# Pattern REcognition-based Statistically Enhanced MT

**PRESE**

## Consortium

**ILSP**
Institute for Language & Speech Processing / R.C. "Athena"

**GFAI**
Gesellschaft zur Förderung der Angewandten Informationsforschung

**NTNU**
Norges Teknisk-Naturvitenskapelige Universitet

**ICCS**
Institute of Communication and Computer Systems

**MU**
Masaryk University

**LCL**
Lexical Computing Ltd.

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. This method attempts to overcome well-known problems of other MT approaches, e.g. compilation of extensive bilingual corpora 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.

**Time frame**
Start date: 1 January 2010
End date: 31 December 2012

www.presemt.eu

## Contact point

**George Tambouratzis,** Coordinator (ILSP)
e-mail: giorg_t@ilsp.gr
Tel: +30 210 6875411
Fax: +30 210 6854270

# Pattern REcognition-based Statistically Enhanced MT

**PRESE**

## S&T objectives

☑ Development of a novel method based on **generalised clustering techniques**, for creating a **language-independent phrase aligner** also adaptable to phrasing principles designated by the end users

☑ Use of **pattern recognition** approaches for defining **syntactic structure**

☑ Employment of techniques inspired by the **functional biological systems** for **disambiguating** translations

☑ Extensive use of **automated optimisation techniques** to define a mature system for methodically **optimising** system parameters

☑ Application of **machine learning** methods for allowing system **adaptation**

☑ Use of **parallel computing** architectures as well as mainstream multi-core architectures for PCs for substantial advances in **translation speed**

## Key innovation

The PRESEMT project proposes a novel approach to the problem of Machine Translation by introducing cross-disciplinary techniques, mainly borrowed from the **machine learning** and **computational intelligence** domains, in the MT paradigm.

To this end, a flexible MT system will be developed, which will be enhanced with (a) **pattern recognition** approaches (such as extended clustering or neural networks) towards the development of a language-independent analysis and (b) **evolutionary computation** (such as Genetic Algorithms or Swarm Intelligence) for system optimisation.

## Result

The PRESEMT project will result in a fully functional system prototype, available both as a stand-alone application as well as a web-based service.
Furthermore, it will provide a language-independent methodology for effectively handling new language pairs.

## Language pairs

### Development phases 1 & 2

Czech ➔ English
German ➔ English
Greek ➔ English
Norwegian ➔ English

Czech ➔ German
Greek ➔ German
English ➔ German
Norwegian ➔ German

### Development phase 3

Czech ➔ Italian
German ➔ Italian
Greek ➔ Italian
English ➔ Italian
Norwegian ➔ Italian

www.presemt.eu