

The Institute for Language and Speech Processing (ILSP) was founded in Athens (Greece) in 1991 under the auspices of the Ministry of Development. ILSP began its Research and Development activities in May 1992.

The goal of ILSP is to support the growth of Language Technology in Greece. For this purpose it has brought together a team of experts and academics, and has created the necessary technical infrastructure in accordance with the EC policy towards safeguarding the European cultural heritage through technology. ILSP aims to be a pole of attraction for the language industry, which is growing both in Greece and in the rest of Europe, thus contributing to the expansion of activities in this particularly important area of modern IT. The industrial direction which it maintains, the experience of its researchers and the close relations which it keeps with key research centres in other European countries, are the three basic elements in the profile of ILSP.

ILSP carries out applied research in Speech Processing, Text Processing and Language Learning Technologies. Expertise used by the Institute includes basic fields as NLP, DSP and Pattern Recognition. Its mission is mainly to support basic research, promoting on the other hand the development of new products in the form of laboratory prototypes.

The research taking place both for Natural Language Processing and Speech Processing gives these areas new impetus, which stems from their complementarity. Taking as a fact that the Institute combines in its activities the research in these two thematic areas, it has the ability to successfully deal with those issues that require know-how from both domains.

The Institute pushes the technologies developed in the framework of projects until the phase of industrial prototypes and products, using the full cycle of industrial development including quality control.

Specifically, ILSP

- develops environments for translating from and into the Greek language, as well as computational tools and products which assist the translation task
- develops CD-ROMs and Websites for computer assisted Greek language learning
- creates electronic dictionaries (monolingual and multilingual), computational lexica and electronic dictionaries for children
- develops prototypes for speech recognition, synthesis and compression
- creates text correction tools
- offers Language Technology consultancy and customisation services and solutions to government, industry and end users



With its expert researchers (90 employees), ILSP will continue its developmental activities, thus contributing to the creation of user-friendly applications and new services by embedding language technology.

ILSP activities have been organised in six departments and a branch office in Northern Greece.

- Electronic Lexicography Department
- Dept. of Language Technology Applications in Office Systems
- Department of Educational Technology
- Department of Speech Technology
- Department of Machine Translation
- Liaison Department
- ILSP Thrace Branch, specialized in network software

All activities are supported by an Administration Office.

Department of Electronic Lexicography

The main objectives of the Department of Electronic Lexicography is to conduct basic and applied research in the field of Computational Lexicology and Lexicography, to create language resources (Text Corpora and Lexica) for the Greek language and to contribute to the development of laboratory prototypes. Research and development in this department focuses on the following domains:

- collection of monolingual and multilingual (parallel) corpora, which serve as the basic language material for linguistic research and application development
- design of lexicographic specifications (criteria for macrostructure design, specifications for microstructure codification, etc.)
- design and development of a lexicographic Workbench
- development of monolingual (Greek) and multilingual electronic dictionaries for human users
- development of computational lexica for Natural Language Processing
- creation of terminological databases

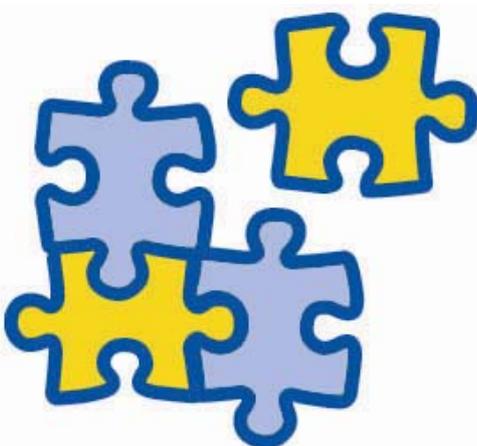


Department of Language Technology Applications in Office Systems

The Department conducts basic and applied research in the field of Natural Language Processing. Its main aim is to design computational models for natural language recognition and "understanding". In particular, ILSP/LTA designs, implements and integrates human language technologies in systems and applications dealing with structured data as well as with unstructured data processing. Applications include language-aware information retrieval and extraction, natural language interfaces, text mining, knowledge management, etc.

LTA research and development efforts aim at developing technology amenable to multilingual and multimedia information processing. Multilinguality is catered for by resorting to processing of parallel and comparable text corpora and elicitation of multilevel translation equivalents. Multimodality is handled by integrating conversion technologies, able to convert speech and image modalities to text via automatic speech recognition and image processing.

Tools are designed and developed for:



- efficient structural and grammatical text annotation
- shallow parsing
- named entity recognition
- term extraction
- co-reference resolution
- event recognition and their interrelations
- semantic-web oriented information extraction
- text classification
- word sense disambiguation
- multilevel text corpora matching
- multilevel parallel text alignment
- automatic elicitation of lexical data pertinent to monolingual and multilingual applications
- intelligent translation memory applications

Department of Educational Technology

The Department focuses on research and development of products and services in the field of Educational Technology, primarily related to language teaching combined with cultural education.

The main activities of the Department are developed around the following topics:

- Teaching Greek as a mother tongue
- Teaching Greek as a second language
- Teaching Greek as a foreign language
 - Teaching Ancient Greek
 - Teaching the Greek Sign Language
 - Parallel teaching of mother tongue and second language to young learners



The activities of the Department aim at developing innovative e-learning environments for language learning, incorporating language technology tools and methods in order to enhance language learning and foster language acquisition. Basic research in this department is directed towards the successful definition of useful tools while applied research is carried out for the evaluation of their efficiency.

Department of Speech Technology

The department develops basic technology which will be used in speech-enabled applications. Basic and applied research is carried out continuously on speech synthesis, compression and speech quality improvement. The objective is the definition and development of products which incorporate speech technology and can be used within specific applications. The following platforms are being developed:

- speech synthesis from written text
- isolated word and continuous speech recognition
- Audio to MIDI conversion through audio recognition techniques
- Automatic audio / speech discrimination
- Optical music recognition
- Score matcher and score follower
- Greeklsh-to-Greek transliteration
- speech compression techniques
- noise reduction techniques
- spectral analysis techniques



Department of Machine Translation

Work in the Department focuses on machine translation issues. Modern Greek stands at the heart of the Department's activities, which are built around the following three main axis:

- The development of Machine Translation systems which take advantage of the up-to-date technological advances and the existing international experience in the field
- The development of: (i) integrated environments, such as authoring tools for controlled languages, (ii) resources, such as grammars and multilingual lexica thesauri and ontologies. All these activities support and/or facilitate Machine Translation
- The study of Modern Greek both at grammatical and stylistic level in order to further enhance the developmental work undertaken by the Department

The know-how acquired from / integrated in the following projects/technologies can be exploited in the frame of the current strategic objectives of IST 4th and 5th call.

REVEAL THIS - IST: The project addresses a basic need underlying content organisation, filtering, consumption and enjoyment by developing content programming systems that will help European citizens keep up with the explosion of digital content scattered over different platforms (radio, satellite TV, World Wide Web, etc), different medias (speech, text, images, video) and different languages. People should be spending most of their leisure time enjoying the content, not searching for it.

Strategic Objectives: *Multimodal Interfaces – Networked Audio Visual Systems and Home Platforms – Semantic-based Knowledge and Content Systems*

MUSE: MUSE will develop a robust system for analysis, annotation, storage and indexing of vast volumes of audiovisual content concerning specifically business news, aiming at the personalised distribution of content and metadata through two different platforms (web and mobile telephones). This platform will be a powerful tool in the hands of media, TV and news services.

Strategic Objectives: *Networked Audio Visual Systems and Home Platforms*

CIMWOS - IST: CIMWOS is a multimedia, multimodal and multilingual system supporting content-based indexing, archiving, retrieval, and on-demand delivery of audiovisual content. The system uses a multifaceted approach to locate important segments within multimedia material employing state-of-the-art algorithms for text, speech and image processing. The CIMWOS system can be a powerful tool in the hands of the world of media and television, video, news broadcasting, show business, advertisement, and any organization that produces, markets and/or broadcasts video and audio programs.

Strategic Objectives: *Multimodal Interfaces – Networked Audio Visual Systems and Home Platforms – Semantic-based Knowledge and Content Systems*



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

MUSA: MUSA aims to provide the industry with a system that will automatically generate and translate subtitles of multimedia content such as videos and television programmes both prerecorded and live.

Strategic Objectives: *Networked Audio Visual Systems and Home Platforms*

<http://sifnos.ilsp.gr/musa>



Visit MUSA at the IST Event 2004, Communication and Amusement zone

ERMIS - IST: The project conducts a systematic analysis of speech and facial input signals, in separate, as well as in common; the aim is to extract parameters and features which can provide human computer interaction (HCI) systems with the ability to recognize the basic emotional state of their users and interact with them in a more natural and user friendly way. The continuity of emotion space, the uncertainty involved in the feature estimation process and the required ability of the system to use prior knowledge, while being also capable of adapting its behaviour to its users' characteristics, are handled by using intelligent hybrid, neurofuzzy, approaches.

Strategic Objectives: *Cognitive Systems – Multimodal Interfaces*

http://www.ilsp.gr/ermis_eng.html

METIS - IST: The aim of the project was to assess the possibility to obtain free text translations of reasonably high quality from large annotated monolingual corpora with pattern-matching techniques.

Strategic Objectives: *Multimodal Interfaces – Research for Innovative Government*

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

ML-IMAGES - eContent: Development of a system for the multilingual search of digitized images that are in geographically remote databases by searching for appropriate terms in a set of defined keywords. Sophisticated indexing has been applied using multilingual ontologies based on IPTC standard.

Strategic Objectives: *Semantic-based Knowledge and Content Systems – Access to and preservation of cultural and scientific resources*

<http://www.ml-images.gr/>

KAIROS: The project aimed at the automatic composition and generation of weather forecast drafts for two languages, English and Greek, on the basis of numerical meteorological data. The project's innovation lies in the modular architecture design allowing immediate integration of existing linguistic resources, Natural Language Generation (NLG) mechanism, use of XML structures in all processing stages of the system, direct access of expert users to linguistic resources.

Strategic Objectives: *Semantic-based Knowledge and Content Systems*

http://www.ilsp.gr/kairos_eng.html

Projects & Technologies

oikONOMIA: Surface Text Understanding for the Efficient Indexing and Information Extraction from Financial Documents. The project aimed at the construction of a modular system integrating NLP tools in a pipeline, which performed text analysis for the production of a semantic representation suitable for template filling in scenario based information extraction (IE) applications.

Strategic Objectives: *Semantic-based Knowledge and Content Systems*

http://www.ilsp.gr/oikonomia_eng.html

LEXIS: The project LEXIS aimed at the construction of large special and general linguistic resources for the Greek language, as well as at the development of environments for their management. The main deliverables of the project were the following: *a Greek Computational Lexicon* of general language based on corpora, which comprises 60,000 entries containing morphological information, of which 30,000 entries also contains syntactic information and a further 15,000 includes semantic information; *Bilingual dictionaries* of the domain "Environment" (i.e. Ecology, Environmental Chemistry, Nuclear Chemistry, Climatology, Meteorology) and relevant scientific texts' collection; *A user-friendly workbench* which supports the management of the lexical and the textual data.

Strategic Objectives: *Semantic-based Knowledge and Content Systems*

http://www.ilsp.gr/lexis_eng.html

CLIO: The project aims at enhancing effective content management at the business sector, taking into account various cultural and linguistic aspects. It provides a common interface for both ERP (Enterprise Resource Planning) and Workflow Management systems through the use of *ontologies*. The main project goals are to create an integrated framework for content and workflow management, to develop a system for creating and handling customised reports and user interfaces in an interactive way and to enhance the existing workflow management systems by employing ontologies, leading to the formation of higher-level interfaces. Two ontologies were developed for the following processes: *travel management* and *recruitment*.



Strategic Objectives: *Semantic-based Knowledge and Content Systems – Research for Innovative Government*

<http://research.unisoft.gr/clio/index.htm>

READING ENVIRONMENT FOR THE BLIND: the creation of a reading environment for the Blind with the use of speech synthesis for two languages, namely English and Greek. It involved the development of a special software that recognizes and "speaks out" the verbal elements of the computer screen, recognizes the language in which these elements are written and makes a topological description of the active PC Windows applications.

ILSP has developed a specific platform for reading musical scores by the blind, in the framework of the *WEDELMUSIC - IST* project

Strategic Objectives: *eInclusion – Cognitive Systems - Research for Innovative Government*

http://www.ilsp.gr/amea_eng.html

SYNENNOISI – PROKLISI – NOEMA: A complete educational and training suite for users of the *Greek Sign Language* (GSL). It comprises the DVD-ROMs *NOEMA* (Dictionary of the GSL) and *Children's Dictionary of the GSL*; the *PROKLISI* framework of electronic terminology resources of the GSL for the field of window based human-computer interaction; the *SYNENNOISI* educational platform integrating avatar and animation technologies along with exploitation of electronic linguistic resources of GSL. This suite addresses difficulties in job finding and in remaining in the job market that are faced by people with disabilities.



Strategic Objectives: *eInclusion – Cognitive Systems – Research for Innovative Government – Technology enhanced Learning*

http://www.ilsp.gr/noema_eng.html

MYTHE - IST: The *MYTHE* project dealt with the design and development of a multilingual interactive, computer-based language learning environment for young children around the transition-to-literacy age providing additive value to Computer Assisted Language Learning (CALL) with the incorporation of advanced Linguistic and Imaging tools (LIM).

Strategic Objectives: *Technology enhanced Learning*

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

IMUTUS - IST: The main goal of *IMUTUS* is to provide an interactive music tuition multimedia system for training users on traditional instruments with no MIDI (Musical Instrument Digital Interface) output. The selected instrument is the recorder. The system will be based on audio / optical recognition, multimedia, virtual reality and audio-to-MIDI transformation technologies.

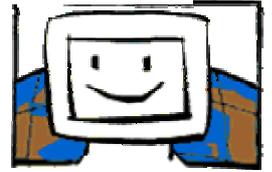


Strategic Objectives: *Technology enhanced Learning*

<http://www.exodus.gr/imutus/>

Projects & Technologies

VIRLAN (Multimedia in Education Task Force): The goal of the project was to provide a virtual real-time language learning network for primary school children. Through the facilities of virtual reality and avatar technology, the children "enter" the central meeting place and from there "travel", via the Internet, to different virtual countries, in order to meet and communicate with other children. The communication is facilitated by means of educational games and dictionaries in various languages.



Strategic Objectives: *Technology enhanced Learning*

http://www.ilsp.gr/virlan_eng.html

LRC - SOCRATES: This project aimed to improve and develop support for language teaching and learning by sharing expertise between established and emerging Language Resource Centres. It brought together a consortium of 17 partner organisations from 16 European countries. The LRC project aimed to help centres expand their access by introducing new languages (especially LWULT languages), new services (e.g. distant access, new media) and increasing the number of users (e.g. opening up to a wider public).

Strategic Objectives: *Technology enhanced Learning*

<http://www.lrcnet.org/>

Tr.AID (Technology): Rapid changes in the global marketplace have given rise to new demands and have provided new opportunities for the translation industry. The incorporation of new technologies in the work environment gives a competitive advantage and is a requirement necessary for the needs of quality and cost-efficient translation. The Tr.AID platform was designed and developed so as to meet the above needs and aims to provide translators with solutions that help them ensure quality, consistency and productivity. Tr.AID enhances the translator's work by deploying and organizing existing translation and terminology assets, and making use of such assets during the translation of new documents.



Strategic Objectives: *Multimodal Interfaces – Research for Innovative Government*

http://www.ilsp.gr/traid_eng.html

EKFONITIS (technology): An automatic text-to-speech synthesis system for Greek. It is the outcome of many years of R&D effort. EKFONITIS+ contains a brand new synthesis engine (time-domain concatenative algorithms and efficient prosody algorithms), the state of the art in this field and a completely upgraded user interface. Moreover, it contains the *Voice Assistant*, an invaluable companion for elderly and visually impaired persons. The Voice Assistant 'keeps an eye' on the typing process and utters the words and sentences just after their completion. ILSP TTS technology includes off-the-self tools for voice-enabled agent applications.



Strategic Objectives: *eInclusion – Cognitive Systems – Multimodal Interfaces – Research for Innovative Government – Technology enhanced Learning*

http://www.ilsp.gr/ekfonitis_plus_eng.html

HELLENIC NATIONAL CORPUS – HNC (technology): The HNC currently contains more than **34,000,000** words of written texts. Users can retrieve parts of these texts in the form of whole sentences by making queries based on one to three words, lemmata or parts of speech. What is more, users can define the maximum distance between search items as well as the specific sub-corpus they wish to make queries in. Finally, users can also look for certain statistical data and word concordances. ILSP is continuously active in all areas of corpora management.



Strategic Objectives: *Semantic-based Knowledge and Content Systems*

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

SYMFONIA – TEXT CORRECTION (technology): SYMFONIA is a spelling checker with 'extended' functionality compared to other commercial spelling checkers. Its basic features include the spelling checking, agreement checking, enrichment of the dictionary with unknown words in their full paradigm and grammatical characterisation of the words.



http://www.ilsp.gr/correct_eng.html

Contact Information

Dr Markos Dendrinou, R&D Project Office

mark@ilsp.gr, Phone: +30210 6875410, Fax: +30210 6856794

Postal Address

Institute for Language and Speech Processing
Artemidos 6 & Epidavrou
151 25 Paradisos Amarousiou
Athens, Greece

Mr Anastasios Patrikakos, Technology Transfer Office

tasos_p@ilsp.gr, Phone: +30210 6875324, Fax: +30210 6856794