Presemt

PRESEMT Fact Sheet

Project details

Project name	**PRESEMT** **P**attern **RE**cognition-based **S**tatistically **E**nhanced **MT**
Grant Agreement number	ICT-248307
Funding Scheme	Small or medium-scale focused research project – STREP – CP – FP – INFSO
Objective	ICT-2009.2.2: Language-Based Interaction
Start date	1.1.2010
Duration	36 months
Coordinating partner	Institute for Language & Speech Processing / R.C. "Athena" www.ilsp.gr Artemidos 6 & Epidavrou GR-15125, Maroussi, Athens, Greece
Coordinator	Dr. George Tambouratzis giorg_t@ilsp.gr Tel: +30 210 6875411 Fax: +30 210 6854270
Website	**www.presemt.eu**

SEVENTH FRAMEWORK PROGRAMME

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 248307.

The **PRESEMT** project is intended to lead to a flexible and adaptable MT system, based on a language-independent method, whose principles ensure easy portability to new language pairs, by utilising widely-available resources. 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 approaches and artificial intelligence techniques towards the development of a language-independent analysis, evolutionary algorithms for system optimisation. It is intended to be of a hybrid nature, combining linguistic processing with the advantages of corpus-based approaches.

The **PRESEMT** project has been conceived as a continuation to a series of MT systems, developed in the past 2 decades at various consortium members, these systems ranging from the EUROTRA project and the SYSTRAN rule-based system to, more recently, statistical-based methods relying on large-scale monolingual corpora for generating translations (such as the METIS and METIS-II IST-supported projects). The main requirements for the **PRESEMT** system are to generate translations fast (a real-time – or near real-time response is of prime importance) and to be able to develop new language pairs in a simple manner, without requiring specialised linguistic tools. In the modern multilingual environment of the European Union as well as beyond the Union, there exists an increased requirement for creating translation systems even for language pairs for which the availability of essential linguistic tools is limited.

Objectives

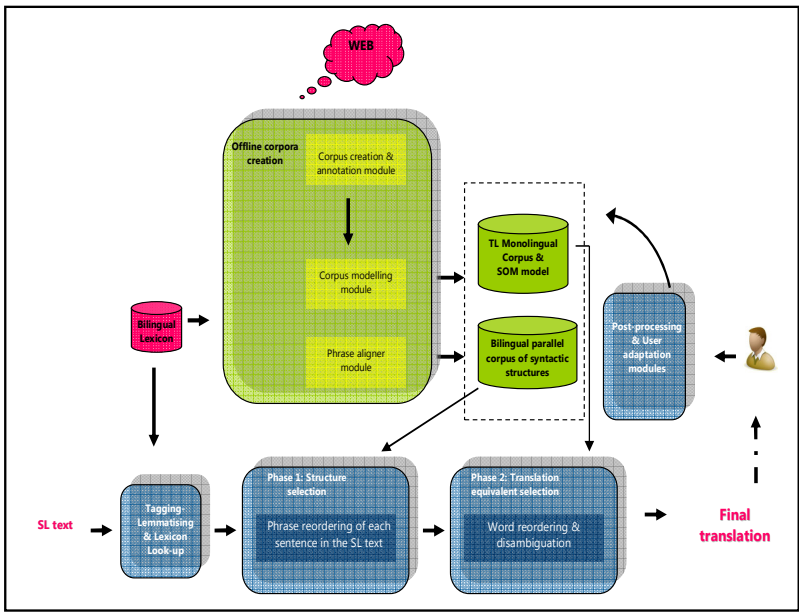
By developing the **PRESEMT** system the consortium intends to meet the following scientific and technical objectives:

1. System flexibility and adaptability
2. Language independence & portability to new language pairs
3. Capacity for self-learning
4. Novel methodology for language modelling & system optimisation

PRESEMT will be available both as a stand-alone application as well as a web-based service.

Presemt

PRESEMT Fact Sheet



System description

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.

PRESEMT Consortium

The consortium comprises 6 partners covering 5 European countries. Out of these partners, two are universities (NTNU and MU), three are non-profit research institutions (ILSP, GFAI and ICCS) and one is an SME (LCL). These partners possess complementary expertises, allowing the consortium to handle the task of machine translation on a Europe-wide basis, providing the desired degree of coverage in terms of languages but more importantly of language families as well as expertise in different areas of the **PRESEMT** project.

Language pairs

Development phases 1 & 2	German → English Greek → English Czech → English Norwegian → English English → German Greek → German Czech → German Norwegian → German
Development phase 3	German → Italian Greek → Italian Czech → Italian Norwegian → Italian English → Italian

PRESEMT Consortium

ILSP	Institute for Language & Speech Processing / R.C. "Athena" www.ilsp.gr	GFAI	Gesellschaft zur Förderung der Angewandten Informationsforshung e.V. www.iai-sb.de/iai/index.php/en/Die-GFAI.html	NTNU Det skapande universitetet	Norges Teknisk-Naturvitenskapelige Universitet www.ntnu.no
ICCS	Institute of Communication and Computer Systems www.iccs.gr	MU	Masaryk University www.muni.cz	LEXCOM Lexical Computing	Lexical Computing Ltd. www.sketchengine.co.uk