# CELI

*Beata Dobrzyńska, Luca Dini*

May 20, 2010
Valletta, Malta

## About us

**CELI** was established in 1999 in Turin, Italy by a group of language technology researchers with the aim to address the growing market in language engineering.

Today, CELI is a provider of software solutions in the field of Natural Language Processing (NLP). Our **mono-, multi-, and cross-language products and services** apply mainly to language processing, semantic analysis, speech technology, text mining and information retrieval.

With **30 employees and collaborators**, CELI operates in **Turin** and, **since 2006 in Grenoble and Paris, France.**

## Our mission

Our mission is to improve the process of extracting meaningful intelligence from unstructured multilingual text by developing the industry's best linguistic resources, software and applications.

# Our business focus

## Technology

Language technology for multiple applications

We provide solutions integrating innovative technologies for natural language comprehension and text analysis, for real-world, business-oriented, multilanguage and cross-language applications

## Consulting

Expertise in the development of linguistic technologies

We offer CELI's competence and experience in the planning and implementation of our innovative solutions. Our team of linguistic experts and computer scientists supports our clients in multiple ways, from advising on technology selection to language analysis to project management

## Research

Innovation, technology transfer and base research

Our research team is an integral part of our organization, keeping our solutions at the forefront of the technology curve. CELI participates in various research projects, in cooperation with European centers of excellence, universities & innovative companies

# Technology

**Sophia** SEMANTIC ENGINE — Software for linguistic analysis

**DocDigger** — Search engine

**BlogMeter** — Platform for analysis of web conversations

**memesphere** Dialoghi, buzz, persone e umori della Rete — Free service of BlogMeter

**CLS** cross library services — Services for multilingual search in digital libraries

**MorSe** — Multilingual Morphological Analysis Web-Service

# Consulting

## Language Consulting

- ↗ CELI provides expertise in language analysis from extraction, to interpretation, to use

## Software Development

- ↗ We have experienced engineers who are able to assist with a wide range of NLP technology

## Resource Development

- ↗ Our experts can guide you in the best practices in resource design, creation and management

## International Cooperation

*We are currently involved in several international collaborative projects, all concerning our core research activities*

## A global prospective

*We are firmly convinced that a crucial role of language technologies is the consideration of all languages, not only the handful of popular languages*

## Research and Engineering

*We attach a great importance to the software quality of components issued by research to ensure transfer to the production environment*

### Recent examples of research activities

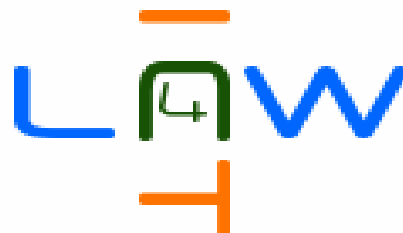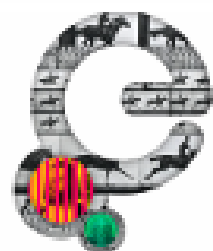| NLP | Cross-Language Information Retrieval |
|---|---|
| We are addressing semantic network Induction semantic role labeling, anaphora resolution, opinion monitoring, sentiment analysis, others | We are investigating Cross-Language Delegated Search which create a multilingual version of a user query irrespective of the underlying corpus |
| **Corpus-based Semantic Networks** | Convergence |
| We are experimenting with state-of-the-art technologies such as Latent Semantic Analysis, Random Indexing, Latent Dirichlet allocation… | Other current research areas are: query log analysis, machine translation, and Markov logic networks for language driven reasoning |

## Research projects



GALATEAS

LAW

SCOOP

LOIS

europeana
pensare la cultura

CACAO
PROJECT
CROSS-LANGUAGE ACCESS TO CATALOGUES AND ON-LINE LIBRARIES

multilingual
tourist information
on the web
mietta
roma//saarland//turku

DeepThought

DALOS

## Laboratory for Innovative Multilanguage Engineering (LIME)

CELI LIME has evolved to meet the multilanguage needs of our clients and is built on the same concept of providing technology, consulting and research as needed.
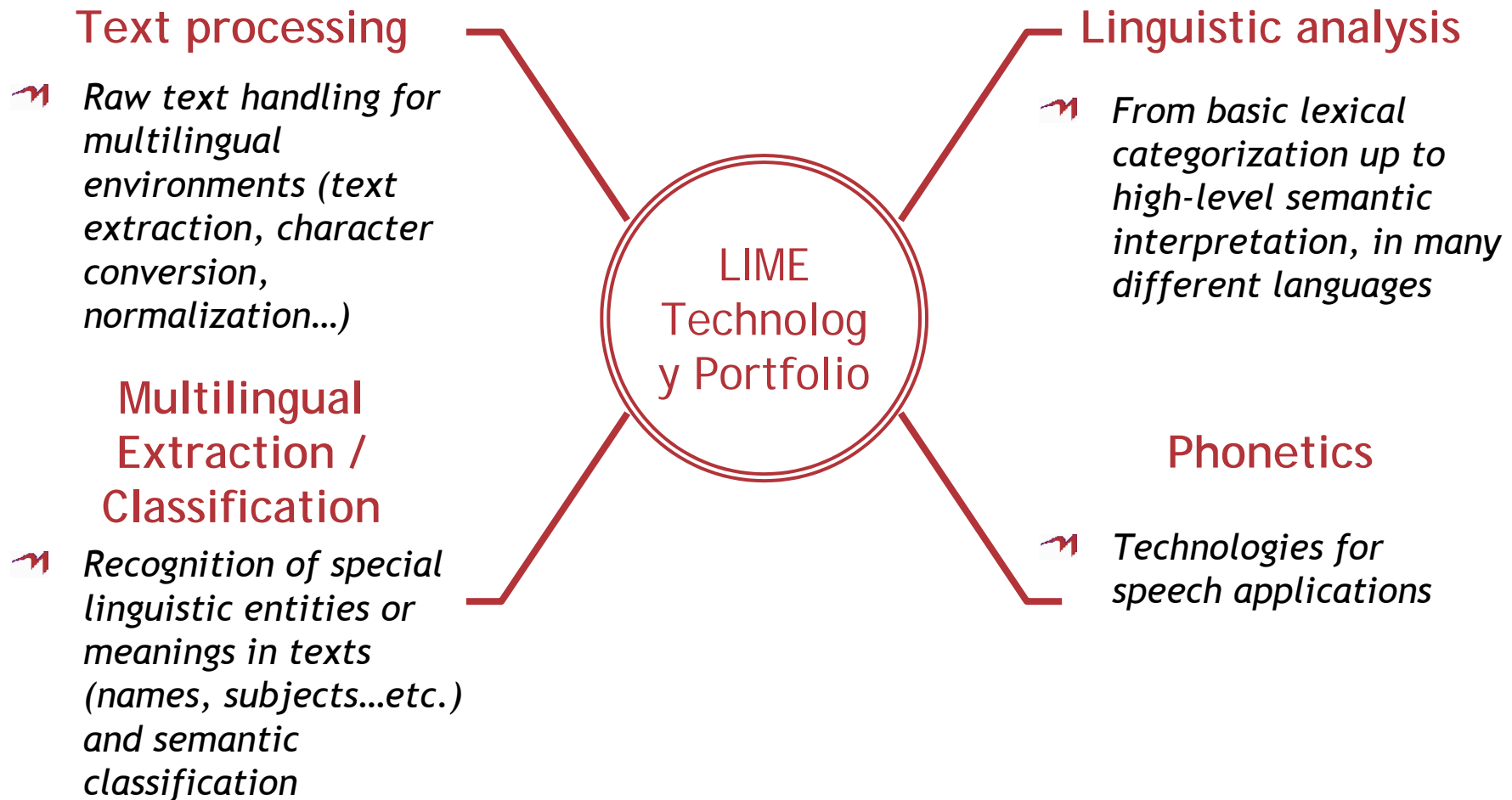
We offer a unique value proposition by having a team of technology engineers, language consultants, and project managers who can deliver a custom solution.

We are aware of the fact that for many applications language resources need to be customized in terms of coverage, domain, style and purpose. The full fledged industrial process of resources development and maintenance in place at CELI, as well as the on-site availability of native speakers for several covered languages, makes this customization process feasible, fast, and cost effective. Moreover our expert language engineer can advise and support the customer in the process of language resources selection  and integration.

# CELI LIME competencies

- LIME expertise cover many levels of language analysis from tokenization to chunking, to phonetics

- LIME systems and modules handle multilinguism and multilanguage systems

- LIME technology engineers analyze and advise on what kind of language technology is best for our client's products

- We design software to accurately and precisely meet our client's language processing needs

- We create lingware solutions for language processing

- We integrate our language processing solutions into more complex systems

# CELI LIME technologies

## Text processing

⚡ *Raw text handling for multilingual environments (text extraction, character conversion, normalization...)*

## Multilingual Extraction / Classification

⚡ *Recognition of special linguistic entities or meanings in texts (names, subjects...etc.) and semantic classification*

LIME Technology Portfolio

## Linguistic analysis

⚡ *From basic lexical categorization up to high-level semantic interpretation, in many different languages*

## Phonetics

⚡ *Technologies for speech applications*

# CELI LIME resources

**Language Corpora**
- *Text corpora, representative of "core" languages or of specific domains sublanguages. Provided with different levels of morphosyntactic annotation.*

**Lexicons**
- *Terminological dictionaries for specific domains. Custom versions can be provided according to specific application requirements.*

**Morphological databases**
- *Lexical resources built for the morphological analysis. Each word form is provided with a full set of morphosyntactic information.*

**Thesauri / Taxonomies / Ontologies / Semantic Networks**
- *Machine-readable representations of knowledge domains. Pre-packaged application-dependent and domain-dependent versions are readily available; all resources can be customized in terms of coverage, domain, style and purpose.*

**Phonetic lexicons**
- *General-purpose and terminological lexicons (e.g. toponymy) provided with standard SAMPA phonetic transcription*

**Bilingual dictionaries**
- *Translation dictionaries for many language pairs; readily integrated with morphological databases, these form the basis for every translation process*

# MorSe: Multilingual Morphological Analysis Web-Service

- MorSe is a set of **monolingual morphological analyzers** for different languages that are built with the same proprietary technology, integrated and placed behind a **common web-service interface**

- Every language-specific morphological analyzer is built on a deep and efficient modeling for each particular language in order to achieve a **precise and reliable analysis**

- In accordance with the **Software-as-a-Service** model, the analysis software is hosted by CELI and is available as a centralized service, which can be accessed remotely by the customer's applications. The client is not required to install any software, not even to maintain or upgrade any software or lexical databases

- The service is currently able to analyze texts in the following languages: Arabic, Danish, English, French, German, Italian, Norwegian , Polish , Romanian, Russian , Swedish, Turkish

- Converges with current European initiatives – compliant with major paradigm shift driven by projects such as MetaNet

# Multilanguage International Team Torino (MITTO)

**The inventory of languages covered includes:**

talian

nglish

rench

panish

atalan

ortuguese

erman

utch

wedish

orwegian

krainian

reek

urkish

rabic

ebrew

rmenian

lbanian

roatian

erbian

zech

## Our Team

- We offer a team of dedicated technology engineers, language experts, team leaders, and engagement specialists
- Our experience in addressing the specific languages needed for our project has helped us develop a specialized recruiting pipeline whereby we can readily find the mother-tongue language experts for any project
- We have a dedicated consultant manager who ensures seamless integration of the team and validation of the required language solution

# Thank you for your attention!



Tel. +39 011 5627115
Cell. +39 320 7405267

CELI srl
Via San Quintino 31 - 10121 Torino
Tel. +39 011.562.71.15 - Fax +39 011.506.40.86
info@celi
www.celi.it