



S M A R T

Statistical Multilingual Analysis
for Retrieval and Translation

Statistical Multilingual Analysis for Retrieval and Translation

Welcome!

May 13, 2009

SMART workshop, Barcelona, Spain

Nicola Cancedda



The SMART project



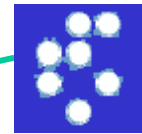
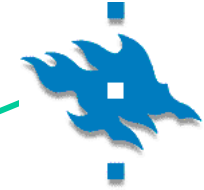
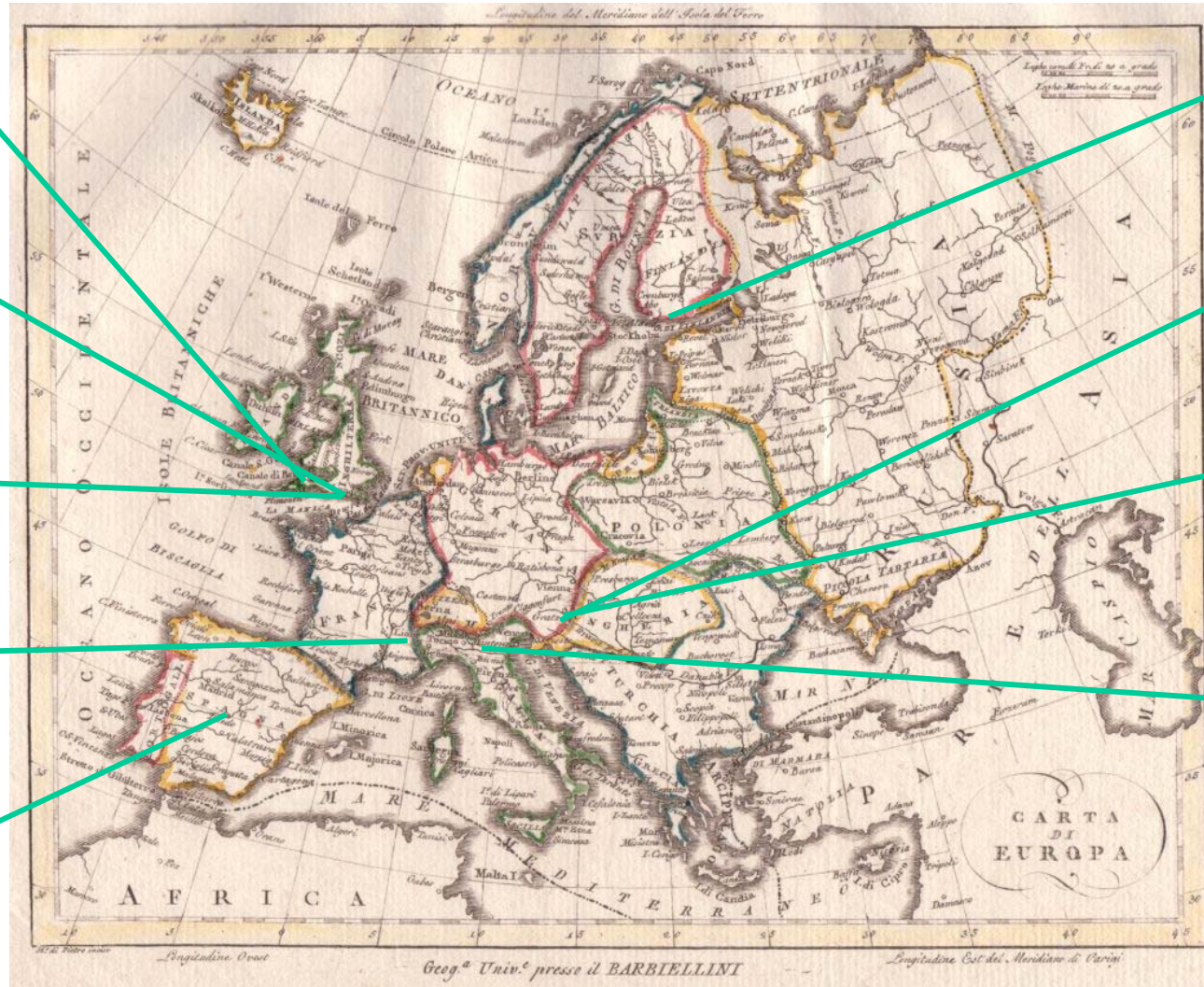
- **Information Society Technologies Programme**
- **Sixth Framework Programme, “Specific Target Research Project” (STReP)**
- **Start date: October 1, 2006**
- **Duration: 3 years**
- **35-40 Researchers involved overall**



**SIXTH FRAMEWORK
PROGRAMME**



The SMART Consortium



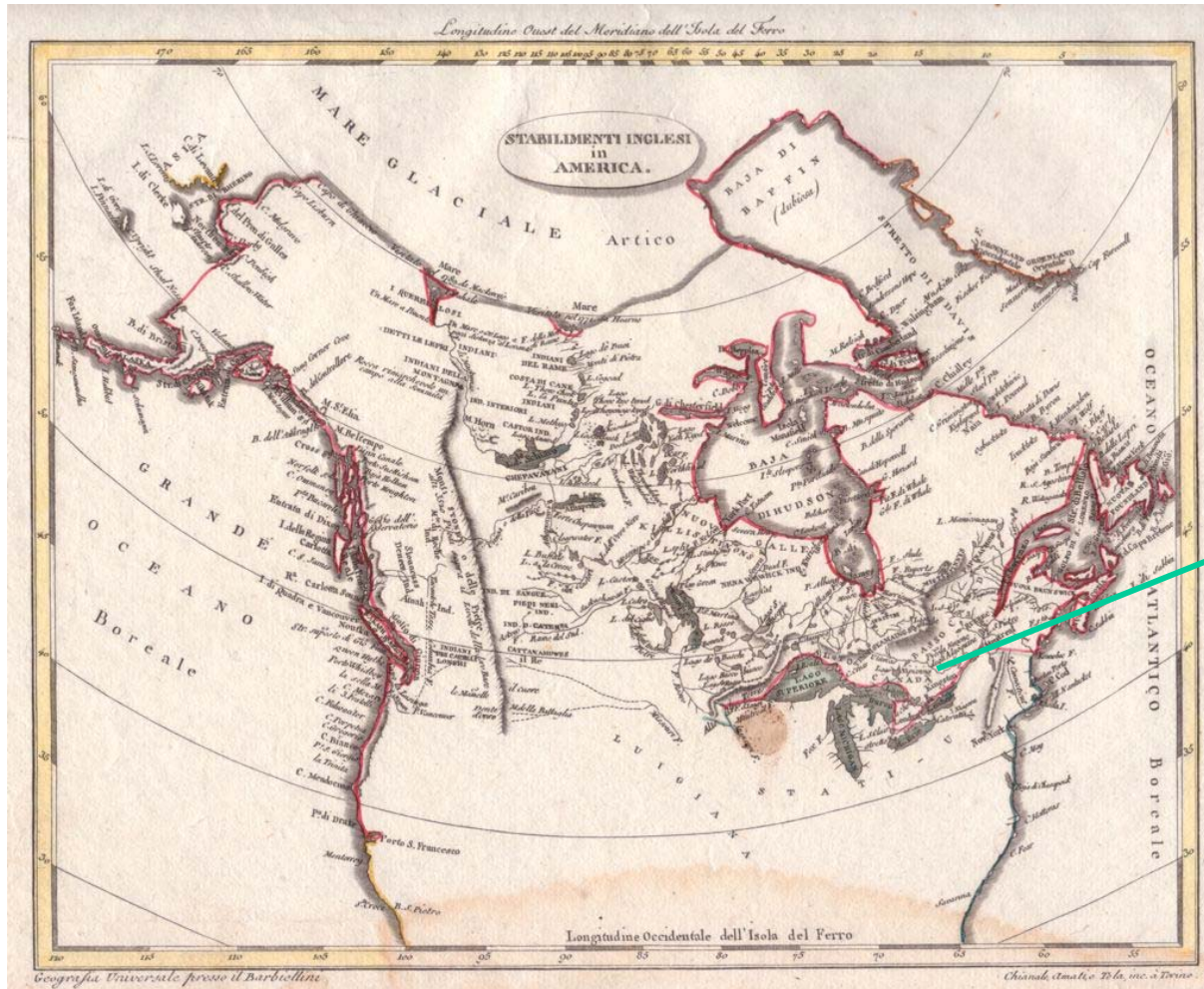
May 13, 2009

SMART workshop, Barcelona, Spain

Nicola Cancedda



The SMART Consortium



May 13, 2009

SMART workshop, Barcelona, Spain

Nicola Cancedda



Motivation

- **Almost half of the citizens of the EU do not speak a second language**
- **Expanding global demand of tools for automatic translation and cross-language retrieval, clustering and categorization**
- **Statistical approaches mainstream in research, but still suffering from shortcomings, preventing their diffusion, e.g.:**
 - *Relatively low fluency/grammaticality of output*
 - *Model training still a somewhat arcane craft*
 - *Difficult to use in new domains*
 - *Trained once for all, do not learn from constant user feedback*
- **Many recent advances in Machine Learning are only starting to hit the field**

**From the proposal
(2005)**

SMART is an attempt to propose some original solutions using the methods of Statistical Learning



- **This workshop is supported by the PASCAL 2 EC Network of Excellence**
 - *Machine Learning, Statistics, Optimization, and applications thereof*
 - *<http://www.pascal-network.org>,
<http://videolectures.net/pascal>*



Some (presented) highlights

- **Two new SMT systems**
 - *Sinuhe (U. Helsinki)*
 - See talk by Matti Kääriäinen
 - *MMR (U. Southampton)*
 - See talk by Sandor Szedmak
- **On-line adaptation of translation models**
 - *Adaptive extension (U. Milan) of NRC's PORTAGE*
 - See talk by Nicolò Cesa-Bianchi
- **A method for discriminative learning of phrase tables**
 - *See talk by Zhuoran Wang (UCL)*



Some (presented) highlights

- **An extensive analysis of an SMT system as a learning system**
 - *See talk by Marco Turchi (U. Bristol)*
- **New methods for sentence-level confidence estimation**
 - *See talk by Lucia Specia (Xerox)*
- **A study on detecting and exploiting translation direction**
 - *See talk by Cyril Goutte (NRC)*
- **Scale-up of methods derived from Canonical Correlation Analysis**
 - *See talk by Blaz Fortuna (Jozef Stefan Institute)*



and also...

- **New discriminatively trained Language Models for SMT**
 - **New lexicon adaptation methods for CLIR**
 - **A regression-based approach to translation**
 - **Methods for extracting bilingual lexica from the Wikipedia**
 - **A prototype-demo for adaptive Computer-Aided Translation**
 - **A prototype-demo for CLIR and MT of the Wikipedia**
- ...check-out our website!**



Project Website

- **Project presentation and deliverables**

– <http://www.smart-project.eu>

SMART
Statistical Multilingual Analysis
for Retrieval and Translation

Search

navigation

- Introduction
- Partners
- Deliverables
- Papers & presentations
- Bibliography
- Internal
- Search

Home

Introduction

More than half of the EU citizens are not able to hold a conversation in a language other than their mother tongue, let alone to conduct a negotiation, or interpret a law. In a time of wide availability of communication technologies, language barriers are a serious bottleneck to European integration and to economic and cultural exchanges in general. More effective tools to overcome such barriers, in the form of software for machine translation and other cross-lingual textual information access tasks, are in strong demand.

May 13, 2009

SMART workshop, Barcelona, Spain

Nicola Cancedda



But first of all...



Invited Speaker

- **Jesús Giménez, Universitat Politècnica de Catalunya**
 - *SVM tools for sequential labelling problems*
 - *Semantic Role labeling*
 - *Empirical Machine Translation*
 - *IQMT, a framework for combining multiple MT metrics*
 - *Today speaking on:*
 - “**Empirical Machine Translation and its Evaluation**”



Backup slides

May 13, 2009

SMART workshop, Barcelona, Spain

Nicola Cancedda



WP7 - Dissemination and Exploitation

- **Platforms for showcasing developed tools:**

The screenshot displays the SMART web interface. At the top, there is a language selector set to 'Slovenščina' and a 'query language' dropdown. The main header features the SMART logo and the text 'Statistical Multilingual Approach for Retrieval and Translation'. Below the header, a search bar shows '1-10 results (took 2.090 seconds)'. The results are categorized into 'Pollution' and 'Radioactive Pollution'. The 'Pollution' section includes a snippet about the waste-return channel for Lachine in Montreal. The 'Radioactive Pollution' section includes a snippet about nuclear waste. On the right side, there is a 'FOUND In Translation' section with a map of Europe and a list of countries with their respective counts. Below the map, there are two news snippets: one about financial crisis and terrorist attacks, and another about Putin wanting Saakashvili hanged. The bottom right corner shows a 'Local-News' and 'EU-News' section with a list of countries and their counts.

SMART
Statistical Multilingual Approach for Retrieval and Translation

Slovenščina query language

FOUND In Translation
Found in Translation monitors news outlets in languages. [More...](#)

EU Activity
0 1 2 3 4 5 6 7 8 9

EU Share

- ★ EU (52.3%)
- ★ Belgium (0.9%)
- ★ Czech Republic (4.5%)
- ★ Other (42.3%)

Weather

- ★ EU (242)
- ★ Belgium (4)
- ★ Czech Republic (21)
- ★ Denmark (3)
- ★ Estonia (3)
- ★ France (50)
- ★ Greece (19)
- ★ Italy (50)
- ★ Latvia (1)
- ★ Netherlands (10)

Weather | **EU News** | **Local News** | **Text Only**

EU (242)
'financial crisis is likely to terrorist attacks grow'
According to the American intelligence agencies, the global economic crisis to political turbulence in conflict areas and an increase in attacks. [Read more...](#)

'Putin wanted Saakashvili hanged'

Search
Search

Local-News

- ★ EU (232)
- ★ Belgium (4)
- ★ Cyprus (2)
- ★ Czech Republic (9)
- ★ Denmark (5)
- ★ France (50)
- ★ Greece (15)
- ★ Italy (50)
- ★ Latvia (1)
- ★ Netherlands (10)
- ★ Portugal (5)
- ★ Slovenia (8)
- ★ Spain (7)
- ★ Sweden (13)
- ★ Switzerland (1)
- ★ United Kingdom (24)

EU-News

- ★ EU (232)
- ★ Belgium (4)
- ★ Cyprus (2)

May 13, 2009

SMART workshop, Barcelona, Spain

Nicola Cancedda