Conference of the International Speech Communication Association, September 14-18, 2014		Singapore	Sep-14			yes	
A semi-Markov model for speech segmentation with an utterance-break prior	Mark Sinclair, Peter Bell, Alexandra Birch, Fergus McInnes	Proceedings of 15th Annual Conference of the International Speech Communication Association	15th Annual Conference of the International Speech Communication Association, September 14-18, 2014		Singapore	Sep-14			yes	
Lexical Access Preference and Constraint Strategies for Improving Multiword Expression Association within Semantic MT Evaluation	Dekai Wu, Chi-kiu Lo, Markus Saers	Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex)	4th Workshop on Cognitive Aspects of the Lexicon (CogALex)		Dublin, Ireland	Aug-14			yes	
Quality Estimation for Automatic Speech Recognition	Matteo Negri, Marco Turchi, José G. C. de Souza, Daniele Falavigna	Proceedings of the 25th International Conference on Computational Linguistics (COLING)	25th International Conference on Computational Linguistics (COLING), August 23-29, 2014		Dublin, Ireland	Aug-14			yes	

XMEANT: Better semantic MT evaluation without reference translations	Chi-kiu Lo, Meriem Beloucif, Markus Saers, Dekai Wu	Association for Computational Linguistics (ACL)	June 22-27, 2014		Baltimore, USA	Jun-14			yes	
The KIT-LIMSI Translation System for WMT 2014	Quoc Khan Do, Teresa Herrmann, Jan Niehues, Alexandre Allauzen, Francois Yvon, Alex Waibel	Association for Computational Linguistics (ACL)	June 22-27, 2014		Baltimore, USA	Jun-14			yes	
EU-Bridge MT: Combined Machine Translation	Markus Freitag, Stephan Peitz, Joern Wuebker, Hermann Ney, Matthias Huck, Rico Sennrich, Nadir Durrani, Maria Nadejde, Philip Williams, Philipp Koehn, Teresa Herrmann, Eunah Cho, Alex Waibel	Association for Computational Linguistics (ACL)	June 22-27, 2014		Baltimore, USA	Jun-14			yes	
Edinburgh's Syntax-Based Systems at WMT 2014	Philip Williams, Rico Sennrich, Maria Nadejde, Matthias Huck, Eva Hasler, Philipp Koehn	Association for Computational Linguistics (ACL)	June 22-27, 2014		Baltimore, USA	Jun-14			yes	
Edinburgh's Phrase-based Machine Translation Systems for WMT-14	Nadir Durrani, Barry Haddow, Philipp Koehn, Kenneth Heafield	Association for Computational Linguistics (ACL)	June 22-27, 2014		Baltimore, USA	Jun-14			yes	

Augmenting String-to-Tree and Tree-to-String Translation with Non-Syntactic Phrases	Matthias Huck, Hieu Hoang, Philipp Koehn	Proceedings of the 9th Workshop on Statistical Machine Translation	9th Workshop on Statistical Machine Translation		Baltimore, USA	Jun-14			yes	
Segmentation for Efficient Supervised Language Annotation with an Explicit Cost-Utility Tradeoff	Matthias Sperber, Mirjam Simantzik, Graham Neubig, Satoshi Nakamura, Alex Waibel	The 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), June 22-27, 2014	52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), June 22-27, 2014		Baltimore, USA	Jun-14			yes	
Euronews: a multilingual speech corpus for ASR	Roberto Gretter	Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC 2014)	9th edition of the Language Resources and Evaluation Conference (LREC 2014), 26-31 May, 2014		Reykjavik, Iceland	May-14			yes	
On the reliability and inter-annotator agreement of human semantic MT evaluation via HMEANT	Chi-kiu Lo, Dekai Wu	Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC 2014)	9th edition of the Language Resources and Evaluation Conference (LREC 2014), 26-31 May, 2014		Reykjavik, Iceland	May-14			yes	
A Corpus of Spontaneous Speech in Lectures : The	Eunah Cho, Sarah Fünfer, Sebastian Stüker, Alex Waibel	Proceedings of the 9th edition of the Language Resources and	9th edition of the Language Resources and Evaluation		Reykjavik, Iceland	May-14			yes	

KIT Lecture Corpus for Spoken Language Processing and Translation		Evaluation Conference (LREC 2014)	Conference (LREC 2014), 26-31 May, 2014							
Vocal Tract Length Normalisation Approaches to DNN-Based Children's and Adults' Speech Recognition	Romain Serizel, Diego Giuliani	2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	May 4-9, 2014		Florence, Italy	May-14			yes	
Euronews: A Multilingual Speech Corpus for ASR	R. Gretter	Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC 2014)	9th edition of the Language Resources and Evaluation Conference (LREC 2014), 26-31 May, 2014		Reykjavik, Iceland	May-14			yes	
Mean-Normalized Stochastic Gradient for Large-Scale Deep Learning	S. Wiesler, A. Richard, R. Schlüter, H. Ney	2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	May 4-9, 2014		Florence, Italy	May-14			yes	
RASR/NN: The RWTH Neural Network Toolkit for Speech Recognition	S. Wiesler, A. Richard, P. Golik, R. Schlütter, H. Ney	2014 IEEE International Conference on Acoustics, Speech and Signal	May 4-9, 2014		Florence, Italy	May-14			yes	

		Processing (ICASSP)								
The magic number 4: Evolutionary pressures on semantic frame structure	Dekai Wu	Proceedings of the 10th International Conference on the Evolution of Language (Evolang X)	10th International Conference on the Evolution of Language (Evolang X), April 2014		Vienna, Austria	Apr-14			yes	
Jane: Open Source Machine Translation System Combination	Markus Freitag, Matthias Huck, Hermann Ney	Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)	14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), April 2014		Gothenborg, Sweden	Apr-14			yes	
Tight Integration of Speech Disfluency Removal into SMT	Eunah Cho, Jan Niehues, Alex Waibel	The 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)	14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014), April 26-30, 2014		Gothenborg, Sweden	Apr-14			yes	
Using Feature Structures to Improve Verb Translation in English-to-German Statistical MT	Philip Williams, Philipp Koehn	Proceedings of the 3rd Workshop on Hybrid Approaches to Machine Translation (HyTra)	3rd Workshop on Hybrid Approaches to Machine Translation (HyTra)		Gothenborg, Sweden	Apr-14			yes	

Integrating an Unsupervised Transliteration Model into Statistical Machine Translation	Nadir Durrani, Hassan Sajjad, Hieu Hoang, Philipp Koehn	Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)	14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)		Gothenborg, Sweden	Apr-14			yes	
Simple and Effective Approach for Consistent Training of Hierarchical Phrase-based Translation Models	S. Peitz, D. Vilar, H. Ney	The 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)	14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), April 26-30, 2014		Gothenborg, Sweden	Apr-14			yes	
Dynamic Topic Adaptation for SMT using Distributional Profiles	Eva Hasler, Barry Haddow, Philipp Koehn	Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)	14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), April 26-30, 2014		Gothenborg, Sweden	Apr-14			yes	
Dynamically Shaping the Reordering Search Space of Phrase-Based Statistical	A. Bisazza, M. Federico	Transactions of the Association for Computational Linguistics Vol. 1	Vol. 1, 2013		Trento, Italy	2013	pp. 327-340		yes	

Machine Translation										
Lightly Supervised Automatic Subtitling of Weather Forecasts	Joris Driesen, Steve Renals	2013 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2013)	December 8-12, 2013		Olomouc, Czech Republic	Dec-13			yes	
Phonetic and Anthropometric Conditioning of MSA-KST Cognitive Impairment Characterization System[1]	Alexei V. Ivanov, Shahab Jalalvand, Roberto Gretter, Daniele Falavigna	2013 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2013)	December 8-12, 2013		Olomouc, Czech Republic	Dec-13			yes	
Building an Arabic News Transcription System With Web-crawled Resources	Arianna Bisazza, Roberto Gretter	Proceedings of the 6th Language & Technology Conference (LTC 2013)	6th Language & Technology Conference (LTC 2013), December 7-9, 2013		Poznan, Poland	Dec-13			yes	
Analyzing the Potential of Source Sentence Reordering in Statistical Machine Translation	Teresa Herrmann, Jochen Weiner, Jan Niehues, Alex Waibel	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	

CRF-based Disfluency Detection using Semantic Features for German to English Spoken Language Translation	Eunah Cho, Thanh-Le Ha, Alex Waibel	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
Parameter Optimization for Iterative Confusion Network Decoding in Weather-Domain Speech Recognition	Shabab Jalalvand, Daniele Falavigna	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
Description of the UEDIN System for German ASR	Joris Driesen, Peter Bell, Mark Sinclair, Steve Renals	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
Edinburgh SLT and MT System Description for the IWSLT 2013 Evaluation	Alexandra Birch, Nadir Durrani, Philipp Koehn	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	

EU-BRIDGE MT: Text Translation of Talks in the EU-BRIDGE Project	Markus Freitag, Stephan Peitz, Joern Wuebker, Hermann Ney, Nadir Durrani, Matthias Huck, Philipp Koehn, Thanh-Le Ha, Jan Niehues, Mohammed Mediani, Teresa Herrmann, Alex Waibel, Nicola Bertoldi, Mauro Cettolo, Marcello Federico	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
FBK @ IWSLT 2013 - ASR tracks	D. Falavigna, R. Gretter, F. Brugnara, D. Giuliani, R. H. Serizel	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
FBK's Machine Translation Systems for the IWSLT 2013 Evaluation Campaign	Nicola Bertoldi, M. Amin Farajian, Prashant Mathur, Nicholas Ruiz, Marcello Federico	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	

Human Semantic MT Evaluation with HMEANT for IWSLT 2013	Chi-kiu Lo, Dekai Wu	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
Improving Machine Translation into Chinese by Tuning Against Chinese MEANT	Chi-kiu Lo, Meriem Beloucif, Dekai Wu	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
Incremental Unsupervised Training for University Lecture Recognition	Michael Heck, Sebastian Stüker, Sakriani Sakti, Alex Waibel, Satoshi Nakamura	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
Maximum Entropy Language Modeling for Russian ASR	Evgeniy Shin, Sebastian Stüker, Kevin Kilgour, Christian Fügen, Alex Waibel	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	

Parameter Optimization for Iterative Confusion Network Decoding in Weather-Domain Speech Recognition	Shahab Jalalvand, Daniele Falavigna	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
Polish - English Speech Statistical Machine Translation Systems for the IWSLT 2013	Krzysztof Wołk, Krzysztof Marasek	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
Report on the 10th IWSLT Evaluation Campaign	Mauro Cettolo, Jan Niehues, Sebastian Stüker, Luisa Bentivogli, Marcello Federico	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
The KIT Translation Systems for IWSLT 2013	Thanh-Le Ha, Teresa Herrmann, Jan Niehues, Mohammed Mediani, Eunah Cho, Yuqi Zhang, Isabel Slawik and Alex Waibel	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	

The RWTH Aachen German and English LVCSR systems for IWSLT-2013	M. Ali Basha Shaik, Zoltan Tüske, Simon Wiesler, Markus Nußbaum-Thom, Stephan Peitz, Ralf Schlüter and Hermann Ney	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
The RWTH Aachen Machine Translation Systems for IWSLT 2013	Joern Wuebker, Stephan Peitz, Tamer Alkhouli, Jan-Thorsten Peter Minwei Feng, Markus Freitag and Hermann Ney	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
The UEDIN English ASR System for the IWSLT 2013 Evaluation	Peter Bell, Fergus McInnes, Siva Reddy Gangireddy, Mark Sinclair, Alexandra Birch, Steve Renals	Proceedings of the 10th International Workshop for Spoken Language Translation (IWSLT 2013)	10th International Workshop for Spoken Language Translation (IWSLT 2013), December 5-6, 2013		Heidelberg, Germany	Dec-13			yes	
Learning Bilingual Categories in Unsupervised Inversion Transduction Grammar Induction	Markus Saers, Dekai Wu	13th International Conference on Parsing Technologies (IWPT 2013)	13th International Conference on Parsing Technologies (IWPT 2013), November 27-29, 2013		Nara, Japan	Nov-13			yes	

Simultaneous Unsupervised Learning of Flamenco Metrical Structure, Hypermetrical Structure, and Multipart Structural Relations[1]	Dekai Wu	14th International Society for Music Information Retrieval Conference (ISMIR 2013)	14th International Society for Music Information Retrieval Conference (ISMIR 2013), November 4-8, 2013		Curibata, Brazil	Nov-13			yes	
Polish to English Statistical Machine Translation	Krzysztof Wołk	XV International Phd Workshop OWD 2013	XV International Phd Workshop OWD 2013, October 19-22, 2013		Wisla, Poland	Oct-13			yes	
Improving Statistical Machine Translation with Word Class Models	Joern Wuebker, Stephan Peitz, Felix Rietig and Hermann Ney	EMNLP 2013, Conference on Empirical Methods in Natural Language Processing	October 18-21, 2013		Seattle, USA	Oct-13			yes	
Learning to Freestyle: Hip Hop Challenge-Response Induction via Transduction Rule Segmentation[1]	Dekai Wu, Karteek Addanki, Markus Saers, Meriem Beloucif	EMNLP 2013, Conference on Empirical Methods in Natural Language Processing	October 18-21, 2013		Seattle, USA	Oct-13			yes	
Bayesian Induction of Bracketing Inversion Transduction Grammars	Markus Saers, Dekai Wu	6th International Joint Conference on Natural Language Processing	October 14-18, 2013		Nagoya, Japan	Oct-13			yes	

		(IJCNLP 2013)								
Improving Language Model Adaptation using Automatic Data Selec- tion and Neural Network	Shahab Jalalvand	Proceedings of the Student Research Workshop associated with RANLP	September 9-11, 2013		Hissar, Bulgaria	Sep-13			yes	
Segmenting vs. Chunking: Unsupervised ITG Induction via Minimum Conditional Description Length	Markus Saers, Karteek Assanki, Dekai Wu	Recent Advances in Natural Language Processing (RANLP 2013)	September 7-13, 2013		Hissar, Bulgaria	Sep-13			yes	
Can Informal Genres Be Better Translated by Tuning on Automatic Semantic Metrics?	Chi-kiu Lo, Dekai Wu	Machine Translation Summit XIV (MT Summit 2013), September 2-6, 2013	Machine Translation Summit XIV (MT Summit 2013), September 2-6, 2013		Nice, France	Sep-13			yes	
Modeling Hip Hop Challenge-Response Lyrics as Machine Translation[1]	Dekai Wu, Karteek Addanki, Markus Saers	Machine Translation Summit XIV (MT Summit 2013), September 2-6, 2013	Machine Translation Summit XIV (MT Summit 2013), September 2-6, 2013		Nice, France	Sep-13			yes	

A Lecture Transcription System Combining Neural Network Acoustic and Language Models	Peter Bell, Hitoshi Yamamoto, Pawel Swietojanski, Youzheng Wu, Fergus McInnes, Chiori Hori and Steve Renals	Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)	14th Annual Conference of the International Speech Communication Association (Interspeech 2013), August 25-29, 2013		Lyon, France	Aug-13			yes	
A Real-World System for Simultaneous Translation of German Lectures	Eunah Cho, Christian Fügen, Teresa Hermann, Kevin Kilgour, Mohammed Mediani, Christian Mohr, Jan Niehues, Kay Rottmann, Christian Saam, Sebastian Stüker, Alex Waibel	Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)	14th Annual Conference of the International Speech Communication Association (Interspeech 2013), August 25-29, 2013		Lyon, France	Aug-13			yes	
An On-line Incremental Speaker Adaptation Technique for Audio Stream Transcription	Diego Giuliani, Fabio Brugnara	Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)	14th Annual Conference of the International Speech Communication Association (Interspeech 2013), August 25-29, 2013		Lyon, France	Aug-13			yes	
Efficient Speech Transcription Through Respeaking	Matthias Sperber, Graham Neubig, Christian Fügen, Satoshi Nakamura, Alex Waibel	Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech	14th Annual Conference of the International Speech Communication Association (Interspeech 2013), August		Lyon, France	Aug-13			yes	

		2013)	25-29, 2013							
Freestyle: A Challenge-Response System for Hip Hop Lyrics via Unsupervised Induction of Stochastic Transduction Grammars[1]	Dekai Wu, Karteek Addanki, Markus Saers	Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)	14th Annual Conference of the International Speech Communication Association (Interspeech 2013), August 25-29, 2013		Lyon, France	Aug-13			yes	
Multilingual Hierarchical MRASTA Features for ASR	Zoltán Tüskea, Ralf Schlütera, Hermann Ney	Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)	14th Annual Conference of the International Speech Communication Association (Interspeech 2013), August 25-29, 2013		Lyon, France	Aug-13			yes	
Relative Error Bounds for Statistical Classifiers Based on the f-Divergence	Markus Nussbaum-Thom, Eugen Beck, Tamer Alkhouli, Ralf Schlüter, Hermann Ney	Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)	14th Annual Conference of the International Speech Communication Association (Interspeech 2013), August 25-29, 2013		Lyon, France	Aug-13			yes	

Training Log-Linear Acoustic Models in Higher-Order Polynomial Feature Space for Speech Recognition	M. Tahir, H. Huang, R. Schlüter, H. Ney, L. t. Bosch, B. Cranen, L. Boves	Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)	14th Annual Conference of the International Speech Communication Association (Interspeech 2013), August 25-29, 2013		Lyon, France	Aug-13			yes	
Slightly Supervised Adaptation of Acoustic Models on Captioned BBC Weather Forecasts	Christian Mohr, Christian Saam, Kevin Kilgour, Jonas Gehring, Sebastian Stüker, Alex Waibel	SLAM 2013 (First Workshop on Speech, Language and Audio in Multimedia)	First Workshop on Speech, Language and Audio in Multimedia, August 22-23, 2013		Marseille, France	Aug-13			yes	
Efficient solutions for word reordering in German-English phrase-based statistical machine translation	Arianna Bisazza, Marcello Federico	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Dynamically Shaping the Reordering Search Space of Phrase-Based Statistical Machine Translation	Arianna Bisazza, Marcello Federico	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
A Phrase Orientation Model for Hierarchical Machine	Matthias Huck, Joern Wuebker, Felix Rietig, Hermann Ney	ACL 2013, Proceedings of the Eighth Workshop on Statistical	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	

Translation		Machine Translation								
An MT Error-driven Discriminative Word Lexicon using Sentence Structure Features;Jan Niehues and Alex Waibel	Jan Niehues and Alex Waibel	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Edinburgh's Machine Translation Systems for European Language Pairs	Nadir Durrani, Barry Haddow, Kenneth Heafield, Philipp Koehn	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Edinburgh's Syntax-Based Machine Translation Systems	Maria Nadejde, Philip Williams, Philipp Koehn	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
MEANT at WMT 2013: A Tunable, Accurate Yet Inexpensive Semantic Frame Based MT Evaluation Metric	Chi-kiu Lo, Dekai Wu	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	

Munich-Edinburgh-Stuttgart Submissions at WMT13: Morphological and Syntactic Processing for SMT	Marion Weller, Max Kisselew, Svetlana Smekalova, Alexander Fraser, Helmut Schmid, Nadir Durrani, Hassan Sajjad, Richárd Farkas	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Munich-Edinburgh-Stuttgart Submissions of OSM Systems at WMT13	Nadir Durrani, Helmut Schmid, Alexander Fraser, Hassan Sajjad, Richárd Farkas	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
QCRI-MES Submission at WMT13: Using Transliteration Mining to Improve Statistical Machine Translation	Hassan Sajjad, Svetlana Smekalova, Nadir Durrani, Alexander Fraser, Helmut Schmid	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
The Feasibility of HMEANT as a Human MT Evaluation Metric	Alexandra Birch, Barry Haddow, Ulrich Germann, Maria Nadejde, Christian Buck, Philipp Koehn	ACL 2013, Proceedings of the Eighth Workshop on Statistical Machine Translation	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
The Karlsruhe Institute of Technology Translation Systems for the	Eunah Cho, Thanh-Le Ha, Mohammed Mediani, Jan Niehues, Teresa	ACL 2013, Proceedings of the Eighth Workshop on Statistical	8th Workshop on Statistical Machine Translation, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	

WMT 2013	Herrmann, Isabel Slawik, Alex Waibel	Machine Translation								
Letter N-Gram-based Input Encoding for Continuous Space Language Models	Henning Sperr, Jan Niehues, Alexander Waibel	ACL 2013, Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality	ACL 2013, Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality, August 4-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Can Markov Models Over Minimal Translation Units Help Phrase-Based SMT?	Nadir Durrani, Alexander Fraser, Helmut Schmid, Hieu Hoang, Philipp Koehn	ACL 2013, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics	51st Annual Meeting of the Association for Computational Linguistics, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Improving Machine Translation by Training Against an Automatic Semantic Frame Based Evaluation Metric	Chi-kiu Lo, Karteek Addanki, Markus Saers, Dekai Wu	ACL 2013, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics	51st Annual Meeting of the Association for Computational Linguistics, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Learning to Prune: Context-Sensitive Pruning for Syntactic MT	Wenduan Xu, Yue Zhang, Philip Williams, Philipp Koehn	ACL 2013, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics	51st Annual Meeting of the Association for Computational Linguistics, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	

Scalable Modified Kneser-Ney Language Model Estimation	Kenneth Heafield, Ivan Pouzyrevsky, Jonathan H. Clark, Philipp Koehn	ACL 2013, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics	51st Annual Meeting of the Association for Computational Linguistics, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Analysing Lexical Consistency in Translation	Liane Guillou	Proceedings of the ACL Workshop on Discourse in Machine Translation (DiscoMT 2013)	August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Unsupervised Transduction Grammar Induction via Minimum Description Length	Markus Saers, Karteek Addanki, Dekai Wu	Second Workshop on Hybrid Approaches to Translation (HyTra 2013), at ACL 2013	Second Workshop on Hybrid Approaches to Translation (HyTra 2013), at ACL 2013, August 8-9, 2013		Sofia, Bulgaria	Aug-13			yes	
Iterative Rule Segmentation under Minimum Description Length for Unsupervised Transduction Grammar Induction	Markus Saers, Karteek Addanki, Dekai Wu	First International Conference on Statistical Language and Speech Processing (SLSP 2013)	First International Conference on Statistical Language and Speech Processing (SLSP 2013), July 29-31, 2013		Tarragona, Spain	Jul-13			yes	

Unsupervised Rhyme Scheme Identification in Hip Hop Lyrics Using Hidden Markov Models[1]	Karteek Addanki, Dekai Wu	First International Conference on Statistical Language and Speech Processing (SLSP 2013)	First International Conference on Statistical Language and Speech Processing (SLSP 2013), July 29-31, 2013		Tarragona, Spain	Jul-13			yes	
A Performance Study of Cube Pruning for Large-Scale Hierarchical Machine Translation	Matthias Huck, David Vilar, Markus Freitag, Hermann Ney	Proceedings of SSST-7, Seventh Workshop on Syntax, Semantics and Structure in Statistical Translation (at NAACL HLT 2013)	SSST-7, 7th Workshop on Syntax, Semantics and Structure in Statistical Translation (at NAACL HLT 2013), June 13, 2013		Atlanta, GA, USA	Jun-13			yes	
Combining Top-down and Bottom-up Search for Unsupervised Induction of Transduction Grammars	Markus Saers, Karteek Addanki, Dekai Wu	Proceedings of SSST-7, Seventh Workshop on Syntax, Semantics and Structure in Statistical Translation (at NAACL HLT 2013)	SSST-7, 7th Workshop on Syntax, Semantics and Structure in Statistical Translation (at NAACL HLT 2013), June 13, 2013		Atlanta, GA, USA	Jun-13			yes	
Measuring the Structural Importance through Rhetorical	Narine Kokhlikyan, Alex Waibel, Yuqi Zhang, Joy Ying Zhang	The 2013 Conference of the North American Chapter of the	June 09-15, 2013		Atlanta, GA, USA	Jun-13			yes	

[1] This incorporates work performed under EU-BRIDGE, but was no costs of the work were charged to EU-BRIDGE

Structure Index		Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2013)								
Combining Word Reordering Methods on different Linguistic Abstraction Levels for Statistical Machine Translation	Teresa Herrmann, Jan Niehues, Alex Waibel	Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2013)	June 09-15, 2013		Atlanta, GA, USA	Jun-13			yes	
Grouping Language Model Boundary Words to Speed K?Best Extraction from Hypergraphs	Kenneth Heafield, Philipp Koehn, Alon Lavie	Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2013)	June 09-15, 2013		Atlanta, GA, USA	Jun-13			yes	
Model With Minimal Translation Units But Decode With Phrases	Nadir Durrani, Alexander Fraser, Helmut Schmid	Proceedings of the Conference of the North American Chapter of the Association for	June 09-15, 2013		Atlanta, GA, USA	Jun-13			yes	

		Computational Linguistics: Human Language Technologies (NAACL 2013)								
A Critical Evaluation of Stochastic Algorithms for Convex Optimization	S. Wiesler, A. Richard, R. Schlütter, H. Ney	2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)	May 26-31, 2013		Vancouver, BC, Canada	May-13			yes	
Comparison of Feedforward and Recurrent Neural Network Language Models	M. Sundermeyer, I. Oparin, J.-L. Gauvain, B. Freiberg, R. Schlüter, H. Ney	2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)	May 26-31, 2013		Vancouver, BC, Canada	May-13			yes	
Source-Side Discontinuous Phrases for Machine Translation: A Comparative Study on Phrase Extraction and Search	Matthias Huck, Erik Scharwächter, Hermann Ney	The Prague Bulletin of Mathematical Linguistics (PBML)	The Prague Bulletin of Mathematical Linguistics (PBML), Number 99, April 2013		Prague Czech Republic/Aachen, Germany	Apr-13	pp. 17-38		yes	
WIT3: il Corpus dei Sottotitoli Multilingue degli Interventi alle Conferenze	Mauro Cettolo, Christian Girardi, Marcello Federico	Proceedings of AISV 2013	Jan-13		Venezia, Italy	Jan-13			yes	

TED										
Esperimenti di identificazione della lingua parlata in ambito giornalistico	Diego Giuliani, Roberto Gretter	Proceedings of AISV 2013	Jan-13		Venezia, Italy	Jan-13			yes	
Accuracy and Robustness in Measuring the Lexical Similarity of Semantic Role Fillers for Automatic Semantic MT Evaluation	Anand Karthik Tumuluru, Chi-kiu Lo, Dekai Wu	Pacific Asia Conference on Language, Information and Computation	Nov-12		Indonesia, 2012	Nov-12			yes	
From Finite-State to Inversion Transductions: Toward Unsupervised Bilingual Grammar Induction	Markus Saers, Karteek Addanki, Dekai Wu	International Conference on Computational Linguistics	2012		Mumbai, India	2012			yes	
Fully Automatic Semantic MT Evaluation	Chi-kiu Lo, Anand Karthik Tumuluru, Dekai Wu	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	

Unsupervised vs. supervised weight estimation for semantic MT evaluation metrics	Chi-kiu Lo, Dekai Wu	Proceedings of SSST-6 Sixth Workshop on Syntax and Structure in Statistical Translation	SSST-6, 6th Workshop on Syntax and Structure in Statistical Translation, July 2012		Korea	Jul-12			yes	
Towards a Predicate-Argument Evaluation for MT	Ondřej Bojar, Dekai Wu	Proceedings of SSST-6 Sixth Workshop on Syntax and Structure in Statistical Translation	SSST-6, 6th Workshop on Syntax and Structure in Statistical Translation, July 2012		Korea	Jul-12			yes	
LTG vs. ITG Coverage of Cross-Lingual Verb Frame Alternations	Karteek Addanki, Chi-kiu Lo, Markus Saers, Dekai Wu	EAMT 2012	May-12		Trento, Italy	May-12			yes	
Building a Turkish ASR system with minimal resources	Arianna Bisazza, Roberto Gretter	Proceedings of LREC Workshop on Spoken Language Resoureces and Technologies for Turkic Languages	2012		Istanbul, Turkey	2012			yes	
FBK @ IWSLT 2012 - ASR track	D. Falavigna, R. Gretter, F. Brugnara, D. Giuliani	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	

FBK's Machine Translation Systems for IWSLT 2012's TED Lectures	N. Ruiz, A. Bisazza, R. Cattoni, M. Federico	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
Focusing Language Models For Automatic Speech Recognition	Daniele Falavigna, Roberto Gretter	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
Overview of the IWSLT 2012 Evaluation Campaign	M. Federico M. Cettolo, L. Bentivogli, M. Paul, S. Stüker	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
WIT3 : Web Inventory of Transcribed and Translated Talks	Mauro Cettolo, Marcello Federico, Christian Girardi	EAMT 2012	May-12		Trento, Italy	May-12			yes	
Spoken Language Translation Using Automatically Transcribed Text in Training	Stephan Peitz, Simon Wiesler, Markus Nußbaum-Thom, Hermann Ney	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	

The RWTH Aachen Speech Recognition and Machine Translation System for IWSLT 2012	Stephan Peitz, Saab Mansour, Markus Freitag, Minwei Feng, Matthias Huck Joern Wuebker, Malte Nuhn, Markus Nußbaum-Thom and Hermann Ney	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
TED Polish-to-English translation system for the IWSLT 2012	Krzysztof Marasek	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
The UEDIN Systems for the IWSLT 2012 Evaluation	Eva Hasler, Peter Bell, Arnab Ghoshal, Barry Haddow, Philipp Koehn, Fergus McInnes, Steve Renals, Pawel Swietojanski	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
Simulating Human Judgment in Machine Translation Evaluation Campaigns	Philipp Koehn	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
Sparse Lexicalised Features and Topic Adaptation for	Eva Hasler, Barry Haddow, Philipp Koehn	Proceedings of the International Workshop for Spoken Language	December 6-7, 2012		Hong Kong	Dec-12			yes	

SMT		Translation (IWSLT 2012)								
Detailed Analysis of different Strategies for Phrase Table Adaptation in SMT	Jan Niehues, Alex Waibel	Proceedings of the American Machine Translation Association (AMTA)	October 28 - November 1, 2012		San Diego, CA, USA	October-November 2012			yes	
Evaluation of Interactive User Corrections for Lecture Transcription	Henrich Kolkhorst, Kevin Kilgour, Sebastian Stüker, Alex Waibel	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
The KIT-NAIST (Contrastive) English ASR System for IWSLT 2012	Michael Heck, Keigo Kubo, Matthias Sperber, Sakriani Sakti, Sebastian Stüker, Christian Saam, Kevin Kilgour, Christian Mohr, Graham Neubig, Tomoki Toda, Satoshi Nakamura, Alex Waibel	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
The 2012 KIT and KIT-NAIST English ASR Systems for the IWSLT Evaluation	Christian Saam, Christian Mohr, Kevin Kilgour, Michael Heck, Matthias Sperber, Keigo Kubo, Sebastian Stüker,	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	

	Sakriani Sakti, Graham Neubig, Tomoki Toda, Satoshi Nakamura, Alex Waibel									
Segmentation and Punctuation Prediction in Speech Language Translation Using a Monolingual Translation System	Eunah Cho, Jan Niehues, Alex Waibel	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
The KIT Translation systems for IWSLT 2012	Mohammed Mediani, Yuqi Zhang, Thanh-Le Ha, Jan Niehues, Eunah Cho, Teresa Herrmann, Alex Waibel	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	
Continuous Space Language Models using Restricted Boltzmann Machines	Jan Niehues, Alex Waibel	Proceedings of the International Workshop for Spoken Language Translation (IWSLT 2012)	December 6-7, 2012		Hong Kong	Dec-12			yes	

TABLE A2.1: LIST OF CONFERENCES AND WORKSHOPS

NO.	Type of activities[2]	Main leader	Title	Date/Period	Place	Type of audience[3]	Size of audience	Countries addressed
1	Conference		SLT: IEEE Spoken Language Technology Workshop	December 6-9, 2014	Lake Tahoe, California, US	Scientists		Worldwide
2	Conference		IWSLT: 11th International Workshop on Spoken Language Translation	December 4-5, 2014	Lake Tahoe, California, US	Scientists		Worldwide
3	Conference		MLMI: 16th International Conference on Multimodal Interaction	November 2-16, 2014	Istanbul, Turkey	Scientists		Worldwide
4	Conference		EMNLP: Conference on Empirical Methods in Natural Language Proceedings	October 25-29, 2014	Doha, Qatar	Scientists		Worldwide
5	Conference		SPECOM:	October 5-9,	Novi Sad,	Scientists		Worldwide

[2] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[3] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

			16th International Conference on Speech and Computer	2014	Serbia			
6	Conference		PolTAL: 9th International Conference on Natural Language Processing	September 17-19, 2014	Warsaw, Poland	Scientists		Worldwide
7	Conference		Interspeech: 15th Annual Conference of the International Speech Communication Association	September 14-18, 2014	Suntec City, Singapore	Scientists		Worldwide
8	Conference		MTM: 9th Machine Translation Marathon	September 9-14, 2014	Trento, Italy	Scientists		Worldwide
9	Conference		FA: 7th Forum Acusticum	September 7-12, 2014	Krakow, Poland	Scientists		Worldwide
10	Conference		TSD: 17th International Conference on Text, Speech and Dialogue	September, 8-12, 2014	Brno, Czech Republic	Scientists		Worldwide
11	Conference		ECCV: European Conference on Computer Vision	September 6-12, 2014	Zürich, Switzerland	Scientists		Worldwide
12	Conference		COLING: 25th International	August 23-29, 2014	Dublin, Ireland	Scientists		Worldwide

			Conference on Computational Linguistics					
13	Conference		STATMT: 9th Workshop on Statistical Machine Translation	June 26-2, 2014	Baltimore, US	Scientists		Worldwide
14	Conference		ACL-WMT: 52th Annual Meeting of the Association for Computational Linguistics	June 22-27, 2014	Baltimore, US	Scientists		Worldwide
15	Conference		EAMT, MT-Summit: 15th Conference of the European Association for Machine translation	June 16-18, 2014	Dubrovnic, Croatia	Scientists		Worldwide
16	Conference		LREC: 9th international conference on Language Ressources and Evaluation	May 26-31, 2014	Reykjavik, Iceland	Scientists		worldwide
17	Conference		SLTU: 4th International Workshop on Spoken Language Technologies for Under-resourced Languages	May 14-16, 2014	St. Petersburg, Russia	Scientists		Worldwide

18	Conference		ICASSP: IEEE International Conference on Acoustics, Speech and Signal Processing	May 4-9, 2014	Florence, Italy	Scientists		Worldwide
19	Conference		EACL: 14th Conference of the European Chapter of the Association for Computational Linguistics	April 26-30, 2014	Gothenburg, Sweden	Scientists		Worldwide
20	Conference		WorldCIST: World Conference on Information Systems and Technologies	April 15-18, 2014	Madeira, Portugal	Scientists		Worldwide
21	Conference		AISV: Aspetti prosodici e testuali del raccontare	April 15-18, 2014	Madeira, Portugal	Scientists		Worldwide
22	Conference		ASRU: Automatic Speech Recognition and Understanding Workshop	December 7-9, 2013	Olomouc, Czech Republic	Scientists		Worldwide
23	Conference		LTC: 6th Language & Technology Conference	December 7-9, 2013	Poznan, Poland	Scientists		Worldwide

24	Conference		IWSLT: 10th International Workshop on Spoken Language Translation	December 5-6, 2013	Heidelberg, Germany	Scientists		Worldwide
25	Conference		EMNLP: Conference on Empirical Methods in Natural Language Processing	October 18-21, 2013	Seattle, US	Scientists		Worldwide
26	Conference		MTM: 8th Machine Translation Marathon	September 9-14, 2013	Prague, Czech Republic	Scientists		Worldwide
27	Conference		EAMT, MT Summit: 14th MT Summit, organised by EAMT	September 2-6, 2013	Nice, France	Scientists		Worldwide
28	Conference		TSD: 16th International Conference on Text	September 1-5, 2013	Czech Republic	Scientists		Worldwide
29	Conference		Interspeech: 14th Annual Conference of the International Speech Communication Association	August 25-29, 2013	Lyon, France	Scientists		worldwide
30	Conference		ACL-WMT: 8th Workshop on	August 8-9, 2013	Sofia, Bulgaria	Scientists		Worldwide

			Statistical Machine Translation					
31	Conference		AISV: Multimodalità: la Sfida più Avanzata della Comunicazione Orale 9° Convegno Nazionale	January 21-23, 2013	Venezia, Italy	Scientists		Worldwide
32	Conference		COLING: 24th International Conference on Computational Linguistics	December 8-15, 2012	Mumbai, India	Scientists		Worldwide
33	Conference		IWSLT 2012: 9th Inaternational Workshop on Spoken Language Translation	December 6-7, 2012	Hong Kong	Scientists		worldwide
34	Conference		AMTA: 10th Biennial Conference of the Association for Machine Translation in the Americas	October 28 – November 1, 2012	San Diego, USA	Scientists		Worldwide
35	Conference		HLT: Human Language Technology Day	September 27-28, 2012	Warsaw, Poland	Scientists		Worldwide
36	Conference		Interspeech: 13th Annual	September 9-13, 2013	Portland, Oregon, US	Scientists		Worldwide

			Conference of the International Speech Communication Association					
37	Conference		MTM: 7th Machine Translation Marathon	September 3-8, 2012	Edingburgh, Scotland	Scientists		Worldwide
38	Conference		TSD: 15th International Conference on Text, Speech and Dialogue	September 3-7, 2012	Brno, Czech Republic	Scientists		Worldwide
39	Conference		WMT: 7th Workshop on Statistical Machine Translation	June 7-8, 2012	Montreal, Canada	Scientists		Worldwide
40	Conference		EAMT: 16th Annual Conference of the European Assoc. for Machine Translation	May 28-30, 2012	Trento, Italy	Scientists		Worldwide
41	Conference		LREC: 8th international conference on Language Resources and Evaluation	May 21-27, 2012	Istanbul, Turkey	Scientists		Worldwide

TEMPLATE A2.2: LIST OF FAIRS AND EXHIBITIONS								
NO.	Type of activities[4]	Main leader	Title	Date/Period	Place	Type of audience[5]	Size of audience	Countries addressed
1	Exhibition	RBM	National Association of Broadcasters (NAB)	April 11-16, 2015	Las Vegas, US	Media and entertainment professionals	100.000	worldwide
2	Exhibition	KIT	CeBIT 2015	March 16-19, 2015	Hannover, Germany	Computer scientists, industry, general public	201.000	worldwide
3	Exhibition	KIT	Lust auf Technik	Nov. 20-23, 2014	Stuttgart, Germany	General public	• 180.000	Germany
4	Exhibition	RBM/KIT	International Broadcast Convention (IBC)	Sep. 11-16, 2014	Amsterdam, The Netherlands	Media and entertainment professionals	55.000	Worldwide
5	Exhibition	KIT	HLT-village @ LREC	May 28-30, 2014	Reykjavik, Iceland	Scientists	1200	Worldwide
6	Exhibition	RBM	National Association of Broadcasters (NAB)	April 05-10, 2014	Las Vegas, US	Media and entertainment professionals	100.000	Worldwide
7	Exhibition	KIT	ICT	Nov. 06-08,	Vilnius,	Experts	4000	Europe

[4] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[5] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

				2013	Lithuania	around digital technology		
8	Exhibition	KIT	CeBIT	March 05-09, 2013	Hannover, Germany	Computer scientists, industry, general public	285.000	Worldwide
9	Exhibition	KIT	HLT-village @ LREC	May 23-25, 2012	Istanbul, Turkey	Scientists	1000	Worldwide

TABLE A2.3: LIST OF TV AND RADIO CLIPS								
NO.	Type of activities[6]	Main leader	Title	Date/Period	Place	Type of audience[7]	Size of audience	Countries addressed
	BBC world (radio)	KIT	Simultaneous Translation of University Lectures	March 25, 2015	UK	General public		Worldwide

TABLE A2.4: LIST OF FLYERS								
NO.	Type of activities[8]	Main leader	Title	Date/Period	Place	Type of audience[9]	Size of audience	Countries addressed
	Flyer	KIT	Invitation postcard for EU-BRIDGE	Jan. 2015	Germany	Scientists, industry, media, policy		Worldwide

[6] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[7] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

[8] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[9] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

			Technology Day			makers		
	Newsletter	KIT	EU-BRIDGE Newsletter (special topic: IWSLT)	Dec. 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Newsletter	KIT	EU-BRIDGE Newsletter (special topic: events)	Dec. 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Flyer	KIT	EU-BRIDGE Technology Catalogue	Dec. 2014	Germany	Scientists, industry, media, politicians		Worldwide
	Flyer	KIT	EU-BRIDGE, 3rd edition	Dec. 2014	Germany	Scientists, industry, media, politicians		Worldwide
	Flyer	KIT	Invitation postcard for IBC	Sep. 2014	Germany	Industry, scientists		Worldwide
	Fact Sheet "Technology Support for High-quality Automatic Speech Recognition Engines"	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet "Automatic Transcription: Audio into Text in Real Time"	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet "Euronews: A	KIT	Fact Sheet	July 2014	Germany	Scientists, industry,		Worldwide

	Multilingual ASR Benchmark”					media, policy makers		
	Fact Sheet “Technology Support for High-quality Speech Translation Engines”	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet “ Statistical Machine Translation”	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet “Open Source Statistical Machine Translation”	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet “Polish Spoken Language Translation”	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet “Punctuation”	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet “Service Architecture”	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet “Automatic Simultaneous Translation Service for University	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide

	Lectures"							
	Fact Sheet "Automated Captioning of Multimedia Content"	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet "European Parliament Interpreter Support"	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet "Automatic Simultaneous Translation Service for Voting Sessions"	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Fact Sheet "Serenty a Webinar Platform for Enhanced Multi-lingual Business Communication"	KIT	Fact Sheet	July 2014	Germany	Scientists, industry, media, policy makers		Worldwide
	Flyer (German version)	KIT	EU-BRIDGE, 3rd edition	July 2014	Germany	General public		Worldwide
	Newsletter	KIT	EU-BRIDGE Newsletter (special topic: IWSLT)	Dec. 2013	Germany	Scientists, industry, media, policy makers		Worldwide
	Newsletter	KIT	EU-BRIDGE Newsletter (special topic: ICT Vilnius)	Nov. 2013	Germany	Scientists, industry, media, policy makers		Worldwide
	Flyer	KIT	EU-BRIDGE,	Oct. 2013	Germany	Scientists, industry,		Worldwide

			2nd edition			media, politicians		
	Flyer (French version)	KIT	EU-BRIDGE, 3rd edition	Oct. 2013	Germany	General public		Worldwide
	Flyer	KIT	EU-BRIDGE, 3rd edition	Oct. 2013	Germany	General public		Worldwide
	Flyer	KIT	EU-BRIDGE 2nd edition	2013	Germany	General public		Worldwide
	Newsletter	KIT	2nd EU-BRIDGE Newsletter	2012	Germany	Scientists, industry, media, policy makers		Worldwide
	Newsletter	KIT	1st EU-BRIDGE Newsletter	2012	Germany	Scientists, industry, media, policy makers		Worldwide
	Flyer	KIT	EU-BRIDGE 1st edition	2012	Germany	Scientists, industry, media, politicians		Worldwide
	Flyer	KIT	EU-BRIDGE 1st edition	April 2012	Germany	General public		Worldwide
	Flyer	KIT	EU-BRIDGE fact sheet	March 2012	Germany	Scientists, industry, media, politicians		Worldwide

TABLE A2.5: LIST OF MEDIA BRIEFINGS AND INTERVIEWS

NO.	Type of activities[10]	Main leader	Title	Date/Period	Place	Type of audience[11]	Size of audience	Countries addressed
	Interview	KIT	FAZ weekend	March 2015	Telephone	Media		Europe
	Interview	KIT	BBC	March 2015	Telephone	Media		Worldwide
	Interview	KIT	dpa	March 2015	Karlsruhe	Media		Worldwide
	Interview	KIT	Die Zeit	January 2015	Telephone	Media		
	Media Briefing	KIT	FKT	Sep. 2014	Telephone	Media		Europe
	Media Briefing	KIT	During IBC	Sep. 2014	Amsterdam	Media		Europe
	Interview	KIT	dpa, BNN, Stuttgarter Zeitung, SWR	June 2012	Karlsruhe	Media		Worldwide

TABLE A2.6: LIST OF POSTERS

[10] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[11] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

NO.	Type of activities[12]	Main leader	Title	Date/Period	Place	Type of audience[13]	Size of audience	Countries addressed
1	Poster	KIT	EU-BRIDGE	March 2012		General public, scientists		Worldwide
2	Poster	KIT	EU-BRIDGE (update)	August 2013		General public, scientists		Worldwide
3	Poster	KIT	EU-BRIDGE use cases	June 2014		Scientists		Worldwide

TABLE A2.7: LIST OF PRESENTATIONS

NO.	Type of activities[14]	Main leader	Title	Date/Period	Place	Type of audience [15]	Size of audience	Countries addressed
	Presentation	KIT	EU-BRIDGE Technology Day	March 2015	Hannover, Germany	Scientists, Media, Industry,		worldwide

[12] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[13] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

[14] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[15] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

						Policy makers		
	Presentation	KIT	Interpreter Support	March 2015	European Parliament, Strasbourg, France	Policy makers		Europe
	Presentation	KIT	„Simultaneous translation of university lectures“ (Deutsch-japanisches Forum)	March 2015	KIT, Germany	Scientists, Industry		Japan
	Presentation	HKUST	“Why Structural Relationships Between Human Representation Languages Are Efficiently Learnable: The Magic Number 4”. HKU Spring Symposium, Science of Learning.	Feb. 2015	Hong Kong	Scientists		worldwide
	Presentation	KIT	Voting-session	Dec. 2014	European Parliament, Strasbourg, France	Policy makers		Europe
	Presentation	UEDIN	„Learning representations of speech in neural network acoustic models“, IEEE GlobalSIP Symposium on Machine Learning Applications in	Dec. 2014	Atlanta, USA	Scientists		Worldwide

			Speech Processing,					
	Presentation	FBK/ KIT	IWSLT Evaluation	Dec. 2014	Lake Tahoe, US	Scientists		Worldwide
	Presentation	PJIIT	„Problemy, metody i obliczenia wielkoskalowe oraz wyzwania modelowania inżynierskiego i biznesowego”	Dec. 2014	Warsaw Technical University	Scientists		worldwide
	Presentation	KIT	Keynote at ICMI 2014	Nov. 2014	Istanbul, Turkey	Scientists		worldwide
	Presentation	KIT	Support for Interpreters	Nov. 2014	European Parliament, Strasbourg, France	Policy makers		Europe
	Presentation	KIT	„Automatic simultaneous translation“, EP-Rectors-Conference	Nov. 2014	European Parliament, Brussels Belgium	Policy makers		Europe
	Presentation	KIT	Automatic simultaneous translation (Wirtschaftsjunionren Karlsruhe)	Oct 2014	KIT, Germany	Industry		Germany
	Presentation	KIT	Automatic simultaneous translation	Oct 2014	KIT, Germany	Industry		Europe
	Presentation	KIT	Simultaneous translation of university lectures (Universities from China)	Oct 2014	KIT, Germany	Scientists		China
	Presentation	KIT	Automatic simultaneous	Oct 2014	KIT, Germany	Scientists		US

			translation of university lectures (Universities from the US)					
	Presentation	KIT	Breaking the language barrier (TAUS Conference)	Oct 2014	Vancouver, Canada	Scientists		Worldwide
	Presentation	PJIIT	„Spoken Language Translation for Polish“, Forum Acousticum 2014,	Aug 2014	Krakow, Poland	Scientists		Worldwide
	Presentation	HKUST	“Augmenting Human Communication: What Doesn't Translate, and What is the Cost of Not Translating?”. SummerDavos, World Economic Forum—Annual Meeting of the New Champions	Sep. 2014	Tianjin, China,	Scientists		worldwide
	Presentation	KIT	„Breaking the speech and language barrier“, (Leopoldina Jahrestagung)	Sep. 2014	Rostock, Germany	Scientists, Industry, Media, Policy makers		Worldwide
	Presentation	KIT	„EU-BRIDGE“ (Press- and communication officers of KIT)	July 2014	KIT, Germany	Scientists		Germany
	Presentation	UEDIN	„Neural Networks for Distant Speech Recognition“, HSCMA-2014 Workshop,	May 2014	Nancy, France	Scientists		Worldwide

	Presentation	PJIIT	„Statistical speech translation and recognition (with special focus on Polish)“, Samsung R&D	May 2014	Warszawa, Poland	Scientists		worldwide
	Presentation	PJIIT	„Prozodia w semantyce - semantyka w prozodii, Statystyczne tłumaczenie mowy“ ((w szczególności polskiej)	May 2014	Warszawa, Poland	Scientists		worldwide
	Presentation	KIT	„Bridging the language divide“ (Keynote at LREC)	May 2014	Reykjavik, Iceland	Scientists, policy makers, media		worldwide
	Presentation	KIT	„Automatic Translation“ (Joint EP/EC interpreter/translator event)	March 2014	European Parliament, Brussels Belgium	Policy makers		Europe
	Presentation	HKUST	“The GAGO Principle”. QTLeap.,	March 2014	Lisbon, Portugal	Scientists		worldwide
	Presentation	KIT	„Breaking the speech and language barrier“ (Waseda-Symposium)	March 2014	Tokyo, Japan	Scientists		Japan
	Presentation	KIT	„The lecture translator“ (Forum Anthropomatik und Robotik)	Jan 2014	KIT, Germany	Scientists		Germany
	Presentation	KIT	„Breaking the language barrier“ (Ministry of	January 2014	Stuttgart, Germany	Policy makers		Germany

			Science, Technology and the Arts, Baden-Württemberg					
	Presentation	FBK/ KIT	IWSLT Evaluation	Dec 2013	Heidelberg, Germany	Scientists		Worldwide
	Presentation	KIT	„Bridges across the language divide“ (STOA workshop)	Dec 2013	European Parliament, Brussels Belgium	Policy makers		Europe
	Presentation	KIT	„EU-BRIDGE networking session“ (ICT, Vilnius)	Nov 2013	Vilnius, Lithuania	Scientists, Policy makers		Europe
	Presentation	HKUST	“Semantic SMT Without Hacks”. 4th Workshop on South and Southeast Asian NLP (WSSANLP)	Oct 2013	Nagoya, Japan	Scientists		worldwide
	Presentation	UEDIN	„(Deep) Neural Networks for Speech Recognition“, UK-Speech Conference	Sep 2013	Cambridge UK	Scientists		UK
	Presentation	HKUST	“Re-Architecting The Core: What SMT Should Be Teaching Us About Machine Learning”. Recent Advances in Natural Language Processing (RANLP)	Sep 2013	Hissar, Bulgaria	Scientists		worldwide
	Presentation	KIT	Keynote (Interspeech)	Aug 2013	Lyon, France	Scientists		Worldwide
	Presentation	UEDI	„Neural Networks	June 2013	Mons	Scientist		Worldwide

		N	for Speech Recognition“, Non-linear Speech Processing Workshop		Belgium,	s		
	Presentation	KIT	Keynote at ICRA	May 2013	Karlsruhe	Scientists		Worldwide
	Presentation	UEDIN	„Multi-domain acoustic modelling for speech recognition“, Microsoft Machine Learning Summit	April 2013	Paris France	Scientists		Worldwide
	Presentation	PJIIT	„Statystyczne tłumaczenie mowy polskiej – wstępne eksperymenty“, NLP seminar Institute of Computer Science,	March 2013	Warszawa, Poland	Scientists		Worldwide
	Presentation	KIT	"Socially Aware Ineractive Assistants" (German Speechtechnologyday)	Jan 2013	Berlin, Germany	Scientists, Policy makers		Germany
	Presentation	FBK/ KIT	IWSLT Evaluation	Dec 2012	Hong Kong	Scientists		Worldwide
	Presentation	KIT	„Automatic simultaneous translation“, EP-Rectors-Conference	Oct. 2012	European Parliament, Brussels Belgium	Policy makers		Europe
	Presentation	PJIIT	„Od rozpoznawania do tłumaczenia mowy polskiej“, HLT-days	Sep 2012	Poland	Scientists		worldwide

	Presentation	KIT	„Bridging the language divide“ (10th anniversary Carnegie Mellon University, Silicon Valley Campus)	June 2012	California, US	Scientists, Policy makers, Industry		US
	Presentation	UEDIN	„Natural speech technology“, JST CREST Symposium on Human-Harmonized Information Technology	April 2012	Kyoto Japan	Scientists		Worldwide

TABLE A2.8: LIST OF PRESS ARTICLES

NO.	Type of activities[16]	Main leader	Title	Date/Period	Place	Type of audience[17]	Size of audience	Countries addressed
1	Press articles, online		Vorlesungsübersetzer: “Very, very, very many possible word follow”	March 22, 2015	Spiegel online, Germany	General public		Germany
2	Press articles		Nix verstanden? Macht nix!	March 16, 2015	Stuttgarter Nachrichten, Germany	General public		Germany
3	Press		Der	March 14,	Badisches Tagblatt,	General		Germany

[16] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[17] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

	articles		elektronische Studentenversteher	2015	Germany	public		
4	Press articles		Unverständliches wird verständlich	March 14, 2015	Badische Neueste Nachrichten, Germany	General public		Germany
5	Press articles, online		Der Studentenversteher: Uni der Zukunft mit Vorlesungsübersetzer	March 13, 2015	dpa, Germany	General public		Worldwide
6	Press articles		Sie haben verstanden	March 5, 2015	Die Zeit, No. 10, Germany	General public		Germany
7	Press articles		Überwindung der Sprachbarrieren	February, 2015	FKT, Germany	General public		Germany
8	Press articles, online		Forschungsprojekte unter der Lupe: vom Fahrsimulator bis zum Übersetzungssystem	November 19, 2014	mese-stuttgart.de, Germany	General public		Germany
9	Press articles, online		Anche FBK nello studio delle traduzioni automatiche con EU-Bridge	June 12, 2014	lavocedeltrentino.it, Italy	General public		Italy
10	Press articles		Übersetzungsprojekt "EU-Bridge" präsentiert sich: Mit Computerhilfe die Sprachbarrieren überwinden	October 10, 2013	EU-Nachrichten, No. 16, Germany	General public		Europe

11	Press articles, online		Confindustria e Fbk: innovazione per le aziende	July 23, 2013	Trentino	General public		Italy
12	Press articles		Vorlesungen werden per Rechner simultan übersetzt	August 3, 2012	Staatsanzeiger Baden-Württemberg	General public		Germany
13	Press articles, online		German university to stream subtitled lectures	June 25, 2012	Deutsche Welle.de, Germany	General public		Worldwide
14	Press articles		New Automatic Translator Could End Language Barrier in Lectures	June 12, 2012	The Chronicle, US	General public		US
15	Press articles		Schon Deutsch versteht der Computer nur schwer	June 12, 2012	Stuttgarter Zeitung, Germany	General public		Germany
16	Press articles		Live-Übersetzer für Vorlesungen	June 12, 2012	Die Welt, Germany	General public		Germany
17	Press articles		Live-Übersetzer für Vorlesungen	June 12, 2012	Bote vom Untermain, Germany	General public		Germany
18	Press articles		Mehr als nur Bahnhof verstehen – Weltweit erster Vorlesungsüber setzer	June 12, 2012	dpa, Germany	General public		Europe
19	Press articles		Weltneuheit aus dem KIT	June 12, 2012	BNN, Germany	General public		Germany
20	Press articles		Ein Computer als Simultan-Dolmetscher	June 12, 2012	BNN, Germany	General public		Germany
21	Press		Mehr als nur	June 12,	Ka-news, Germany	General		Germany

	articles , online		Bahnhof verstehen – Weltweit erster Vorlesungsüber setzer am KIT	2012		public		
22	Press articles , online		Uni-Übersetzungs-Automat	June 12, 2012	Spiegel online, Germany	General public		Germany
23	Press articles , online		Vorlesungsüber setzer	June 12, 2012	N24, Germany	General public		Germany
24	Press articles , online		Mobile Technologies Unveils World's First Real-Time Lecture Translation System	June 12, 2012	Globalmedianews	General public		Worldwide
25	Press articles , online		KIT stellt System zur automatischen Vorlesungsüber setzung vor	June 12, 2012	Portel, Germany	General public		Germany
26	Press articles , online		Live-Übersetzer für Vorlesungen	June 12, 2012	Main-Netz	General public		Germany
27	Press articles , online		Ein Computer als Dolmetscher	June 11, 2012	Stuttgarter Zeitung, Germany	General public		Germany
28	Press articles , online		Mehr als nur Bahnhof verstehen – weltweit erster Vorlesungsüber	June 11, 2012	dpa, Germany	General public		Worldwide

			setzer					
29	Press articles, online		Computer übersetzt Vorlesungen am KIT	June 11, 2012	SWR3 online, Germany	General public		Germany
30	Press articles, online		Simultane Übersetzung: Lehre ohne Sprachbarrieren	June 11, 2012	Informationsdienst Wissenschaft, Germany	General public		Worldwide
31	Press articles, online		Mehr als nur Bahnhof verstehen – weltweit erster Vorlesungsüber setzer am KIT	June 11, 2012	Reutlinger General-Anzeiger	General public		Germany
32	Press articles, online		Mehr als nur Bahnhof verstehen – weltweit erster Vorlesungsüber setzer	June 11, 2012	Südwest Presse, Germany	General public		Germany
33	Press articles, online		Erster Vorlesungsüber setzer für Studenten entwickelt	June 11, 2012	Hamburger Abendblatt, Germany	General public		Germany
34	Press articles, online		Mehr als nur Bahnhof verstehen – weltweit erster Vorlesungsüber setzer	June 11, 2012	Schwäbische Zeitung, Germany	General public		Germany
35	Press articles, online		Simultane Übersetzung: Lehre ohne Sprachbarrieren	June 11, 2012	Technologiewerte, Germany	General public		Germany
36	Press articles		Mobile Technologies	June 11, 2012	Virtualizationconferenc e, US	General public		US

	, online		Unveils World's First Real-Time Lecture Translation System					
37	Press articles , online		Mobile Technologies Unveils World's First Real-Time Lecture Translation System	June 11, 2012	The Vancouver Sun, US	General public		US
38	Press articles , online		Menschlicher Übersetzer (noch) nicht überflüssig	June 11, 2012	Badisches Tagblatt, Germany	General public		Germany
39	Press articles , online		Mobile Technologies Unveils World's First Real-Time Lecture Translation System	June 11, 2012	FinanzNachrichten, Germany	General public		Germany
40	Press articles , online		Mobile Technologies Unveils World's First Real-Time Lecture Translation System	June 11, 2012	Businesswire, US	General public		Worldwide
41	Press articles , online		KIT präsentiert weltweit ersten Vorlesungsüber setzer	June 11, 2012	Pforzheimer-Zeitung (PZ) News, Germany	General public		Germany
42	Press articles , online		Mobile Technologies Unveils World's First Real-Time	June 11, 2012	Marketwatch, US	General public		US

			Lecture Translation System					
43	Press articles, online		Mobile Technologies Unveils World's First Real-Time Lecture Translation System	June 11, 2012	Live PR, US	General public		US
44	Press articles, online		Mehr als nur Bahnhof verstehen weltweit erster Vorlesungsüber setzer	June 11, 2012	CIO online, Germany	General public		Germany
45	Press articles, online		Simultane Übersetzung: Lehre ohne Sprachbarrieren	June 11, 2012	Uni-Protokolle, Germany	General public		Germany
46	Press articles, online		Technischer Meilenstein für den Hörsaal	June 11, 2012	Mainpost, Germany	General public		Germany
47	Press articles, online		Mehr als nur Bahnhof verstehen – weltweit erster Vorlesungsüber setzer präsentiert	June 11, 2012	Südtirol online, Italy	General public		Italy
48	Press articles, online		Weltweit erster Vorlesungsüber setzer entwickelt	June 11, 2012	Die-Mark-Online, Germany	General public		Germany
49	Press articles		Technologie: Institut lässt per Computer	June 4, 2012	Die Rheinpfalz, Germany	General public		Germany

			übersetzen					
50	Press articles		University without Language Barriers	June 2012	APRnews Newsletter of the KIT Focus Anthropomatics and Robotics	General public		Germany

TABLE A2.9: LIST OF PRESS RELEASES

NO.	Type of activities[18]	Main leader	Title	Date/Period	Place	Type of audience[19]	Size of audience	Countries addressed
1	Press release	KIT	CeBIT: Security in a Smart World (KIT)	March 05, 2014		Media		
2	Press release	KIT	IBC: Breaking the Speech and Language Barrier in Media	September 2014		Media		
3	Press release	KIT	CeBit: IT hilft bei Operationen, Pflege und Übersetzung	February 26, 2013		Media		
4	Press release	KIT	Simultane Übersetzung: Lehre ohne Sprachbarrieren	June 11, 2012		Media		
5	Press release	KIT	”EU-BRIDGE”: Bridges across the language	February 6, 2012		Media		

[18] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[19] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

			divide					
6	Press release	KIT	“EU-BRIDGE”: Sprachbarrieren überbrücken	Feburary 6, 2012		Media		
7	Press release	PJIIT	“EU-BRIDGE”: Mosty ponad podziałami językowymi	February 6, 2012		Media		

TABLE A2.10: LIST OF THESES

NO.	Type of activities[20]	Main leader	Title	Date/Period	Place	Type of audience [21]	Size of audience	Countries addressed
	Thesis	*KIT*	Linguistic Structure in Statistical Machine Translation (PHD)	*March 2015*	*Germany, Karlsruhe*	*Scientists*		Worldwide
	Thesis	*RWTH*	"Phrase-Based Lattice Decoding and its Application in Machine Translation" (Bachelor thesis,)	*March 2015*	*Aachen, Germany*	*Scientists*		Worldwide
	Thesis	*PJIIT*	Statistical machine	*2015*	*Krakow,*	*Scientists*		Worldwide

[20] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[21] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

			translation enhanced by comparable corpora (PHD)		*Poland*			
	Thesis	KIT	Source Sentence Reordering for English to Japanese Machine Translation (Bachelor Thesis)	November 2014	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Rule-Based Preordering on Multiple Syntactic Levels in Statistical Machine Translation (Master Thesis)	August 2014	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Analyse von Methoden zum Graphemclustering bei der automatischen Spracherkennung (Bachelor Thesis)	May 2014	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	An Optimization of Deep Neural Networks in ASR Using Singular Value Decomposition (Bachelor Thesis)	May 2014	*Karlsruhe, Germany*	*Scientists*		Worldwide

	Thesis	KIT	Deep Neural Network Language Models for Low Resource Languages (Bachelor thesis)	April 2014	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Pronominal Anaphora in Machine Translation (Master thesis)	January 2014	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Language Model Adaptation using Temporal Information (Master thesis)	January 2014	*Germany, Karlsruhe*	*Scientists*		Worldwide
	Thesis	KIT	Adaptation in Machine Translation (PHD)	January 2014	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Automatische Segmentierung und Gruppierung natürlicher Sprache anhand verschiedener Sprecher (Master thesis)	December 2013	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	RWTH	Improved Optimization of Neural Networks" (Master thesis)	Dec 2013	*Aachen, Germany*	*Scientists*		Worldwide

	Thesis	KIT	Letter N-Gram-based Input Encording for Continuous Space Language Models (Master thesis)	June 2013	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Discriminative Maximum Entropy Language Model in the Context of Large-Vocabulary Speech Recognition for Russian (Master thesis)	June 2013	*Germany, Karlsruhe*	*Scientists*		Worldwide
	Thesis	RWTH	"Comparison of Feedforward and Recurrent Neural Network Language_Models" (Diploma thesis)	June 2013	*Aachen, Germany*	*Scientists*		Worldwide
2	Thesis	FBK	Linguistically Motivated Reordering Modeling for Phrase-Based Statistical Machine Translation (PHD)	April 2013	Trento, Italy	Scientists		Worldwide
	Thesis	RWTH	„Pruning Strategies for	April 2013	*Aachen,*	*Scientists*		Worldwide

			Phrase-based Statistical Machine Translation" (Bachelor thesis)		*Germany*			
3	Thesis	RWTH	"Lexicalized Reordering Models for Phrase-Based Statistical Machine Translation" (Bachelor thesis)	Feb. 2013	*Aachen, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Unsupervised Acoustic Model Training for Simultaneous Lecture Translation in Incremental and Batch Mode (Master thesis)	December 2012	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Extending Phrase-Based Machine Translation with Topic Models (Master thesis)	December 2012	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Training Deep Neural Networks for Bottleneck Feature Extraction (Master Thesis)	December 2012	*Karlsruhe, Germany*	*Scientists*		Worldwide

	Thesis	KIT	Automatic Segmentation and Summarization of Spoken Lectures (Master Thesis)	November 2012	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	A Study of Distance Measures for Clustering Generalized Polyphones (Bachelor thesis)	November 2012	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	Temporal Patterns (TRAPs) in Janus Recognition Toolkit (Bachelor thesis)	October 2012	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	KIT	High-Accuracy Frequency, Phase and Amplitude Estimation for Robust Speech Recognition (Master thesis)	August 2012	*Germany, Karlsruhe*	*Scientists*		Worldwide
	Thesis	KIT	Japanese-English Machine Translation for a Humanoid Robot	July 2012	*Karlsruhe, Germany*	*Scientists*		Worldwide

			Moderator (Bachelor thesis)					
	Thesis	KIT	Getting Bilingual Information from the Web (Bachelor thesis)	May 2012	*Karlsruhe, Germany*	*Scientists*		Worldwide
	Thesis	RWTH	"Discontinuous Phrases for Statistical Machine Translation" (Bachelor thesis)	March 2012	*Aachen, Germany*	*Scientists*		Worldwide

TABLE A2.11: LIST OF VIDEOS

NO.	Type of activities[22]	Main leader	Title	Date/Period	Place	Type of audience[23]	Size of audience	Countries addressed
	Video	*RWTH*	*Voting Session*	*September 2014*	*Germany*	*Scientists, Media, General public,*		Worldwide

[22] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[23] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

						Politicians		
	Video	*FBK*	*Euronews*	*August 2014*	*Italy*	*Scientists, Media, General public, Politicians*		Worldwide
	Video	ADX	Unified Communication Technologies	May 2014	France	*Scientists, Media, General public, Politicians*		Worldwide
	Video	RBM	Captioning and translation of subtitles for TV programs	May 2014	UK	*Scientists, Media, General public, Politicians*		Worldwide
	Video	KIT	The Lecture Translator	June 2012	Germany	*Scientists, Media, General public, Politicians*		Worldwide
	Video	KIT	EU-BRDIGE (update)	June 2013	Germany	*Scientists, Media, General public, Politicians*		Worldwide
	Video	KIT	EU-BRDIGE	July 2012	Germany	*Scientists, Media, General public, Politicians*		Worldwide

TABLE A2.12: LIST OF WEBSITES								
NO.	Type of activities[24]	Main leader	Title	Date/Period	Place	Type of audience[25]	Size of audience	Countries addressed
1	Website	PEV	http://eubridge.pervoice.com/EBPortal/	January 2015		*General public, Scientific Community, Industry, Policy Makers, Media*		
2	Website	KIT	http://eu-demo.ira.uka.de/#/subtitles	December 2014		*General public, Scientific Community, Industry, Policy Makers, Media*		
3	Website	*KIT*	*www.eu-bridge.eu*	*March 2012*		*General public, Scientific Community, Industry, Policy Makers, Media*		

[24] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[25] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).