PROGRESS REPORT

PUBLISHABLE SUMMARY

Grant Agreement number: 270915

Project acronym: BOLOGNA

Project title: Bologna Translation Service

Project type: ☐ **Pilot A** ☑ **Pilot B** ☐ **TN** ☐ **BPN**

Periodic report: **1st** ☐ **2nd** ☑ **3rd** ☐ **4th** ☐

Period covered: **from March 2012 to February 2013**

Project coordinator name, title and organisation:

Heidi Depraetere
Director
CrossLang

Tel: +32 498 90 65 91

Fax: +32 9 335 22 01

E-mail: heidi.depraetere@crosslang.com

Project website address: http://www.bologna-translation.eu

Project Objectives

There is a continuing increasing need for educational institutes to provide course syllabi documentation and other educational information in English. Access to translated course syllabi and degree programmes plays a crucial role in the degree to which universities effectively attract foreign students and, more importantly, has an impact on international profiling. To present all education information in English is a major challenge for most higher education institutes. Figures and trends show that investment in traditional human translation services is prohibitive, consequently course materials and degree programmes are often provided in the local language only.

The Bologna project aims at providing a solution to this problem by offering a low-cost, web-based, high-quality machine translation (MT) service, the Bologna Translation Service (BTS). The project covers 9 languages (Chinese, Dutch, English, Finnish, French, German, Portuguese, Spanish and Turkish). Existing statistical as well as rule-based solutions have been tailored to the educational domain and are offered through the translation portal. The service is accessible from a web-interface and can be made accessible from the working environments of the universities, typically from the database where course syllabi and degree programmes are produced and edited, by means of a web service API.

Project Progress

Areas of focus in the second year have been the revamping of the portal interface, the continued development and evaluation of advanced machine translation systems for all languages, the creation of automatic post-editing modules for all languages, and the exploration of exploitation options for the service.

Revamping Portal Interface

During the first year, a basic front-end had been developed which tied together the various components of the Bologna Translation Service platform. The focus of the development activities in year one was mostly on the technical integration of the components. By the end of year one, we were able to demonstrate how translations could be requested through the portal and processed by the back-end components.

During the second year remaining planned features were implemented, but more importantly, focus shifted to improving the user experience of the end-user. A usability expert company was hired to collect and document the requirements in a field study. Look-and-feel as well as information flows in the portal were subject to a thorough reassessment, in which also user group members were involved.

The main recommendations from the study involved a simplified registration flow, a redesign of the collaboration and post-editing environment, and a simplification of the organizational structure. A prototype was built that implemented these recommendations and the prototype was presented to a selected set of user for a final usability test. The fine-tuned prototype then formed the basis for the redesign of the portal. Figure 1 and Figure 2 give an impression of the new design.

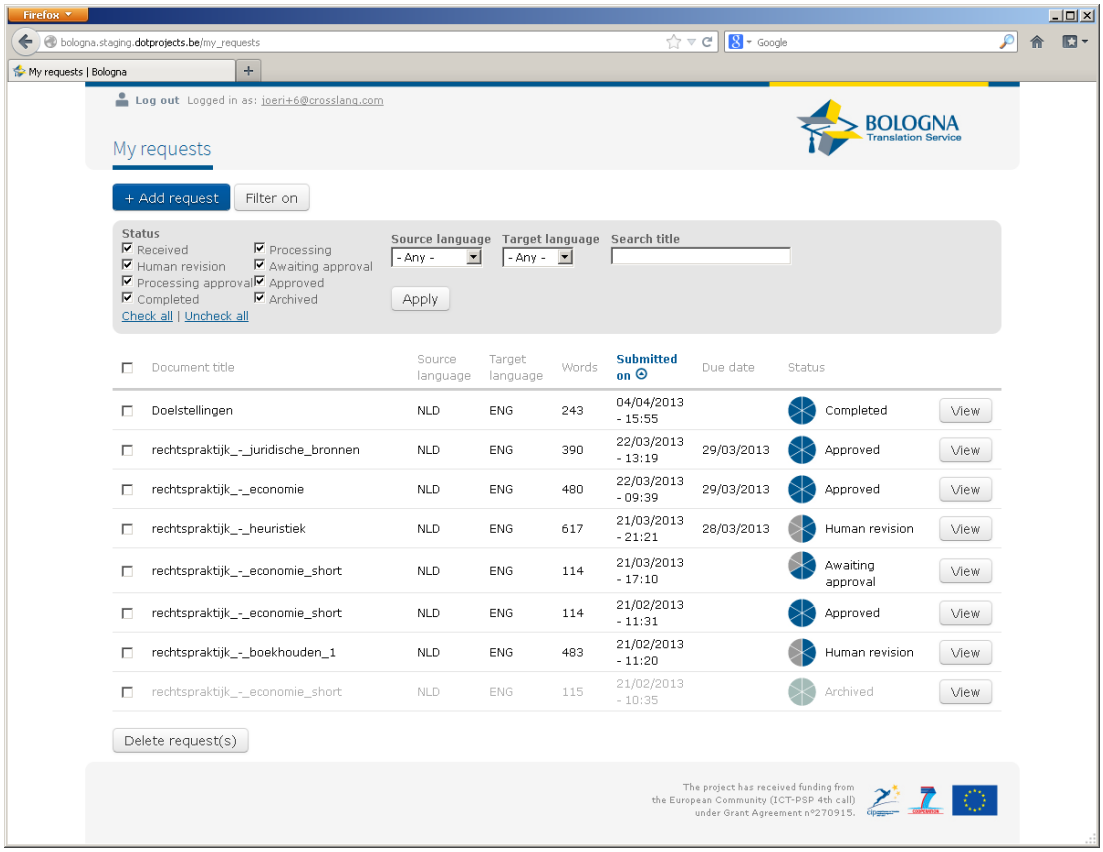

Figure 1: Request Overview in Redesigned Collaboration Portal

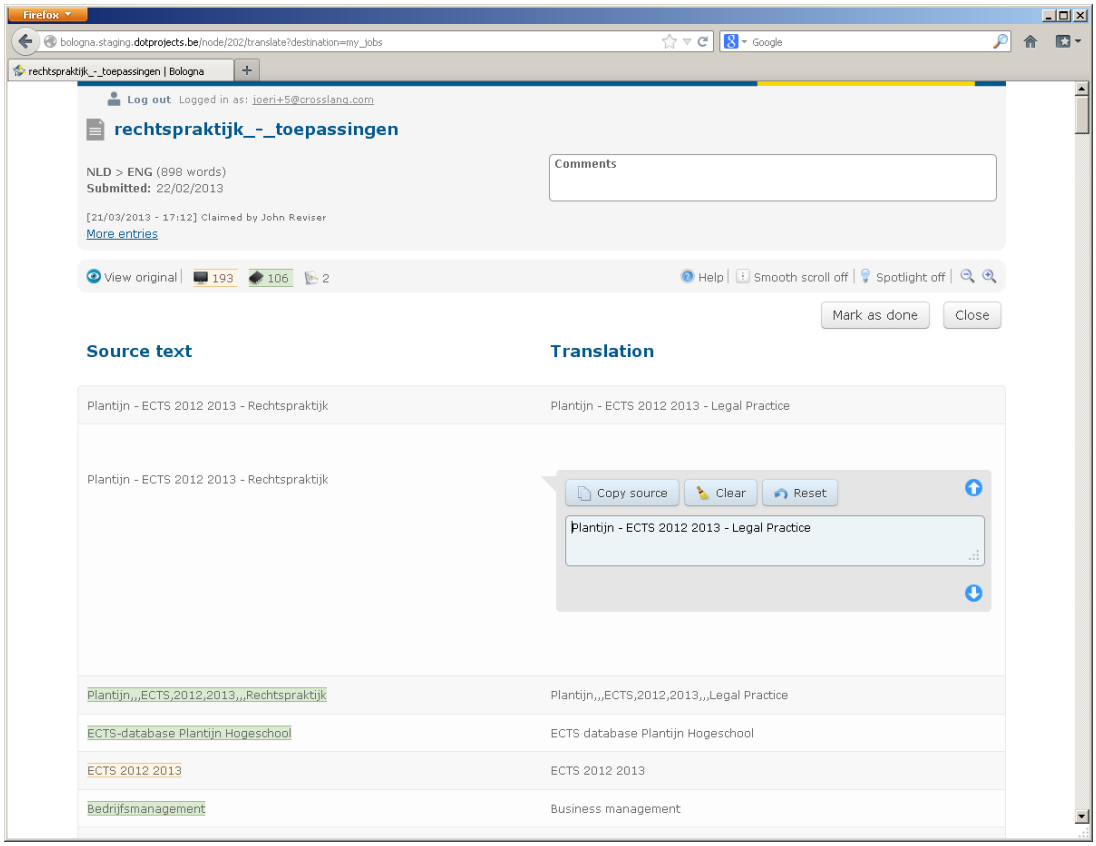

Figure 2: Post-editing Environment in Redesigned Collaboration Portal

Machine Translation Systems

In the course of the project, different types of systems with different levels of quality were built. The first systems that were built were what we called the baseline systems, i.e. systems trained with general domain data. These systems were used as benchmarks to compare the systems against that we would build in year two of the project.

In year two advanced systems were built, i.e. systems that were trained with the education-specific data that was collected in the course of year one. Various experiments were tried, e.g. training with in-domain data only vs. training with a combination of in-domain and general domain data.

Language Pair	Advanced			Baseline		
	BLEU	METEOR	TER	BLEU	METEOR	TER
German—English	43.4	35.0	48.4	22.3	26.4	63.8
French—English	51.1	42.9	32.9	40.8	38.6	40.2
Spanish—English	34.8	37.7	45.1	30.0	34.7	49.5
Portuguese—English	40.1	37.3	45.7	29.9	33.5	52.0
Turkish—English	26.7	28.9	61.2	18.9	26.2	68.6
Finnish—English	42.8	35.1	48.4	15.2	20.3	76.0
Dutch—English	35.8	32.6	54.3	16.9	23.7	65.0
English—Chinese	67.4	48.0	25.8	20.9	25.9	65.3

Table 1 below shows the best scores that were obtained per language pair. We include the baseline scores of last year for reference.

Language Pair	Advanced			Baseline		
	BLEU	METEOR	TER	BLEU	METEOR	TER
German—English	43.4	35.0	48.4	22.3	26.4	63.8
French—English	51.1	42.9	32.9	40.8	38.6	40.2
Spanish—English	34.8	37.7	45.1	30.0	34.7	49.5
Portuguese—English	40.1	37.3	45.7	29.9	33.5	52.0
Turkish—English	26.7	28.9	61.2	18.9	26.2	68.6
Finnish—English	42.8	35.1	48.4	15.2	20.3	76.0
Dutch—English	35.8	32.6	54.3	16.9	23.7	65.0
English—Chinese	67.4	48.0	25.8	20.9	25.9	65.3

Table 1: Automatic Evaluation Scores (Advanced vs. Baseline)

As Table 1 indicates we were able to improve on the baseline scores for all languages. Automatic evaluations were complemented by human evaluations with which we tried to assess the potential of the engines as aids for gisting and boosting translation productivity. These evaluations suggested that all but two engines (German—English and English—Chinese) were suited for gisting purposes and that the translation output of four systems (Spanish—English, Dutch—English, German—English, and French—English) increases translation productivity.

Automatic Post-editing

As there are limits to the quality improvements that can be achieved by just increasing the amounts of training data, the consortium also explored other ways of boosting the translation quality. Therefore, another area of focus in the second year of the project was the creation of modules for automatic post-editing of the MT output. The idea behind automatic post-editing is that you look for recurring mistakes in the output and formulate rules for correcting those mistakes as post-processing step to the output.

Post-editing modules were built for all SMT systems and all RBMT systems that were built in this project. Different types of post-editing rule sets (rule-based rules and statistical rules) were created and applied to the various types of systems. Best results were obtained by combining the RBMT systems with the statistical rules with BLEU score improvements up to 20 BLEU points. The improvements obtained from the application of the rule-based rule components was found to be only marginal, with BLEU score improvements ranging from 0.2 to 2.2 BLEU points. More significant improvements are expected, though, when rule-based APE modules will be created from real post-edited data. The APE modules built during the project were derived from human reference translations that were not the result of post-editing. Because the edit distance between MT output and this type of human reference translations is expected to be bigger than the edit distance between MT output and its corresponding post-edited output, it was expected that the impact of the rules would be limited. We are convinced that the impact will be more significant when rules can be extracted from human translations that are the result of post-editing. Such data will only become available once users start making use of the BTS platform.

Exploitation

In view of exploitation of the Bologna Translation Service past the project's end date, exploitation discussions between partners were started early in the second year of the project. Three out of the four commercial partners showed an interest in continuing the service beyond the project's life cycle, i.e. CrossLang, Convertus, and Eleka. Capita Translation and Interpreting, previously ALS decided to discontinue their involvement.

The respective sales channels of CrossLang, Convertus and Eleka will be used for exploiting the project results. A commercial agreement between these three partners has been set up.

Other Highlights

Dissemination

In year two of the project, partners intensified their efforts to make the Bologna Translation Service known in the European educational space by visiting more educational events across Europe. Encouraged by the positive response received at the EAIE Conference in Copenhagen, Denmark, last year, partners participated at similar events with a presentation, a poster, or a booth.

Date	Event	Location
March 2012	INTED 2012 (International Technology, Education and Development Conference)	Valencia, Spain
May 2012	EAMT (European Association for Machine Translation)	Trento, Italy
May 2012	LREC 2012 (Language Resources and Evaluation Con-	Istanbul, Turkey

	ference)	
June 2012	MosesCore workshop	Paris, France
June 2012	LOCALIZATION WORLD 2012	Paris, France
July 2012	EDULEARN 2012 (Education and New Learning Technologies)	Barcelona, Spain
September 2012	EAIR 2012 Forum (European Higher Education Society)	Stavanger, Norway
September 2012	EAIE 2012 (European Association for International Education)	Dublin, Ireland
November 2012	ICERI 2012 (International Conference of Education, Research and Innovation)	Madrid, Spain
November 2012	Nordic Translation Industry Forum	Copenhagen, Denmark
November 2012	Translating and the Computer	London, UK
March 2013	INTED 2013 (International Technology, Education and Development Conference)	Valencia, Spain

Table 2: Overview of Attended Conferences

Collaterals including a project fact sheet, a flyer, a poster and a number of press releases have been distributed to help raise awareness of the Bologna project. The project website (www.bologna-translation.eu), which has been live since March 2011, is Bologna's main dissemination channel.

Through the website we have had 95 applications for membership of the Bologna User Group.

More than 20 workshops took place at local universities in Belgium, Turkey and Spain. During these workshops users were introduced to the Bologna project and to the current state of the Bologna translation platform. Once the platform had reached a level where it could be released to a broader audience, workshop participants also got a chance to be hands-on during the workshop and register to test more.

The portal (demo.bologna-translation.eu) is now live with dedicated instances for 4 user group members giving a total of 34 users access to the BTS platform

Project Meetings

The Bologna consortium has held a number of on-site meetings at various stages to address technical issues and discuss progress. These meetings have proven to be very productive and have helped speed up progress as well as encourage cooperation between partners. Whereas during the first year, these meetings were more operation-oriented, during the second year they focused more on matters relating to exploitation.

Outlook

The Bologna platform is now available for prospective users. Active dissemination will continue after the project comes to an end. The Bologna Translation Service will be represented with a booth at the main international educational conference, EAIE taking place in Istanbul, September 2013.