



Multilingual On-Line Translation

Olga Caprotti <caprotti@chalmers.se>

Cristina España <cristinae@lsi.upc.edu>

Lluís Màrquez <lluism@lsi.upc.edu>

Aarne Ranta <aarne@chalmers.se>

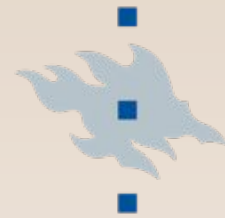
Project summary

MOLTO's goal is to develop a set of tools for translating texts between multiple languages in real time with high quality. Languages are separate modules in the tool and can be varied; prototypes covering a majority of the EU's 23 official languages will be built.

Consortium



GÖTEBORGS UNIVERSITET



HELSINGIN YLIOPISTO



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

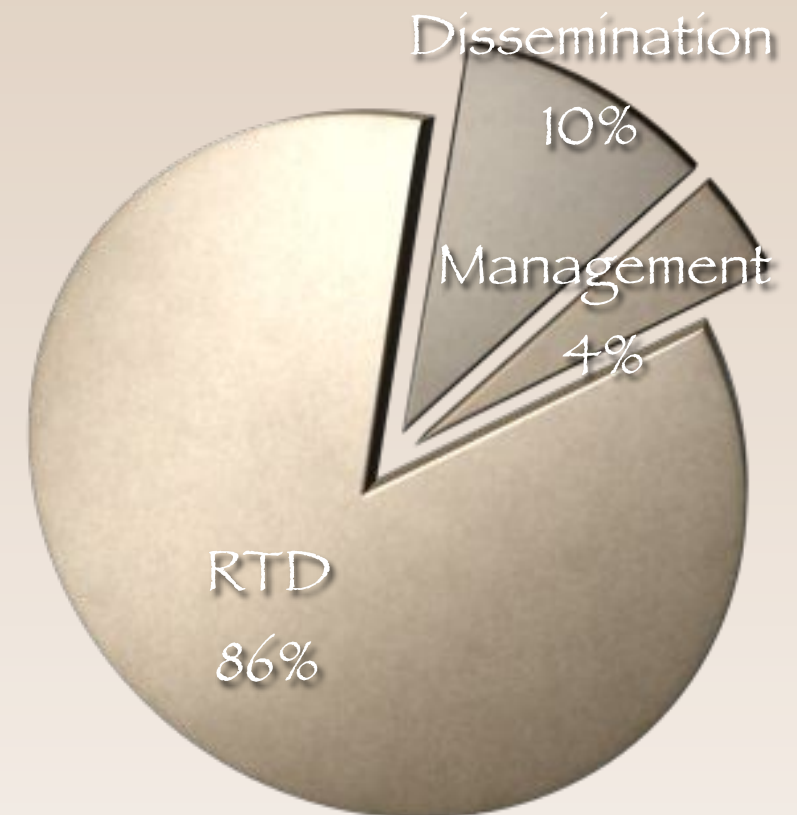


ontotext
Semantic Technology Lab



How much?

- ♦ Total: 3,000,000 EUR, EC contribution 2,375,000 EUR
- ♦ 90% for work (390 person months)
- ♦ 1 March 2010 – 28 February 2013



What's new

	Google / Babelfish	MOLTO
target user	consumer	producer
input	unpredictable	predictable
coverage	unlimited	limited
quality	browsing	publishing

MOLTO domains

- ♦ Mathematical exercises (WebALT)
- ♦ Biomedical and pharmaceutical patents
- ♦ Museum object descriptions

More potential uses

- ♦ Wikipedia articles
- ♦ E-commerce sites
- ♦ Medical treatment recommendations
- ♦ Tourist phrasebooks
- ♦ Social media
- ♦ SMS

MOLTO technologies

GF

grammaticalframework.org

Statistical Machine Translation

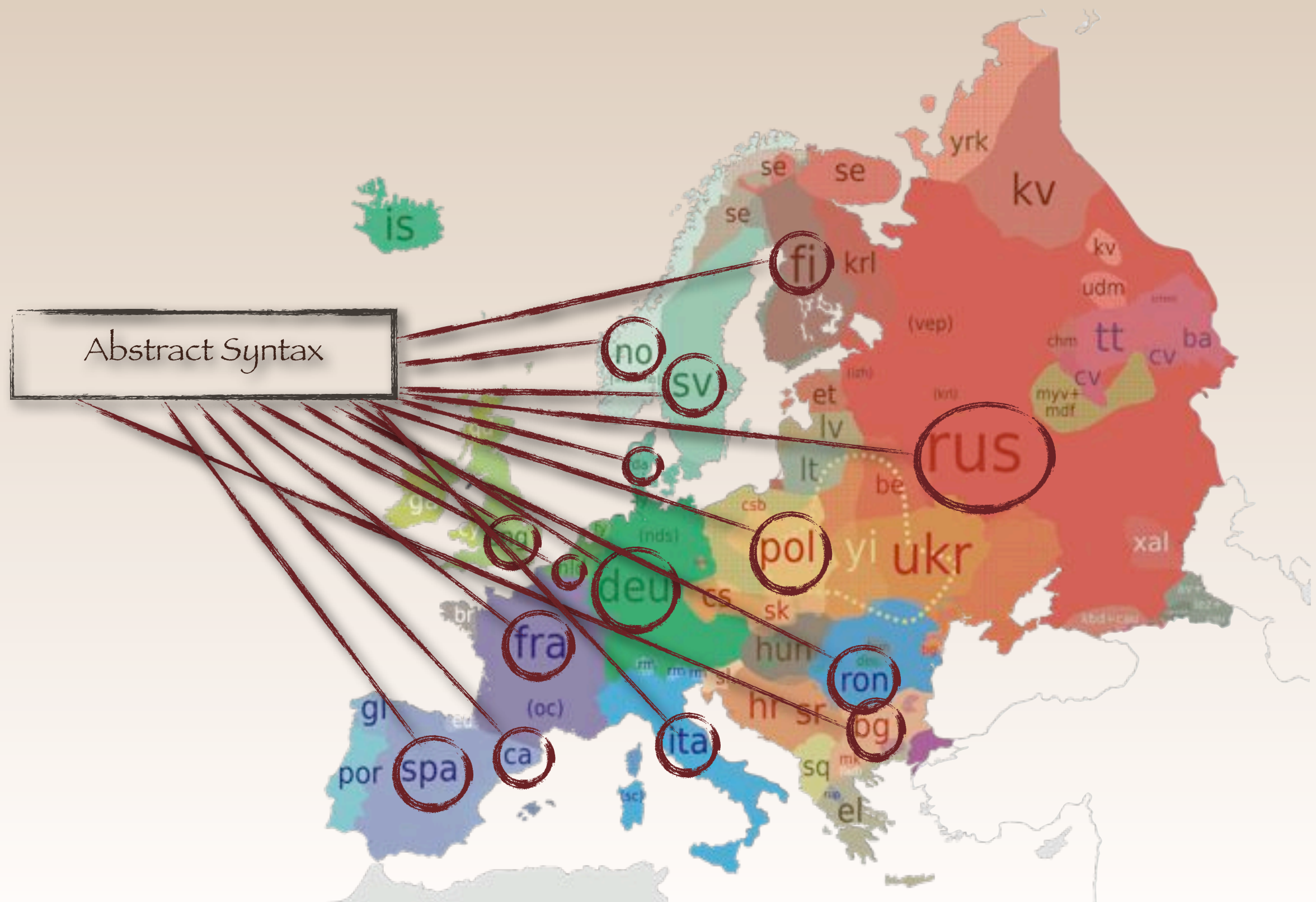
OWL Ontologies

GF – Grammatical Framework

Core of MOLTO is a *multilingual GF grammar*:

- ♦ meaning-preserving translation by composition of parsing and generation
- ♦ abstract syntax as interlingua
- ♦ RGL, GF Resource Grammar Library, for inflectional morphology and syntactic combination functions of 16 languages

MOLTO Languages



Domain-specific interlinguas

The abstract syntax must be formally specified, well-understood

- ❖ semantic model for translation
- ❖ fixed word senses
- ❖ proper idioms

e.g.
a mathematical theory, an ontology

Statistical Machine Translation

Main goal:

improve robustness of raw GF on a quasi-open domain by statistical machine translation

Robustness and Statistics

MOLTO research topics:

- ❖ SMT as fall-back
- ❖ Hybrid systems
- ❖ Learning of GF grammars by statistics
- ❖ Improving SMT by grammars

Innovation: OWL interoperability

OWL as a way to specify interlinguas:

- ❖ 2-way transformation ontology-grammar
- ❖ Web pages with ontologies... will soon be equipped by translation systems
- ❖ Natural language search and inference

NL Knowledge Management

The MOLTO infrastructure will

- ❖ semi-automatically create abstract grammars from ontologies;
- ❖ derive ontologies from grammars;
- ❖ retrieve instance level knowledge from / in NL by transforming queries to semantic queries, and by expressing the knowledge in NL.

First results

- Online Demo, Jun 2010 at molto-project.eu
- Knowledge Representation Infrastructure, Nov 2010
- GF Grammar Compiler API, Mar 2011