MOLTO

non multa, sed multum

Multilingual On-Line Translation

MOLTO Consortium

SEVENTH FRAMEWORK PROGRAMME

FP7-247914

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

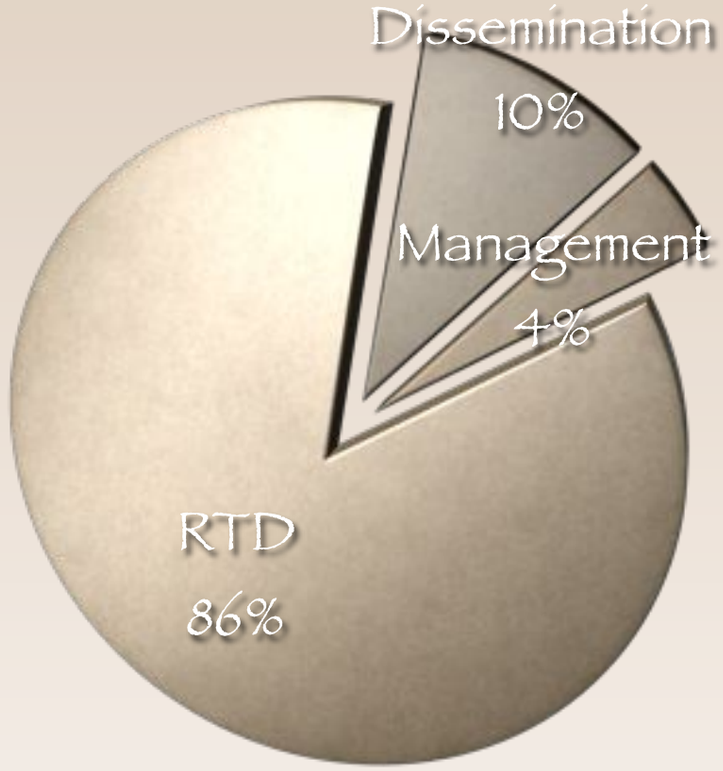
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

ontotext
Semantic Technology Lab

MATRIXWARE
INFORMATION SERVICES

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

Translation directions

Statistical methods

work best to English

- rigid word order
- simple morphology

Grammar-based methods

work equally well for different languages

- German word order
- Finnish cases

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

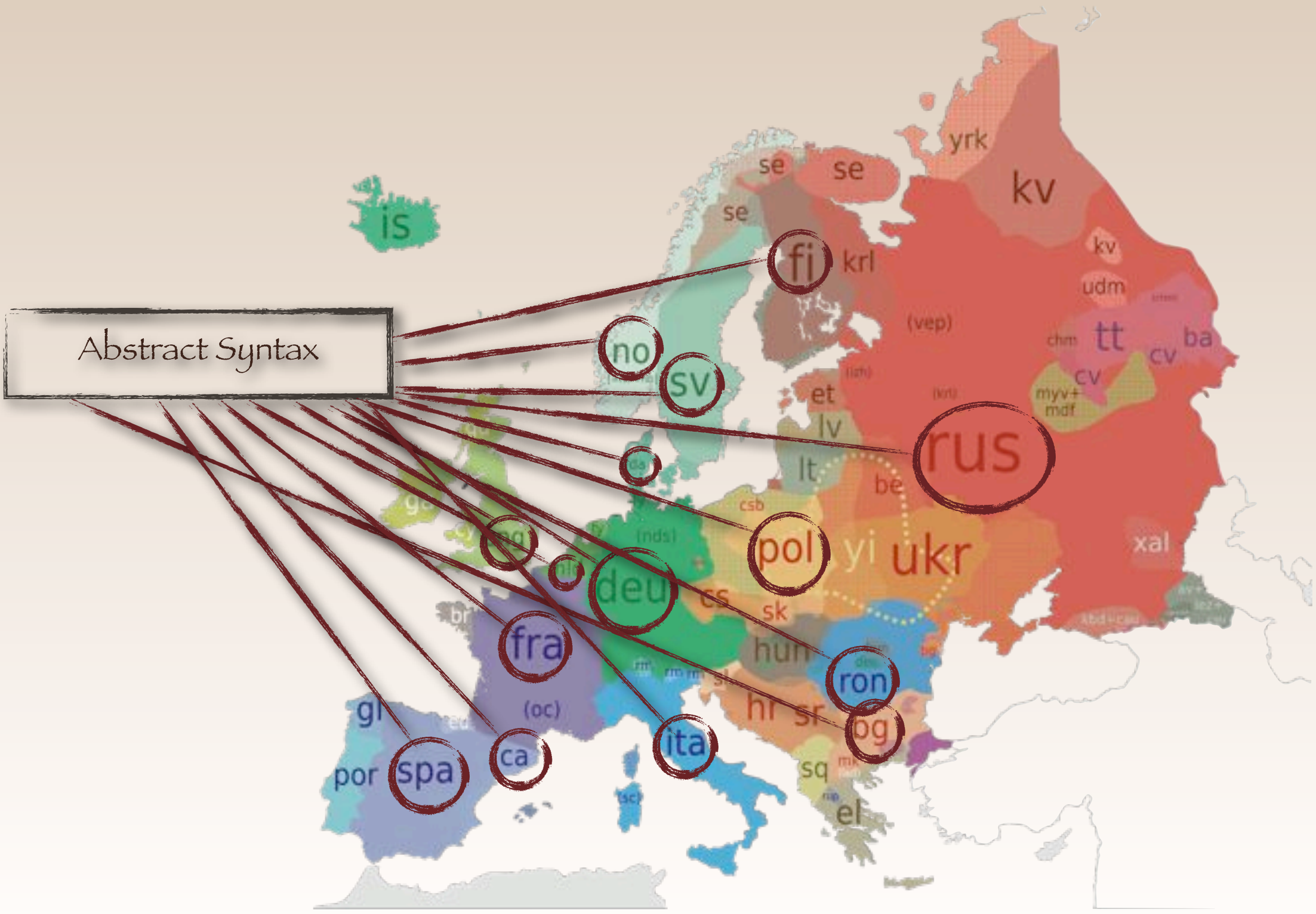
OWL Ontologies

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

Grammar tools

Scale up production of domain interpreters

100's of words
GF experts
months
hand-crafting a grammar

1000's of words
domain experts & translators
days
translating a set of examples

Challenge

Mathematics

Grammar generalization

Abstract syntax

```
Nat : Set
Even : Exp -> Prop
Odd : Exp -> Prop
Gt : Exp -> Exp -> Prop
Sum : Exp -> Exp
```

English concrete syntax (by examples)

```
Nat = "number"
Even x = "x is even"
Odd x = "x is odd"
Gt x y = "x is greater than y"
Sum x = "the sum of x"
...
```

every even number that is greater than 0 is the sum of two odd numbers

German concrete syntax (by examples)

```
Nat = "Zahl"
Even x = "x ist gerade"
Odd x = "x ist ungerade"
Gt x y = "x ist größer als y"
Sum x = "die Summe von x"
...
```

jede gerade Zahl, die größer als 0 ist, ist die Summe von zwei ungeraden Zahlen

Translator's tools

- text input + prediction
- syntax editor for modification
- disambiguation
- on the fly extension
- normal workflows: API for plug-ins in standard tools, web, mobile phones...

Authoring: document edits

Authoring: document edits

Chère Madame X,

j'ai l'honneur de vous informer que vous avez été promue chargée de projet.

Avec mes salutations distinguées, le président.

Authoring: document edits

Madame X → Monsieur Y

Chère Monsieur Y,

j'ai l'honneur de vous informer que vous avez été promue chargée de projet.

Avec mes salutations distinguées, le président.

Authoring: syntax edits

Mrs X → Mr Y

```
Letter (Dear (Mrs "X"))
(Honour (Promote
ProjectManager))
(Formal President)
```

Chère Madame X,

j'ai l'honneur de vous informer que vous avez été promue chargée de projet.

Avec mes salutations distinguées, le président.

```
Letter (Dear (Mr "Y"))
(Honour (Promote
ProjectManager))
(Formal President)
```

Cher Monsieur Y,

j'ai l'honneur de vous informer que vous avez été promu chargé de projet.

Avec mes salutations distinguées, le président.

Statistical Machine Translation

Main goal:

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

Robustness & statistics

- Statistical Machine Translation as fall-back
- Hybrid systems
- Learning of GF grammars by statistics
- Improving SMT by grammars

Challenge

Models of hybrid MT systems

- **baseline**: cascade of independent MT systems;
- **hard integration**: GF partial output is fixed in a regular SMT decoding;
- **soft integration I**: GF partial output, as phrase pairs, is integrated as a discriminative probability feature model in a phrase-based SMT system;
- **soft integration II**: GF partial output, as tree fragment pairs, is integrated as a discriminative probability model in a syntax-based SMT system.

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.

OWL↔Grammar (sketch)

`Class(pp:Nat ...)`	`cat Nat`
`ObjectProperty(pp:Odd domain(pp:Nat))`	`fun Odd: Nat->Prop`
`ObjectProperty(pp:Gt domain(pp:Nat) range(pp:Nat))`	`fun Gt: Nat->Nat->Prop`

First results

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