Publishable Summary

Project objectives

The objective of the PRESEMT project is to develop a flexible and adaptable MT system, based on a language-independent method, which is easily portable to new language pairs. This method attempts to overcome well-known problems of other MT approaches, e.g. bilingual corpora compilation or creation of new rules per language pair. PRESEMT will address the issue of effectively managing multilingual content and is expected to suggest a language-independent machine-learning-based methodology.

The key aspects of PRESEMT involve syntactic phrase-based modelling, pattern recognition techniques towards the development of a language-independent analysis and evolutionary algorithms for system optimisation. PRESEMT is intended to be of a hybrid nature, combining linguistic processing with the positive aspects of corpus-based approaches, such as SMT and EBMT. In order for PRESEMT to be easily amenable to new language pairs, relatively inexpensive, readily available language resources as well as bilingual lexica will be used. The translation context will be modelled on phrases, as they have been proven to improve the translation quality. Phrases will be produced via a semi-automatic and language-independent process of morphological and syntactic analysis, removing the need for compatible NLP tools per language pair.

Parallelisation of the main translation processes will be investigated in order to reach a fast, high-quality translation system. Furthermore, the optimisation and personalisation of the system parameters via automated processes (such as genetic algorithms or swarm intelligence) will be studied. To allow for user adaptability, all the corpora used in PRESEMT will be retrieved from web-based sources. User feedback will be integrated through the use of appropriate interactive interfaces. PRESEMT is expected to be easily customisable to both new language pairs and specific sublanguages.

Work performed since the project start

The work performed by the partners within the context of the PRESEMT project since the start of the project involves the following axes:

∗ **Definition of system specifications:** In this area, the system specifications have been finalised, together with the set-up of the evaluation activities.

∗ **Technical development of the prototype:** In this area, work carried out involves the design and implementation of PRESEMT modules (this process is currently underway) and the collection of the required generally-available as well as dedicated resources and tools.

∗ **Dissemination activities:** A number of dissemination activities have been carried out, these including the launch of the project website, the preparation of dedicated factsheets and presentations and the participation in relevant conferences to inform the community of the PRESEMT project.

∗ **Management:** Within this subject, the entire set of scheduled-for-completion deliverables have been finalised and submitted to the European Commission, while a number of bilateral and project-wide meetings have been held to facilitate the progress of the project.

Main results achieved so far

During the first five months of the PRESEMT project, the main results achieved relate to the design of the PRESEMT system and the methodology for assessing it.

More specifically, the consortium has consolidated the system architecture (cf. Figure 1) and has defined the constituent system modules, their mode(s) of operation and their interaction. In addition, a UML scheme has been generated to facilitate the future development of the PRESEMT prototype software.

In a similar vein, an internal document has been produced, detailing the current view of the consortium on the implementation of the PRESEMT main translation-related algorithmic modules, namely the Phrase aligner and the two phases of the Main translation engine.

Another major result achieved during the current reporting period involves the full specification of the validation and evaluation activities, which will be carried out near the end of the project and concern the assessment of the PRESEMT system in terms of performance & conformance to the system design principles (validation) and of translation quality (evaluation).

Figure 1: PRESEMT system architecture
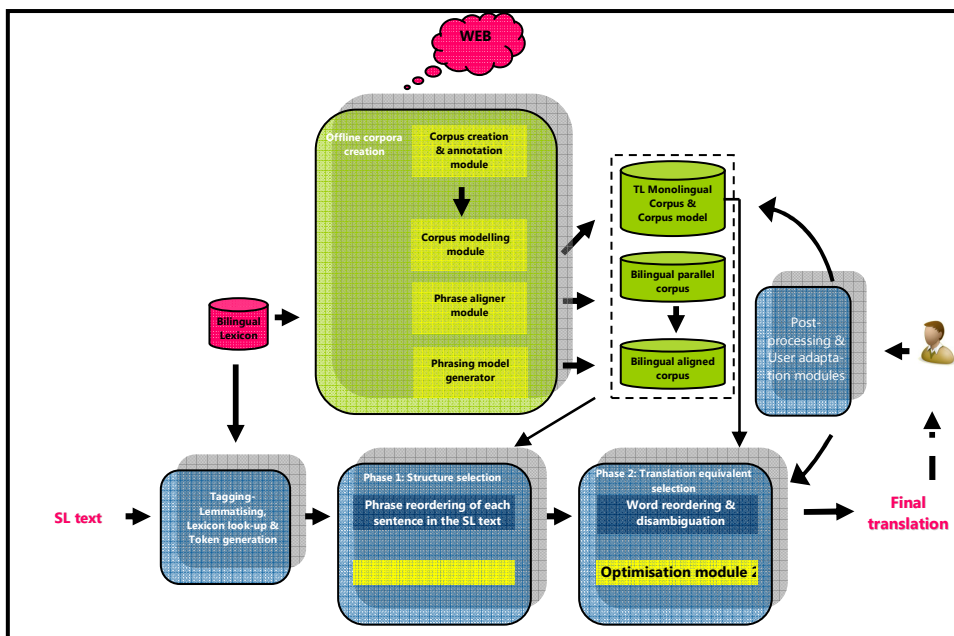
Within the **PRESEMT** system architecture three processing stages have been identified: (a) the Pre-processing stage, where the resources needed for the translation process are created, (b) the Main translation engine, involving the translation process per se and its optimisation and (c) the Post-processing stage, which allows users to modify the system output.

Website

During the first month of the project lifecycle the PRESEMT official website (www.presemt.eu) has been launched and is since maintained by ILSP. Having been created using the Joomla[1] CMS and registered in the "**.eu**" domain, as requested by the Commission, it is kept constantly updated, while it is planned to continue being maintained and updated after the formal completion of the project.

Serving as the principal means of communication and data exchange (reports, deliverables, data in general) between partners, besides functioning as the central dissemination point towards the outside world, the website is additionally foreseen to (a) host the PRESEMT prototypes and the respective user manuals, available for download, (b) provide the users with an online PRESEMT-based translation service and (c) make contact points publicly available to address requests for assistance.

---

[1] http://www.joomla.org/

On the PRESEMT website one can find comprehensive information about the **project** (key features, work packages, deliverables and milestones), the **consortium** and the **FP7 programme** as well as downloadable **dissemination material**. There are also a **news section**, a **login area** for registered users, various **links** of interest (such as other MT systems, MT evaluation metrics, relevant conferences, related projects etc.) and several **events** relevant to the project. Additionally, a **search** facility is provided, while it is foreseen to host a constantly updated list of project **publications**.

The picture is completed with the restricted access section (**Archive**), which is open only to consortium members and the European Commission project officer. The Archive is designed to serve as a repository (for software modules, deliverables, internal reports, draft documents, financial data etc.).

Figure 2: Screenshot of PRESEMT Homepage

The PRESEMT logo (cf. Figure 3), which saliently identifies PRESEMT as being an MT system, is also shown on the project website. The logo is available for download, appears on the cover page of all project reports and deliverables and also features in any info material that has up to this point been produced (or will be produced) during the project lifecycle.

Figure 3: PRESEMT logo

PRESEMT contact point

George Tambouratzis (Coordinator, Head of the Machine Translation Department (ILSP))

6, Artemidos & Epidavrou str.
151 25 Paradissos Amaroussiou
Athens, Greece

giorg_t@ilsp.gr

+30 210 6875411

+30 210 6854270

Homepage: http://www.ilsp.gr/homepages/tambouratzis_eng.html