



Pattern **RE**cognition-based **S**tatistically **E**nhanced **MT**

ANNUAL PUBLIC REPORT 2

Grant Agreement number	ICT-248307
Project acronym	PRESEMT
Project title	Pattern REcognition-based Statistically Enhanced MT
Funding Scheme	STREP – CP-FP-INFSO
Deliverable title	Public Annual Report 2
Dissemination level	Public
Period covered	2011
Responsible partner	ILSP

Project coordinator name & title	Dr. George Tambouratzis
Project coordinator organisation	Institute for Language and Speech Processing / RC 'Athena'
Tel	+30 210 6875411
Fax	+30 210 6854270
E-mail	giorg_t@ilsp.gr
Project website address	www.presemt.eu

PRESEMT consortium & contact persons



Institute for Language and Speech Processing/R.C. "Athena"

Coordinator

<http://www.ilsp.gr/>

Contact person: Dr. George Tambouratzis, giorg_t@ilsp.gr



Gesellschaft zur Förderung der Angewandten Informationsforschung e.V.

<http://www.iai-sb.de/iai/index.php/en/Die-GFAI.html>

Contact person: Dr. Paul Schmidt, paul@iai.uni-sb.de



Norges Teknisk-Naturvitenskapelige Universitet

<http://www.ntnu.no/>

Contact person: Prof. Björn Gambäck, gamback@idi.ntnu.no



Institute of Communication and Computer Systems

<http://www.iccs.gr/eng>

Contact person: Dr. Georgios Goumas, goumas@cslab.ece.ntua.gr



Masaryk University

<http://www.muni.cz/>

Contact person: Prof. Karel Pala, pala@fi.muni.cz



Lexical Computing Ltd.

<http://www.sketchengine.co.uk/>

Contact person: Dr. Adam Kilgariff, adam.kilgariff@gmail.com

Table of Contents

1. PRESEMT overview3

2. PRESEMT system description4

3. Activities within the 2nd year of the project7

4. Dissemination activities11

5. Future work14

6. Further information15

1. PRESEMT overview

PRESEMT (Pattern REcognition-based Statistically Enhanced MT) is an EU-funded project under the FP7 topic "ICT-2009.2.2: Language-based Interaction". It is intended to lead to a flexible and adaptable Machine Translation (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.

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 GAs or swarm intelligence) will be studied.

To allow for user adaptability, all the corpora used in PRESEMT will be retrieved from web-based sources via the system platform, while user feedback will be integrated through appropriate interactive interfaces.

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** techniques (such as extended clustering or neural networks) towards the development of a language-independent analysis and (b) **evolutionary computation** methods (such as Genetic Algorithms or Swarm Intelligence) for system optimisation.

Features

The core features of PRESEMT are listed below:

1. Development of a novel method based on **generalised clustering techniques**, for creating a **language-independent phrase aligner** adaptable to phrasing principles defined by the end users
2. Use of **pattern recognition** approaches for defining **syntactic structure**
3. Employment of techniques inspired by **functional biological systems** for **disambiguating** translations
4. Extensive use of **automated optimisation techniques** to define a mature system for methodically **optimising** system parameters
5. Application of **machine learning** methods for allowing system **adaptation**
6. Use of **parallel computing** architectures as well as mainstream multi-core architectures for PCs for substantial advances in **translation speed**

2. PRESEMT system description

The PRESEMT system, as envisaged and implemented up to this point, is depicted in Figure 1. It roughly comprises 3 components, each of them having a modular structure (cf. Table 1):

1. Pre-processing stage

It involves the compilation of resources needed for the MT system to perform, i.e. the collection and appropriate annotation of corpora, the elicitation of phrasing information as well as the extraction of semantic and statistical data.

2. Main translation engine

This component, being the core part of the system, translates a source language (SL) text to a target language (TL) one, drawing, in stepwise mode, on the information obtained in the Pre-processing stage.

3. Post-processing stage

This stage offers the user the opportunity to modify the system translation output according to their preferences. These modifications can then be endorsed by the system so as to adapt itself to the given input.

Table 1: PRESEMT basic system modules

Pre-processing stage: 4 modules	Main translation engine: 3 modules	Post-processing stage: 2 modules
Corpus creation & annotation module	Structure selection module	Post-processing module
Phrase aligner module	Translation equivalent selection module	
Phrasing model generator	Optimisation module	User adaptation module
Corpus modelling module		

Pre-processing stage

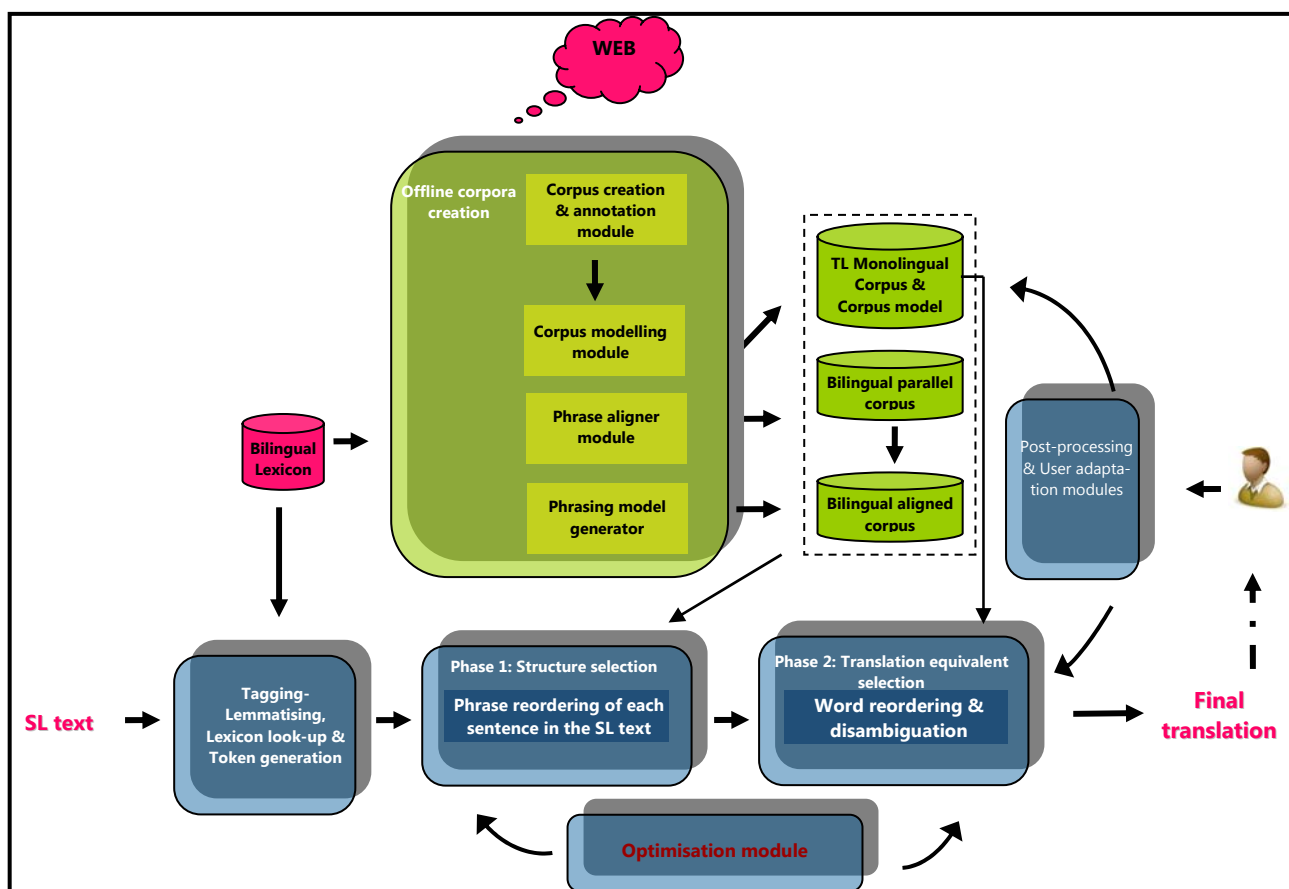
The **Corpus creation & annotation module** entails the compilation and annotation of large monolingual and small bilingual corpora to be utilised by the Main translation engine. The former are collected via web crawling, while the latter are created manually (mainly based on web resources). The collected text resources are submitted to various levels of processing (e.g. monolingual corpora: cleaning and content de-duplication; bilingual corpora: corrections / modifications) and annotation (e.g. Part-of-Speech (PoS) tagging and lemmatisation).

The **Phrase aligner module**, operating on bilingual corpora (cf. the aforementioned ones), performs word-and-phrase-level alignment of a bilingual corpus, one side of which is annotated only with PoS tags and lemmata, while the other one additionally bears phrasing information. In the current implementation the source language is assumed to be the non-parsed member of the language pair, while the target language is fully annotated. After determining lexical correspondences within a given language pair and on the basis of the TL parsing the Phrase aligner proceeds to segmenting the SL corpus side into phrases. It subsequently outputs the bilingual corpus aligned at clause, phrase and word level.

The **Phrasing model generator** takes as input the output of the Phrase aligner and utilises it so as to (a) generate a probabilistic phrasing model for the source language and (b) apply this model for segmenting a given SL text being input for translation. For the first task the module operates offline, whereas the second mode is an online process that forms part of the actual translation procedure.

The last module of this stage, the **Corpus modelling module**, takes as input an annotated TL monolingual corpus (yielded by the Corpus creation & annotation module) and processes it so as to extract semantic-type and statistical-based information (by applying methods such as n-gram models over words and PoS tags, SOM for words and vector space models). This type of information will then be utilised during the translation process to choose the most likely among a number of candidate translations.

Figure 1: PRESEMT System architecture



Main translation engine

The Main translation engine is split into two phases:

The **Structure selection module** determines the optimal structure of an SL sentence, by utilising information residing in a bilingual corpus.

The **Translation equivalent selection module** disambiguates translation equivalents and microstructures, after the SL sentence structure has been established, by utilising information residing in a monolingual corpus.

The **Optimisation module** is responsible for enhancing the performance of the two translation phases, by optimising the values of the parameters employed.

Post-processing stage

The **Post-processing module** is a GUI via which the user can feedback their modifications to the system translation output.

The **User adaptation module** collects the user modifications and "corrects" the translation system accordingly.

Language pairs covered

The language pairs, to be examined as case studies and for evaluation purposes, have been selected on the basis of three criteria, namely (a) availability of a large TL corpus, (b) examination of different language families and (c) coverage of the consortium languages:

The left-hand column of the following table illustrates the language pairs that will be handled for the development of the first two versions of the system prototype, whereas the right-hand column lists the language pairs to be used for system assessment.

Table 2: Language pairs covered by PRESEMT

Language pairs (development phases 1 & 2)	Language pairs (development phase 3)
* Czech \Rightarrow English	* Czech \Rightarrow Italian
* German \Rightarrow English	* English \Rightarrow Italian
* Greek \Rightarrow English	* German \Rightarrow Italian
* Norwegian \Rightarrow English	* Greek \Rightarrow Italian
* Czech \Rightarrow German	* Norwegian \Rightarrow Italian
* English \Rightarrow German	
* Greek \Rightarrow German	
* Norwegian \Rightarrow German	

3. Activities within the 2nd year of the project

During the second year of the project one main phase is evident, corresponding to the twelve months of the reporting period, for which all technical packages are active. These include the four work packages revolving around the design and implementation of PRESEMT modules, namely WP3, WP4, WP5 and WP6, as well as the integration of the system modules into a single platform, i.e. WP7. The relevant activities culminate at the end of the second project year into the release of the final versions of the different modules. The first PRESEMT prototype has been released earlier (at month M19, i.e. just after the middle of the second year), with further improvements continuing till the end of the year (month M24).

During the latter part of the second year, the validation and evaluation activities are also initiated. These involve initially the validation of the modules developed and afterwards the evaluation of the first prototype of the PRESEMT system in the translation task.

In the remainder of this section, the main results obtained and objectives achieved during the 2nd year are summarised per work package.

WP3: Corpus extraction and processing algorithms

WP3 relates to the design and development of the Pre-processing modules.

Work has continued on collecting text resources, both monolingual and bilingual ones, over the web and annotating them with tag, lemma and phrase information. The monolingual corpora have been collected by crawling the web with the Heritrix crawler and have then been processed with the *justText* algorithm (see Pomikálek (forthcoming¹)) for removing duplicate content and non-informative parts outside of the main content of a web page, typically machine-generated and repeated across the web pages of the same website. The main progress in the collection of monolingual corpora during the second year has concerned the collection of monolingual corpora in Greek, Czech and Norwegian. In addition, a new annotation has been performed for English and German, in accordance with the requirements of the PRESEMT translation engine.

The bilingual corpora are substantially smaller in size (~200 sentence pairs each). The collection of these bilingual corpora has been completed during the second year of the project, for all studied language pairs involving English and German as the target languages.

The following tables provide an overview of the text resources collected so far in the project.

¹ Pomikálek J. (forthcoming). Removing Boilerplate and Duplicate Content from Web Corpora (PhD thesis). Masaryk University, Brno, Czech Republic.

Table 3: Statistical data on PRESEMT monolingual corpora (2nd year of the project)

Language	English	Italian	German	Greek	Czech	Norwegian
Corpus name	enTenTen / BiWeC	itTenTen	deTenTen / BigDeWaC	grTenTen	czTenTen	noTenTen
final size (in tokens)	3,658,726,327	3,076,812,674	2,874,779,294	1,073,266,453	2,028,422,159	2,267,156,701
downloaded volume (gzipped)	129 GB	272 GB	291 GB	157 GB	418 GB	364 GB
downloaded volume uncompressed	645 GB	1360 GB	1455 GB	785 GB	2090 GB	1820 GB
downloaded (unique) URLs	13,638,928	33,459,999	43,160,992			
original main content (words)	6,805,296,135	7,740,199,568				
original boilerplate content (words)	7,843,420,305	14,814,859,910				
after removing exact duplicates						
docs	3,357,252	5,335,839	8,237,310	3,357,894	5,982,830	
tokens	4,765,119,530	5,017,409,779	4,880,335,291	1,952,309,530	3,325,287,880	
duplicate 10-grams	212,864,389	295,369,351	331,258,395	131,047,973		
after removing duplicate text blocks						
docs	2,838,738	4,020,968	5,752,857	2,208,243	3,785,229	6,015,349
tokens	3,658,726,327	3,076,812,674	2,874,779,294	1,073,266,453	2,028,422,159	2,267,156,701
duplicate 10-grams	6,757,185	12,230,555	7,196,387			
	words = whitespace separated character strings (wc -w)					
	tokens = unitok.py (universal tokeniser) output					

Table 4: PRESEMT bilingual corpora (2nd year of the project)

		Target Language		
		English	German	Italian
Source Language	Czech	1	1	
	English			
	German	1		
	Greek	1	1	
	Norwegian	1	1	

A new version of the **Phrase aligner module** has been developed and released. The algorithmic part has been extensively revised and now operates in a three-stage mode, (a) performing alignment of single words from SL to TL, (b) extending alignments by using part-of-speech categories and (c) completing the alignment taking into account (i) additional tag information and (ii) the word environment. The new module is faster to run with reduced memory requirements, while also giving an increased accuracy which reaches 94.0% for the Greek → English pair and 87.5% for the German → English pair.

The **Phrasing model generator** has been developed on the basis of the CRF (Conditional Random Fields) implementation in the MALLET package. This has been extensively tested in order to determine the optimal parameters for the Greek language. The optimal accuracy attained was over 90% for the Greek language. In the latter part of the second year, emphasis has been placed on applying the CRF to create models of other source languages, initially German and English.

Development of the **Corpus modelling module** has also proceeded. A number of different methods for modelling have been used, these being (i) n-gram modelling, (ii) vector space modelling and (iii) Self-Organising Maps. Experimentation with these methods has focussed on the two main target languages of the project, namely German and English. In addition, suites designed for the evaluation of the proposed methods in the disambiguation task have been developed.

WP4: Structure selection & WP5: Translation equivalent selection

These two work packages involve the design and development of the two translation phases, the first of which establishes the translation structure based on the information contained in a bilingual corpus, while the second determines word order and introduction/deletion of words in the final translations. The consortium has continued investigating two conceptually-related yet distinct algorithmic implementations in order to identify their comparative strengths and weaknesses. The first approach involves using synchronous grammars and consists of (a) a parser for synchronous grammars and (b) an algorithm that extracts bilingual productions out of the bilingual corpus, and a mix of flat and hierarchical structures in order to account for the productivity of language. The second approach relies on the dynamic programming paradigm, the structure selection process being treated as a sequence alignment, aligning the input source sentence to an SL sentence from the aligned bilingual corpus and assigning a similarity score.

These two approaches are being applied initially to the first batch of translation pairs, namely Greek → English, German → English, English → German and Greek → German. Other language pairs will be introduced at a slightly later time-schedule.

With respect to the optimisation of the translation modules, a set of parameters have been identified and the effect of changing these parameters has been investigated. For the actual optimisation of the parameters, the SPEA2² multi-objective evolutionary algorithm is the first-choice algorithm, with other algorithms to be studied at a later time-frame.

WP6: Post-processing & User adaptation

Within WP6 a first version of the Graphical User Interface for implementing post-processing of translations has been completed. In addition, the related requirements have been discussed and finalised for the subsequent versions of the post-processing interface. In addition work has started on the User adaptation mechanisms, which will be integrated in forthcoming versions of the PRESEMT prototype.

² E. Zitzler, M. Laumanns, L. Thiele, "SPEA2: improving the strength Pareto evolutionary algorithm," Swiss Federal Institute of Technology (ETH). Zurich, Switzerland. Technical report TIK-Report 103, 2001

WP7: Integration

The main aim of WP7 is to integrate the modules developed into the technical work packages (i.e. WP3, WP4, WP5 and WP6) into one working prototype and to exploit parallelisation opportunities towards the efficient application of the prototype to multi-processor/multi-core architectures.

To support this effort, the Apache Subversion revision control server established has been used in joining together all the modules developed by the partners. This has led to the release of the 1st PRESEMT prototype in early autumn, together with the associated documentation and user manual.

As the second main aim of WP7, the parallelisation of the translation engine has been studied, with interactions between algorithms, data structures and execution platforms being analysed to determine the optimal parallelisation strategy. The results of the parallelisation activities will become evident in the second version of the PRESEMT prototype.

WP9: Validation & Evaluation

The purpose of the specific work package is the validation of the PRESEMT system and its individual modules as well as the evaluation of the translation output.

The validation process has just begun at each partner's site. Validators are requested to check two specific system functionalities, i.e. the translation process and the post-processing, and document their experimentation with the system by filling in the respective validation forms, which have been compiled for this purpose.

It is planned to follow the same validation pattern in the future for the residual system functionalities.

In relation to evaluation, specifications have been compiled concerning the data to be used (development and test sets) for the evaluation process, both consortium-internally and consortium-externally. Currently, the developments sets are being prepared based on content drawn automatically from the web.

Besides, a first version of the evaluation GUI, which will be accessed by end users for assessing the PRESEMT translation output, has been implemented, and it is currently being tested and further developed.

4. Dissemination activities

The following tables summarise the main dissemination activities undertaken by the PRESEMT consortium members during the second year of the project, as well as activities planned for the immediate future.

Publications & presentations at conferences / workshops							
Event title	Place	Date	Publication / presentation title	Type	Authors	Details	Status
IEEE Symposium on Computational Intelligence in Multicriteria Decision-Making	Paris, France	April 11 -15, 2011	Studying the SPEA2 Algorithm for Optimising a Pattern-Recognition Based Machine Translation System	Conference paper	Sokratis Sofianopoulos & George Tambouratzis		Published
Contrastive Linguistics – Translation Studies – Machine Translation – what can we learn from each other? Workshop held in conjunction with the annual conference of GSCL	Hamburg, Germany	September 27, 2011	Using annotated corpora for rapid development of new language pairs in MT	Conference paper	Susanne Preuss, Hajo Keffer, Paul Schmidt		Published
India-Norway Workshop on Web Concepts and Technologies 2011 [INWWCT 2011]	Trondheim, Norway	October 3, 2011	A Survey of Domain Adaptation in Machine Translation: Towards a refinement of domain space	Conference paper	Lars Bungum & Björn Gambäck		Published
International Workshop on Using Linguistic Information for Hybrid Machine Translation [LIHMT-2011]	Barcelona, Spain	November 18, 2011	Word Translation Disambiguation without Parallel Texts	Conference paper	Erwin Marsi, André Lynum, Lars Bungum & Björn Gambäck		Accepted for presentation
LREC 2012	Constantinople, Turkey	May 21-27, 2012	Machine Translation Using Mainly Large Monolingual Corpora	Conference paper	Sokratis Sofianopoulos, Marina Vassiliou & George Tambouratzis		Submitted

Publications & presentations at conferences / workshops

Event title	Place	Date	Publication / presentation title	Type	Authors	Details	Status
15 th International Conference of the European Association for Machine Translation [EAMT2011]	Leuven, Belgium	May 30-31, 2011	A resource-light phrase scheme for language-portable MT	Conference paper	George Tambouratzis, Sokratis Sofianopoulos, Marina Vassiliou, Fotini Simistira & Nikos Tsimboukakis		Published
14 th International Conference on Text, Speech and Dialogue [TSD 2011]	Plzeň, Czech Republic	September 1-5, 2011	Effective Parsing Using Competing CFG Rules	Conference paper	Miloš Jakubíček		Presented
2 nd Sketch Engine Workshop [SKEW-2]	Brighton, UK	March 16 – 17, 2011	Corpus Architect developments and CCBC (Comparable Corpus BootCat)	Conference paper	Jan Pomikálek	Jan Pomikálek and Adam Kilgarriff have been involved in the organisation of the event.	Presented
Research Models in Translation Studies II, Panel 6: Rethinking Corpus-based Translation Studies in the Web Era	Manchester, UK	April 29-May 2, 2011	Web Corpora for a Hierarchy of Domains	Conference paper	Adam Kilgarriff	Adam Kilgarriff & Silvia Bernardini have placed a proposal for the specific panel discussion at the conference.	Presented
25 th European Conference on Object-oriented Programming (ECOOP-2011)	Lancaster, UK	July 25-30, 2011	The PRESEMT project for the Machine Translation Task	---	George Tambouratzis	1. Oral presentation of the PRESEMT project by George Tambouratzis, following an invitation received from the ECOOP conference <i>To be published in the Blog section of the Journal of Object Technology</i> (http://www.jot.fm) 2. Poster presentation on PRESEMT project	Presented
15 th International Conference of the European Association for Machine Translation [EAMT2011]	Leuven, Belgium	May 30-31, 2011	PRESEMT	Conference poster presentation	---	EU project presentation session	Presented

Publications & presentations at conferences / workshops

Event title	Place	Date	Publication / presentation title	Type	Authors	Details	Status
Lexicom 2011: Workshop in Lexicography and Lexical Computing	St Petersburg, Russia	June 14-18, 2011	---	---	---	Adam Kilgarriff and Jan Pomikálek were members of the organising committee. Event included PRESEMT presentations	---
18 th Nordic Conference of Computational Linguistics (NoDaLiDa), Workshop on Constraint Grammar Applications	Riga, Latvia	May 11-13, 2011	OBT+Stat: Evaluation of a combined CG and statistical tagger	Conference paper	Janne Bondi Johannessen, Kristin Hagen, André Lynum and Anders Nøklestad		Presented
The Second e-lexicography conference	Bled, Slovenia	November 10-12, 2011	Comparable Corpora Boot-Cat	Conference paper	Adam Kilgarriff & PVS Avinesh		Presented

Info days / Exhibitions / Other events

Event title	Place	Date	Presentation title	Details / Comments
Linguistics Institute 2001	University of Colorado, USA	July 7-August 2, 2011		LCL has sponsored the Linguistics Institute, a prestigious annual four-week summer school organised by the Linguistics Society of America. It is typically attended by around 600 academics and graduate students. This year the Institute focus has been on interdisciplinary, empirically based approaches to language”, which fits well with PRESEMT methods as well as LCL expertise. Adam Kilgarriff has given several lectures, including and showcasing aspects of PRESEMT work.
META-FORUM 2011	Budapest, Hungary	June 27-28, 2011	PRESEMT	Pavel Rychlý and Miloš Jakubíček have presented the PRESEMT project within the META Exhibition forum

Teaching

Title	Type	Lecturer	Place	Date
Machine Translation, Natural Language Interfaces (NTNU)	Lecture	Björn Gambäck (NTNU)	Trondheim, Norway	March 2011 (spring semester 2011)

5. Future work

Within the next period (third year of the project), work in the PRESEMT project will focus to a large extent on improving the PRESEMT prototype by evaluating its effectiveness in the translation task and then performing the required modifications to the relevant modules. In addition, the remaining functionalities that have not been provided in the first version will be added.

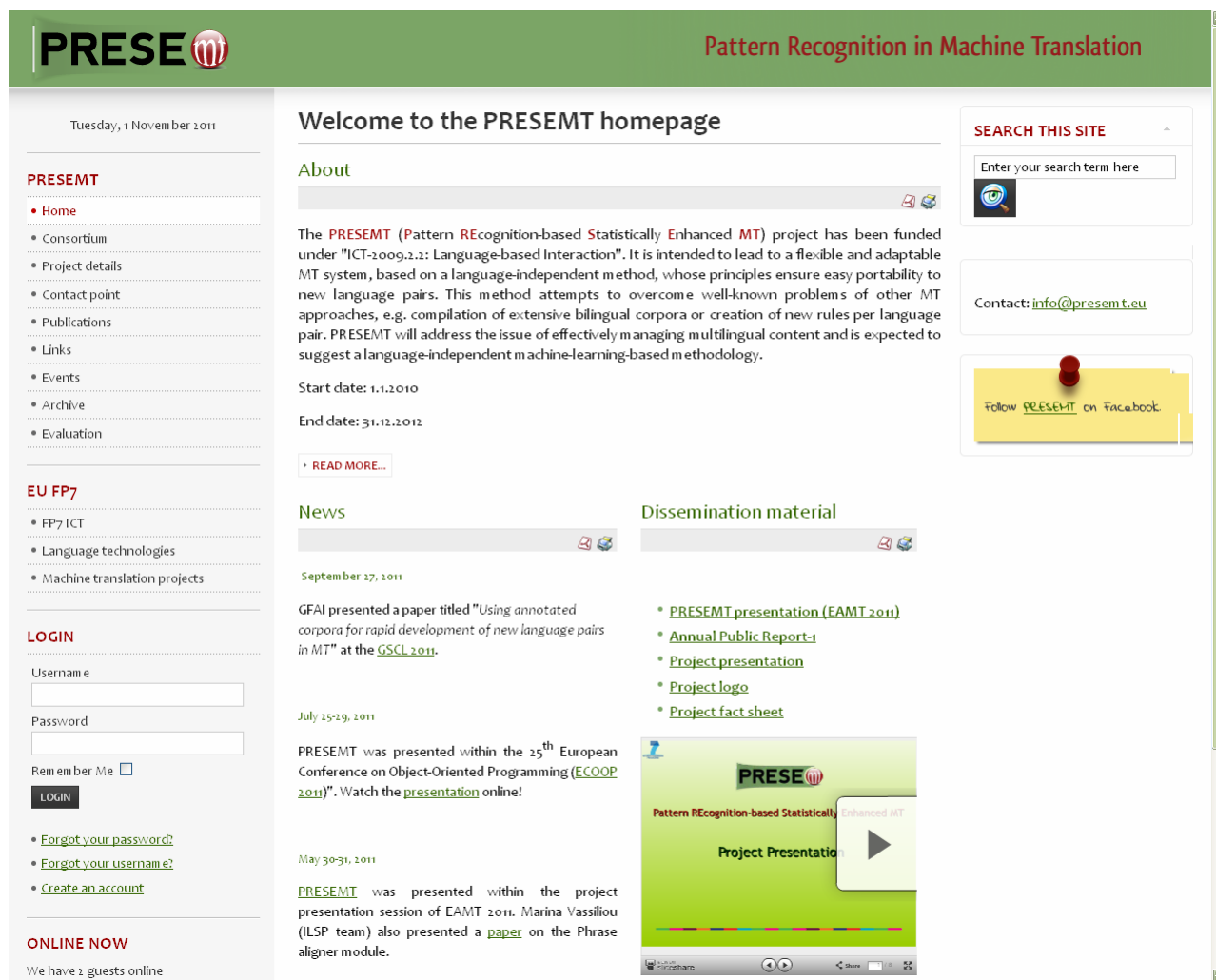
Major improvements, to be carried out within the remainder of the second as well as the third year of the project, involve the two translation phases (structure selection and translation equivalent selection), which will be studied extensively, based on the evaluation activities planned. In addition, the optimisation of parameters for both phases will be studied by applying automated evaluation metrics on evolutionary computation algorithms. Based on these revisions, a sequence of two new prototypes (together with their associated user manual and documentation deliverables) is planned. The forthcoming prototype versions will be further enhanced, by adding the user adaptation functionalities that will allow the system to be tailored towards the requirements of specific users. Finally, the results of the parallelisation activities will be incorporated, leading to substantial improvements in the system response time (the time required to generate a translation).

In addition, during the forthcoming months, the PRESEMT prototype will be extended to include additional language pairs, initially corresponding to translation pairs involving English and German as the target languages, and then Italian as a target language. The release of the MT systems for these translation pairs will be combined with extensive evaluation activities to investigate the translation quality and the system usability. This is expected to enable sound conclusions to be drawn regarding the effectiveness of the PRESEMT translation paradigm.

Furthermore, dissemination activities will be continued and intensified, in order to publicise the project results to the wider academic community as well as the existence of the system to prospective user groups. These activities will also be communicated via the PRESEMT website in order to reach the relevant audience. The prototype will be made available for download and use, while at the same time individual modules developed within PRESEMT will be released for use by other researchers in other relevant applications.

6. Further information

For further information and for keeping up-to-date regarding the PRESEMT project, please visit our web-site at www.presemt.eu.



The screenshot shows the PRESEMT homepage with a green header. The header contains the PRESEMT logo and the text "Pattern Recognition in Machine Translation". Below the header, the date "Tuesday, 1 November 2011" is displayed. The main content area is divided into several sections:

- Welcome to the PRESEMT homepage**: A section titled "About" describing the PRESEMT project, its funding, and its goals. It includes the start date (1.1.2010) and end date (31.12.2012), and a "READ MORE..." link.
- News**: A section titled "News" with two entries:
 - September 27, 2011: GFAI presented a paper titled "Using annotated corpora for rapid development of new language pairs in MT" at the GSCL 2011.
 - July 25-29, 2011: PRESEMT was presented within the 25th European Conference on Object-Oriented Programming (ECOOP 2011). Watch the presentation online!
 - May 30-31, 2011: PRESEMT was presented within the project presentation session of EAMT 2011. Marina Vassiliou (ILSP team) also presented a paper on the Phrase aligner module.
- Dissemination material**: A section titled "Dissemination material" with a list of links:
 - PRESEMT presentation (EAMT 2011)
 - Annual Public Report-1
 - Project presentation
 - Project logo
 - Project fact sheet
- Search this site**: A search bar with the text "Enter your search term here" and a magnifying glass icon.
- Contact**: A contact form with the email address info@presemt.eu.
- Follow PRESEMT on Facebook**: A yellow button with a red pushpin icon.
- Left sidebar**: Contains a navigation menu with links to Home, Consortium, Project details, Contact point, Publications, Links, Events, Archive, and Evaluation. It also includes a login section with fields for Username and Password, a "Remember Me" checkbox, and a "LOGIN" button. Below the login section are links for "Forgot your password?", "Forgot your username?", and "Create an account". At the bottom of the sidebar, it says "ONLINE NOW" and "We have 2 guests online".