Annual Public Report - Project Year 2

November 2012

Grant Agreement number: 247762

Project acronym: FAUST

Project title: Feedback Analysis for User Adaptive Statistical Translation

Funding Scheme: FP7-ICT-2009-4 STREP

Period covered: from 1 February 2010 to 15 November 2011

Project coordinator :

Dr William Byrne

Reader in Information Engineering

Department of Engineering,

University of Cambridge, UK

Tel: +44 (0)1223 332651

Fax: +44 (0)1223 332662

E-mail: bill.byrne@eng.cam.ac.uk

Project website address: http://faust-fp7.eu

SEVENTH FRAMEWORK PROGRAMME

This report summarizes work within the FAUST project midway through its second year.

FAUST aims to develop fluent, web-based MT systems that respond to user feedback.

Web-based machine translation systems are now readily available in many of the world's major and minor languages. The FAUST project was motivated by perceived shortcomings of a fundamental nature in the technological approaches to delivering machine translation to general populations of users. We elaborate here on the two key aspects of our project's aim.

Feedback As translation technology is brought ever closer to its community of users, there is strong potential for creating collaborative interaction between translators, casual active users, and technology developers. For example, on the Reverso.net website (http://www.reverso.net), which translates an average of 30 million text passages each month, users are invited to provide feedback and to suggest improvements to the automatic translation of any given sentence. Unfortunately this feedback cannot yet be exploited because:

•User feedback tends to be very noisy;

•No research published to date makes explicit how statistical translation systems can be adapted to benefit from feedback provided by web users;

•No mechanisms exist to identify user feedback of value and for immediately modifying a statistical MT system so that subsequent users do not run into the same problem.

‣FAUST aims to address these problems to 'close the loop' so that user feedback becomes part of the development and evaluation cycle for machine translation systems deployed in online translation.

Fluency Automatic systems that make basic mistakes in grammar and word sense are perceived as unreliable and unpleasant to use. We take the view that MT systems must become fluent if they are to be accepted and trusted by large communities of users.

‣FAUST aims to improve user satisfaction with online MT by bringing natural language generation into statistical machine translation to improve MT fluency.

The project is organized to address the following technical objectives

1.Enhance the high-volume, Reverso.net translation website with an experimental and evaluation infrastructure for the study of instantaneous user feedback in MT.

2.Deploy novel web-oriented, feedback collection mechanisms that reduce noise in feedback provided by users and increase the utility of the web contributions.

3.Automatically acquire data collections to study translation with user feedback.

4.Develop mechanisms for instantaneously incorporating user feedback into the machine translation engines that are used in production environments, such as those that power the Reverso.net website.

5.Create novel automatic metrics of translation quality that reflect preferences learned from user feedback.

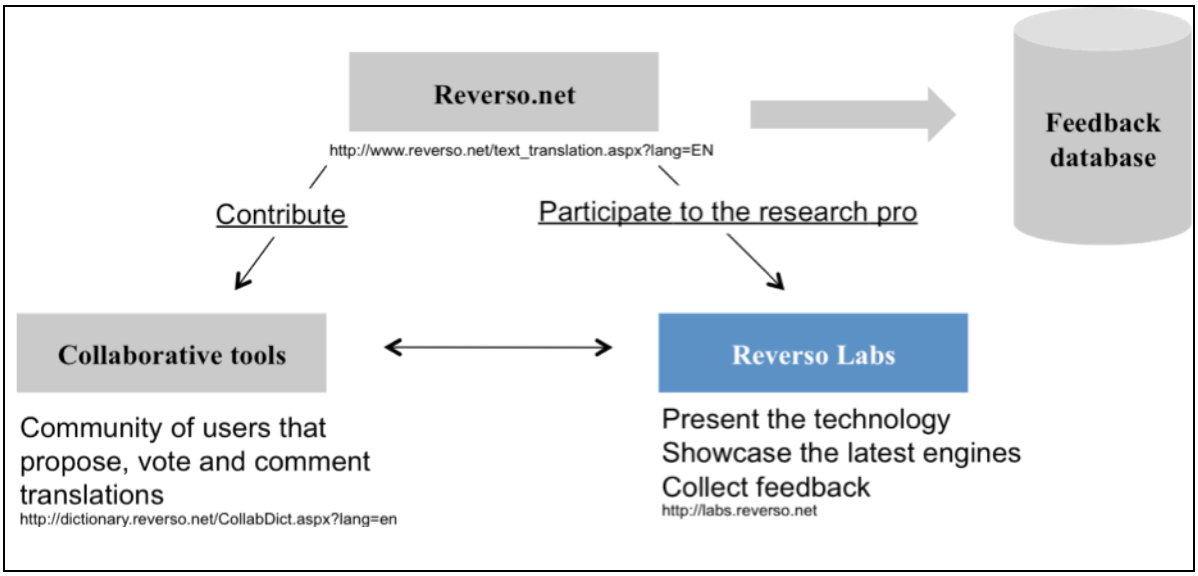
6.Develop translation models driven by user feedback data and integrate natural language generation into MT to improve translation fluency and reduce negative feedback from users.

The project has a focus on the following language pairs: Czech-English ; French-English ; Romanian-English ; Spanish-English ; Spanish-Catalan ; and Arabic and Chinese -> English. We have assembled a team of machine translation researchers from academic and industry:

FAUST Project Academic Partners	**FAUST Project Commercial Partners**
- Department of Engineering, University of Cambridge, UK - Computer Laboratory, University of Cambridge, UK - Center for Language and Speech Technologies and Applications (TALP), Universitat Politecnica de Catalunya, Spain - Institute of Formal and Applied Linguistics, Charles University, Czech Republic	- Language Weaver Inc., USA - Language Weaver SRL, Romania - Softissimo, Inc., France

Interactive translation FAUST is developing new interfaces to the Reverso.net translation services that will provide access to MT systems developed within the project. Users are interacting with systems on http://labs.reverso.net where research MT systems are deployed directly so that researchers can observer users interacting directly with MT systems.

This environment is tightly linked with the Reverso website so that user interaction can be monitored and logged for study. The Labs website is now online and MT systems from the project partners are providing translation services freely to the public.



The Reverso Labs environments will support interaction between MT users and researchers

The project has been organized into Work Packages, according to the interest and expertise of the project partners:

- WP1 – Project Coordination and Management
- WP2 – System Architecture and Integration
- WP3 – Web-Oriented Feedback Collection
- WP4 – Automatic Acquisition, Annotation, and Modeling of User Feedback
- WP5 – User-driven MT Systems
- WP6 – Evaluation and Dissemination

We now focus on descriptions of the technical work packages WP2-6, since these are likely to be of the most general interest.

<u>WP2 – System Architecture and Integration -- Language Weaver SRL</u>

Effort in this work package has focused on developing the common infrastructure to support the research and development efforts throughout the project. The main objectives for this period are:

- Development of an architecture that will permit the evaluation and exploitation of novel web-oriented, feedback collection mechanisms that reduce noise in feedback provided by users and increase the utility of the web contributions.

- The creation of mechanisms for instantaneous incorporation of user feedback into 24x7 operational translation engines.

- The creation of seamless flows and bridges between the production, research, and evaluation environments.

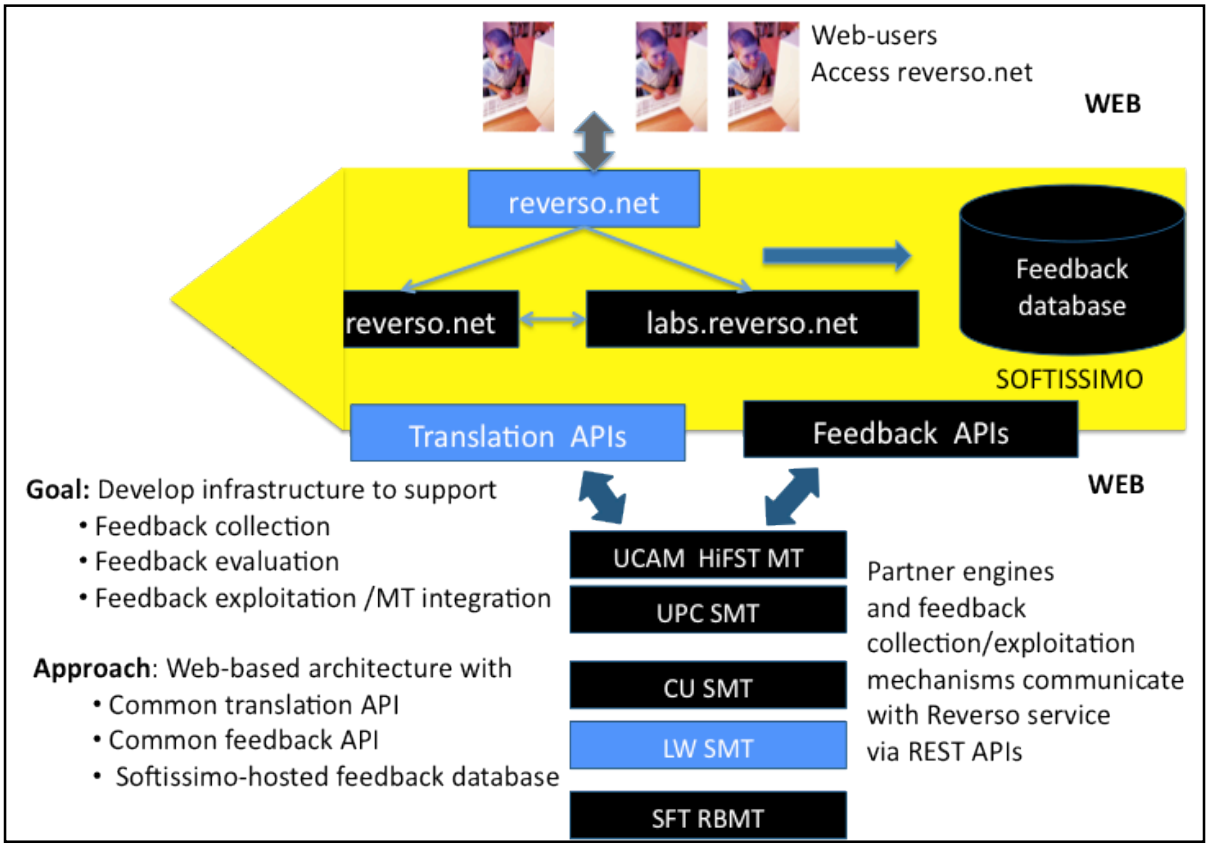
Progress in this work package is necessarily technical in nature, but key achievements concern the development of APIs for integration of SMT systems into the labs.reverso.net environment as well as APIs for collecting and sharing user feedback:

•The Technical Architecture was conceived and delivered to the partners by month 3.

• The Translation APIs among all sites and components have been created, documented, reviewed and approved by all consortium members.

•Access to the fully operational SMT infrastructure has been provided to the partners with the use of the commonly agreed API; the Language Weaver translation infrastructure is fully operational and integrated into the FAUST platform located at http://Reverso.net and at http://labs.reverso.net .

This figure gives a high-level summary of the development issues faced by WP2:



Infrastructure for user feedback collection and deployment of research MT systems

Feedback will be provided to all partners within the project. In addition, a special API and real-time communication channels between the reverso.net website, labs.reverso.net and the SaaS environment at Language Weaver has been developed. These channels is enabling the feedback provided by users to flow to the Language Weaver engines.

Language Weaver has developed 'Term and Brand Management' API that ensures certain terms and/or brands are correctly and consistently used throughout translations. The Term and Brand Management API, is part of the infrastructure that will enable the rapid incorporation of the good feedback into the Translation engines. The API will connect the applications created to filter the good feedback and extract the dictionary entries to the SMT.

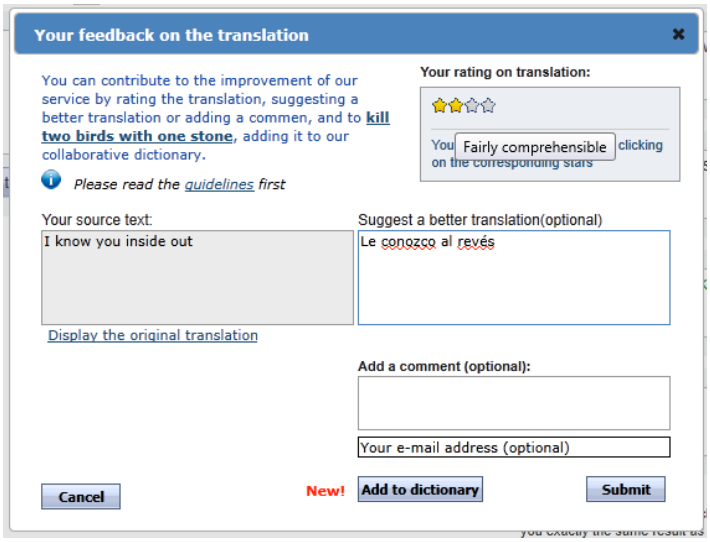
WP3 - Web-Oriented Feedback Collection Mechanisms -- Softissimo Inc

The Labs.reverso.net website allows users to compare translations provided by the partners' engines, see their back-translation, rate, suggest another translation and give a comment on each of them. The Labs.reverso.net is linked to the partners' MT engines. For the time being, 5 language pairs are provided, namely English-French (by UCAM, LW and Softissimo), English-Spanish (UPC, LW, UCAM and Softissimo), English-Romanian (LW), English-Czech (CU) and Spanish-Catalan (UPC). From the home page, users have several options:

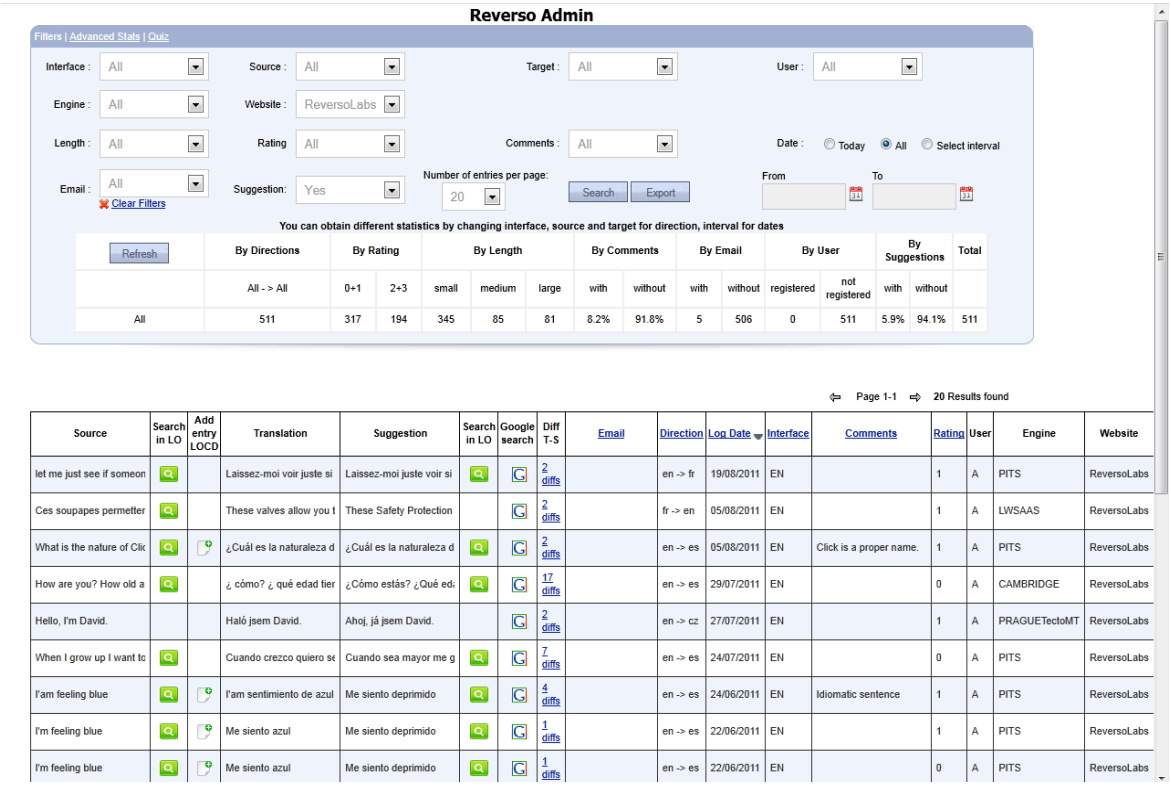
- See by default only one translation engine.
- Otherwise, Display all translation engines in case of the translations being different
- Highlight the differences between 2 translations

An example is shown in the following figure:

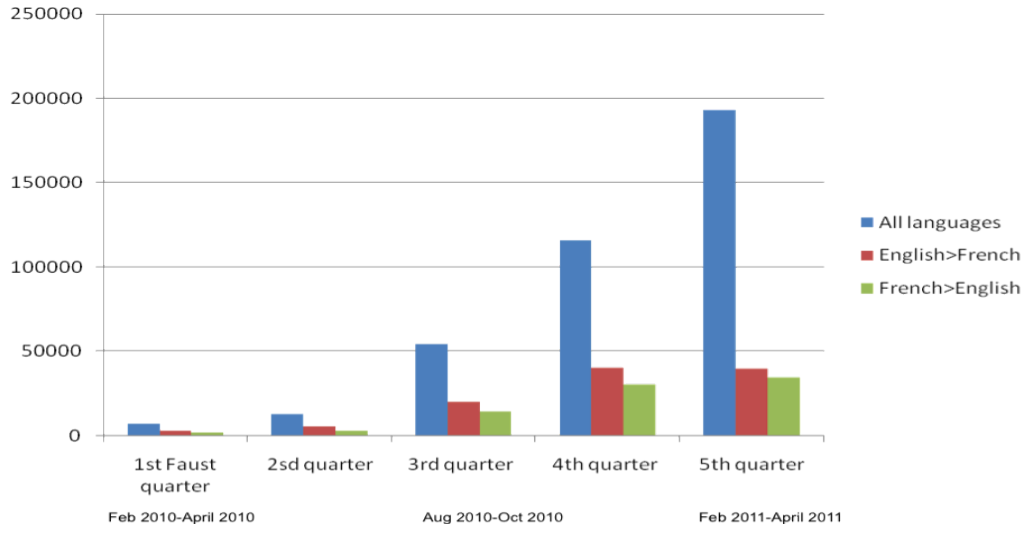
Feedback can be entered via a modal feedback window:



Softissimo has also implemented interfaces to the SQL databases which store translation exchanges and feedback. This is available to all project partners and it is designed to enable feedback filtering on various criteria, such as interface, engine used, time, rating, source/target languages, etc. A screen shot of the interface is in the next figure.



A summary of the feedback traffic at reverso.net to April 2011 is in the following graph.



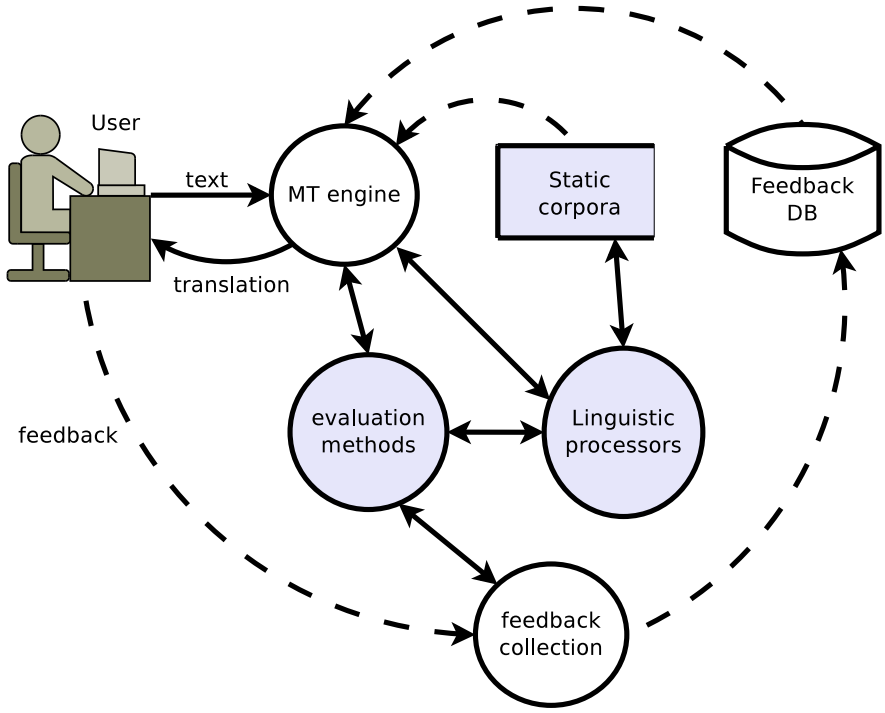
WP4 - Acquisition, Annotation, and Modelling of Feedback -- Universitat Politècnica de Catalunya

This work package has the aim of analyzing feedback provided by users:

•Develop methods for automatically annotating and analyzing translation requests and SMT system outputs

•Produce an open source suite of automatic metrics for SMT.

Produce linguistic annotation tools to be used throughout the project. This includes the development of novel parsing techniques for analyzing noisy translation output



The role of feedback modelling within the FAUST project

The above figure shows, highlighted in blue, the role of feedback annotation and modeling within the project. The aim is to abstract from the data we collect the information needed to characterize the task and to refine our MT engines.

Some highlights from the current year of the project are:

• We have gathered linguistically-annotated corpora (i.e., treebanks) for the various languages involved (English, Spanish, Catalan, French, Czech, and Romanian). We are using these data to train our linguistic processors, and these processors have now been trained and are working well for regular text (i.e., well-formed sentences). One of the challenges in the FAUST project is to deal with possibly poorly-formed sentences, e.g., arising from user inputs or system outputs. For this reason we are adapting linguistic processors to the analysis of noisy text.

• We have compiled a static corpora collection to be used for training and testing our SMT Systems. We have prepared these corpora and performed its basic annotation (tokenization, part-of-speech tagging and lemmatization) for English, Spanish, French, Czech, and Catalan. Extensively annotated data collections are now available at : http://www.faust-fp7.eu/faust/Main/DataReleases

• We have developed the ASIYA Open Toolkit for Automatic Machine Translation Evaluation and Meta-evaluation. The Asiya toolkit is the natural evolution/extension of its predecessor, the IQMT Framework. Asiya is publicly available at http://nlp.lsi.upc.edu/asiya. The developments carried out within the first 18 months of the FAUST project are the following:

> • Refactoring of the previous IQMT Framework to create Asiya, which uses more standard formats and a modular architecture that facilitates the development of new methods:
>
> • Incorporation of meta-evaluation methods
>
> • Update of the *-granularity* option (Section 3.3), which now incorporates the metrics to obtain document-level scores
>
> • Extension of shallow syntax-based metrics to Catalan and Spanish languages
>
> • Incorporation of a set of confidence (or quality) estimation measures
>
> • Creation of the *-learn* option, which now uses a ranking perceptron to learn how to combine quality estimation measures
>
> • Development of the new Asiya website at http://nlp.lsi.upc.edu/asiya and the preliminary version of the on-line prototype.

• A corpus of adequacy assessments for real-world machine translation output is being created. Taking the FAUST feedback corpus as source, we have maually annotated, for 5 alternative translations of the same source sentences, all pairwise quality rankings and the absolute quality scores. These two layers of annotation are complementary and useful in different ways, and they will be exploited to learn models of translation quality in the project. To do the annotation we are use a methodology based on hierarchical graphs that allows us to save a considerable amount of manual effort.

•We have filtered and analyzed the first body of 50,000 user feedback instances collected through the Reverso portal. We have carried out two manual annotations on a portion of it. One aimed at identifying suggestions that, at least in part, improve upon the automatic translations. The other was

focused on the selection of error-free suggestions. The corpus and the resulting study are publicly available.

WP5 - User-driven MT systems -- Charles University

This work package is focused on the development of MT systems which respond to user-feedback:

•Create baseline MT systems for all language pairs in the project

•MT system training and adaptation to user feedback

•MT confidence for guiding user feedback

Baseline systems have been developed for all the language pairs of interest to the project and these systems are being connected to the labs.reverso.net website. There are a mix of commercial systems, such as the Language Weaver and Softissimo systems already running at Reverso.net, and academic systems based on phrase-based, Hiero, and TectoMT technologies. These systems are now stable and being used throughout the project.

Some highlights from the current year of the project are:

- Improvement of baseline MT systems by using additional language models built from target language public web translation requests.
- Creation of a feedback filtering mechanism that filters useful userfeedback based on linear binary classification techniques, using a variety of simple surface-level features.
- Creation of an automatic term dictionary augmentation mechanism that improves translation quality by applying standard unsupervised statistical machine translation learning methods to user feedback to identify mistranslated phrases and their corrections.

WP6 – Evaluation and Dissemination -- University of Cambridge

The objectives of this work package within the current project year are

•Common and consistent use of resources

•Consistent evaluation of all component technologies

•Internal evaluation of integrated systems, research and real-time demonstrators

•Demonstrate that the quality of the research within the project is at the state-of-the art by participation in international evaluations of translation technology

•Release Open Source modelling and analysis tools developed within the project

Some highlights of the project so far are :

• Specification of the size and selection criteria for the English-to-Czech/French/Romanian/Spanish development test sets. There was extensive discussion prior to data set selection by Language Weaver from web logs provided by Softissimo. Source segments in Czech, English, French, Romanian, and Spanish have been randomly selected from translation requests submitted to the LW Translation On-Demand Infrastructure during the months of Jan-Feb 2010.

• Distribution and support for translation software within the project. LW has provided translation support software through a modified version of Olifant.

• Development of translation guidelines and conventions for translation from English. Development of these guidelines was an iterative process guided by early feedback from the translators working with the project partners.

• Collection and initial release of the Open Source annotation tools which will be used within the project. Links to these tools are available at the FAUST website: http://www.faust-fp7.eu/faust/Main/PublicTools .

• An initial release of the user-feedback corpus has been made available at the FAUST website http://www.faust-fp7.eu/faust/Main/DataReleases . This is a large sampling of the weblogs from Reverso.net collected as of 20 October 2010 -- the initial release is 11.3MB compressed. This release contains entries filtered so that the source text (i.e. the original translation request) is no more than twenty words in length; no other processing was done to this data. We will continue to add to this corpus over the course of the project, as more user transactions data becomes available. We also intend to refine the release by using the tools we develop to identify and include the most relevant and useful material.

Outreach Activities

Publications in the technical literature are a key part of our dissemination strategy. A comprehensive list is available at http://www.faust-fp7.eu/faust/Main/FaustPapers .

FAUST has also been described in the following venues:

•Language Technology Days : 22-23 March 2010, Luxembourg, W Byrne

•DISI seminar series organized by the Language, Speech, and Interfaces group, University of Trento: 17 September 2010, L Màrquez. This presentation was also attended by staff from Fondazione Bruno Kessler .

• On 19 July 2011, The FAUST project was featured on the BBC World Service Radio Programme `Click'. BBC World Service English has total weekly reach of 43m listeners. An extended version of the interview broadcast on Click on BBC World Service Radio, 19th July 2011 is available on the The Open University website.

- BBC Programme website: http://www.bbc.co.uk/programmes/p00hvfgc
- Open University Website: http://bit.ly/np6iVV

We have participated in the international evaluation campaign associated with the ACL 2010 Fifth Workshop on Statistical Machine Translation (http://www.statmt.org/wmt10/) organized by the ACL special interest group for MT (http://www.sigmt.org) . This involved three shared tasks: a shared translation task for eight pairs of European languages; a shared evaluation task to test automatic evaluation metrics; and a system combination task combining the output of all the systems entered into the shared translation task. Systems submitted by FAUST researchers were judged as very competitive under the various evaluation criteria.