

MOLTO = Multilingual On-Line Translation

Aarne Ranta, MOLTO Kick-Off Meeting, Barcelona, 9 March 2010

FP-7 Project nr 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.

...

# Gothenburg University News, 23 December

## Multilingual translation system receives 25 millions in EU funding

All citizens, regardless of native tongue, shall have the same access to knowledge on the Internet. The MOLTO project, coordinated by University of Gothenburg, receives more than 25 million SEK (2.375 million euro) in project support from the EU to create a reliable translation tool that covers a majority of the EU languages.

Google Translator is a widely spread translation programme that gradually improves the quality of translations through machine learning ...

In contrast, MOLTO is being developed in the opposite direction, meaning it begins with precision and grammar, while wide coverage comes later. We wanted to work with a translation technique that is so accurate that people who produce texts can use our translations directly.

# The Local, 19 January

## Swedish translation system gets EU funding

A research group led by the University of Gothenburg in the west of Sweden, has been granted 25 million kronor (\$3.5 million) in EU funding to develop an online multilingual translation system covering most European languages.

The system, which can translate a number of languages simultaneously, will be more accurate and reliable than existing programmes, says Aarne Ranta, professor of computer science and engineering at the University of Gothenburg.

"The current tools out there, such as Google Translator, do not give good quality translations. They are okay for browsing and to get an idea of what is in a text but most people would not use them to translate something important," he told The Local.

<http://www.thelocal.se/24468/20100119/>

## Comments in The Local

- It would be nice to have translations that didnt sound like instructions written in China.
- This is a very good EU funded project, which will increase communication across borders and make doing business in other countries a lot easier.

## Comments in The Local, 2

Can't wait to see this. Try this from English to Swedish: "After given a ballpark figure, the aforementioned rubber stamper decided to 86 the whole schabang, which in turn led the subject putting a foot in her mouth. Her father then drove a wooden stake thru heart of her paramour. A frothy confection that melts in your hand ensued." – Can a software program do this?

## Comments in The Local, 3

I'll bet anybody this little EU project is a waste of money and will not create anything as good as the private sector.

In 5 years they might have something comparable to google translate. and google translate will be 5 years ahead.

## **Linkedin (Automatic Translation Group)**

Sorry to be cynical, but the only novelty here is the absence of the substring "euro" in the name. How many Europrojects related to MT ran for ages and ended with another promising start?

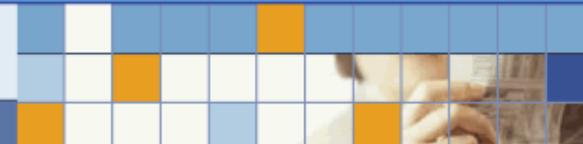
I remember talking to a participant of one of those, which started in mid-1980s. When asked about his experiences there, he reacted, "I wasn't there too long. Only the first 5 years."

Of course, I can't say with 100% certainty that it will be exactly the same, but then, it doesn't look different...

**Cordis main page 20 January**



# CORDIS News



Search articles :

Detailed search

Latest on...

- FP7
- FP6
- EU Presidency
- Lisbon Strategy
- Calls
- Events
- Interviews
- EU R&D Newsroom

research\*eu

CORDIS Express

## News highlights

### Global emissions impact springtime ozone increase above western North America

[Date: 2010-01-21]

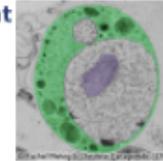
Western North America is recording higher springtime ozone levels, triggered mostly by the air that flows eastward from the Pacific... [read more](#)



### Research throws new light on evolution of bacteria

[Date: 2010-01-21]

Researchers in Germany have discovered that a relatively obscure group of related bacteria possesses proteins that were previously... [read more](#)



### Key gene helps plants adapt to climate change

[Date: 2010-01-21]

A group of researchers from the UK has discovered through genetic experiments how plants control their development through their ... [read more](#)



### EU funds effective translation tool

[Date: 2010-01-20]

Europeans recognise the importance of communicating in other languages as well as their native tongue, and the availability of ef... [read more](#)



Presidencia Española *eu* 2010.es

Presidency News

research\*eu



CORDIS *Wire*

→ Virtual tour of the Parthenon Frieze

## Some more media

La UE destina más de 2 millones a financiar un traductor online (Hoytecnologia Mon, 25 jan 2010 10:51)

La UE destina más de 2 millones de euros a financiar una herramienta de traducción eficiente en Internet (Granada Digital Sun, 24 jan 2010 13:58)

Multilingual Translation System Receives Over 2 Million Euro in EU Funding (Communications of the ACM Wed, 20 jan 2010 21:41)

L'UE finance un outil de traduction efficace (CORDIS Nouvelles Wed, 20 jan 2010 15:11)

Swedish translation system gets EU funding (European Journalism Centre Wed, 20 jan 2010 10:40)

Multilingual Translation System Receives over 2 Million Euro in EU Funding (Resource Shelf Mon, 18 jan 2010 23:12=)

University of Gothenburg Multilingual translation system receives over 2 million euro in EU funding (Pressrelations.de Mon, 18 jan 2010 15:29)

**Analytics 22-29 January 2010**



## 3,097 visits came from 82 countries/territories

Detail Level: [City](#) | [Country/Territory](#) | [Sub Continent Region](#) | [Continent](#) Dimension: [None](#) ▾

Site Usage

Goal Set 1

Views:

<b>Visits</b> <b>3,097</b> % of Site Total: 100.00%	<b>Pages/Visit</b> <b>1.12</b> Site Avg: 1.12 (0.00%)	<b>Avg. Time on Site</b> <b>00:00:28</b> Site Avg: 00:00:28 (0.00%)	<b>% New Visits</b> <b>92.12%</b> Site Avg: 92.12% (0.00%)	<b>Bounce Rate</b> <b>91.44%</b> Site Avg: 91.44% (0.00%)
---	---	---	--	---

Detail Level: <a href="#">Country/Territory</a> ▾		Visits ↓	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
1.	<a href="#">Germany</a>	933	1.08	00:00:21	95.07%	93.68%
2.	<a href="#">Spain</a>	289	1.15	00:00:23	91.35%	88.58%
3.	<a href="#">Italy</a>	222	1.12	00:00:50	87.84%	90.99%
4.	<a href="#">United Kingdom</a>	220	1.12	00:00:20	97.27%	91.36%

## **Who?**

UGOT: University of Gothenburg, Sweden (coordinator)

UHEL: University of Helsinki, Finland

UPC: Universitat Politècnica de Catalunya, Barcelona, Spain

Ontotext: Ontotext AD, Sofia, Bulgaria

Mxw: Matrixware GmbH, Vienna, Austria

## How much?

Total: 3,000,000 EUR, EC contribution 2,375,000 EUR

90% for work (390 person months)

86% for RTD, 10% dissemination, 5% management

1 March 2010 – 28 February 2013

## What's new?

<b>Tool</b>	<b>Google, Babelfish</b>	<b>MOLTO</b>
target	consumers	producers
input	unpredictable	predictable
coverage	unlimited	limited
quality	browsing	publishing

## Producer's quality

Cannot afford translating

- *prix 99 euros*

to

- *pris 99 kronor*

## Producer's quality

Cannot afford translating

- *I miss her*

to

- *je m'ennuie d'elle*  
("I'm bored of her")

# The translation directions

Statistical methods (e.g. Google translate) work the best *to* English

- rigid word order
- simple morphology
- focus of DARPA-funded research

Grammar-based methods work equally well for different languages

- Finnish cases
- German word order

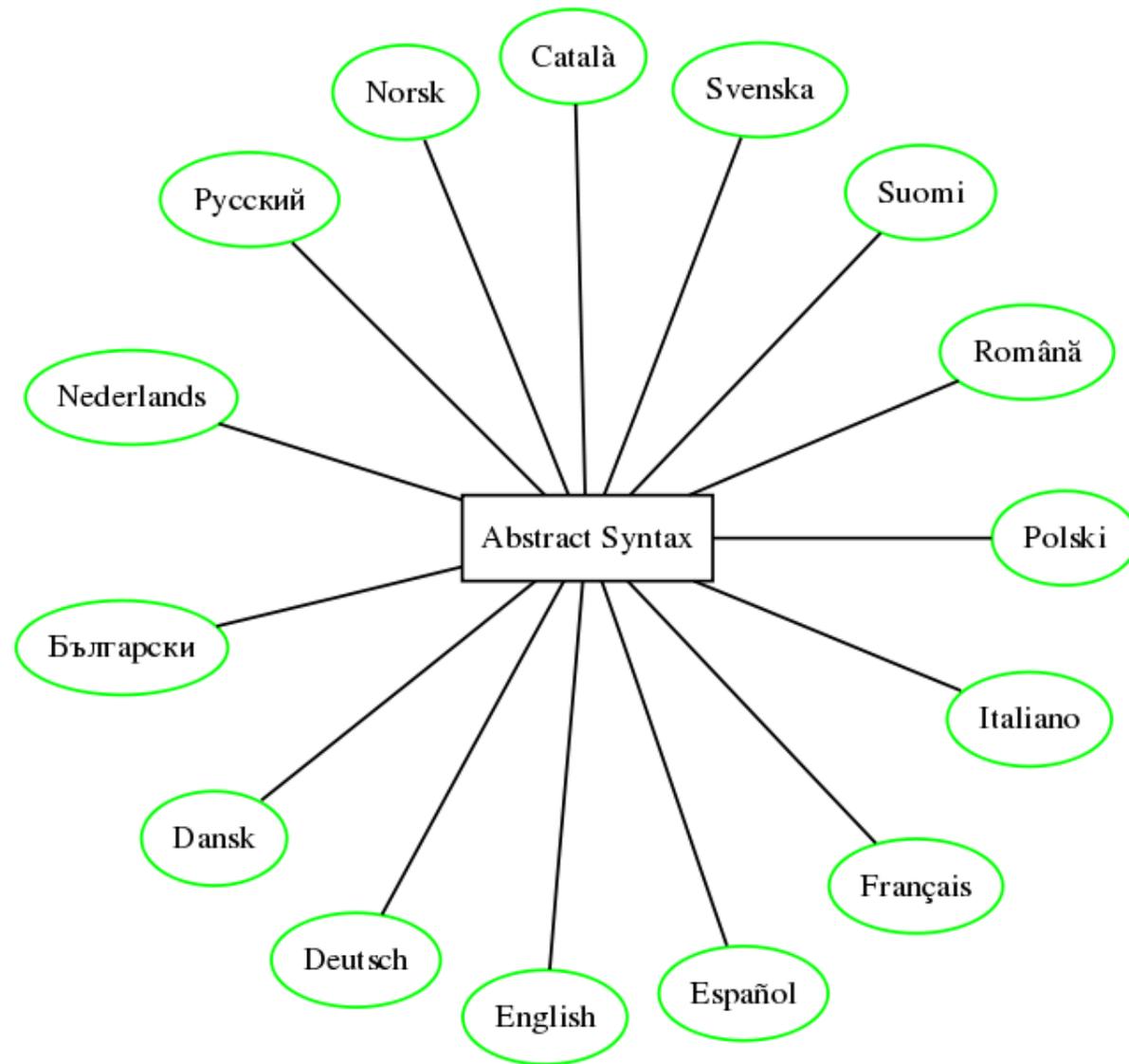
# Main technologies

GF, [grammaticalframework.org](http://grammaticalframework.org)

- Domain-specific interlingua + concrete syntaxes
- GF Resource Grammar Library
- Incremental parsing
- Syntax editing

OWL Ontologies

Statistical Machine Translation



**MOLTO languages**

## Domain-specific interlinguas

The abstract syntax must be formally specified, well-understood

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

For instance: a mathematical theory, an ontology

## Two things we do better than before

**No** universal interlingua:

- *The Rosetta stone is not a monolith, but a boulder field.*

**Yes** universal concrete syntax:

- no hand-crafted *ad hoc* grammars
- but a general-purpose resource grammar library

## **Domains for case studies**

Mathematical exercises (<- WebALT)

Patents in biomedical and pharmaceutical domain

Museum object descriptions

## **Other potential uses**

Wikipedia articles

E-commerce sites

Medical treatment recommendations

Tourist phrasebooks

Social media

SMS

## Challenge: grammar tools

Scale up production of domain interpreters

- from 100's to 1000's of words
- from GF experts to domain experts and translators
- from months to days
- writing a grammar  $\approx$  translating a set of examples

## Example: mathematics

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"

## 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"

## Resulting generalization

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

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

# Challenge: translator's tools

Transparent use:

- 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 the source: incremental parsing

The screenshot shows a web-based translation interface. At the top, a text input field contains the French sentence "a femme qui remplit le formulaire est co", with the word "est" underlined in red. To the right of the input is a "Translate" button. Below the input field, a dropdown menu is open, displaying three suggestions: "la femme qui remplit le formulaire est connue", "la femme qui remplit le formulaire est contente", and "la femme qui remplit le formulaire est correcte". To the right of the suggestions, there are two dropdown menus: one labeled "toFre" and another labeled "To: All languages".

Live demo in [grammaticalframework.org:41296](http://grammaticalframework.org:41296)

## Authoring: modifications of a document

*Chère Madame X,*

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

*Avec mes salutations distinguées, le président.*

**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.*

## Modification by syntax editing

*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 (Mrs "X")) (Honour (Promote ProjectManager)) (Formal President)

-- change Mrs "X" -> "Mr Y"

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.*

## **Innovation: OWL interoperability**

Transform web ontologies to interlinguas

Pages equipped with ontologies... will soon be equipped by translation systems

Natural language search and inference

## Example of OWL-GF conversion

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

## **Scientific challenge: robustness and statistics**

1. Statistical Machine Translation (SMT) as fall-back
2. Hybrid systems
3. Learning of GF grammars by statistics
4. Improving SMT by grammars

## Availability of MOLTO tools

Open source, LGPL

- *except* parts of the patent case study

Web demos from June 2010

# The Consortium

UGOT: grammar development tools, museum case

UHEL: translation tools, evaluation

UPC: statistical methods, mathematics case

Ontotext: ontology tools, web interfaces

Matrixware: patents case

## **Previous EU projects**

TYPES 1985-2008

TALK 2004-2006

WebALT 2005-2006

JEM 2006-2009

# The Call

ICT-2009.2.2: Language-Based Interaction

Majority of EU languages

Improvement by an order of magnitude

Use of existing linguistic resources