

# Two Years of MOLTO

Aarne Ranta

MOLTO Meeting, Zurich, 7 March 2012



# Multilingual Online Translation

Non multa, sed multum not quantity but quality

[ABOUT](#)

[NEWS](#)

[EVENTS](#)

MOLTO's mission is to develop a set of tools for translating texts between *multiple languages* in *real time* with *high quality*. MOLTO will use multilingual grammars based on semantic interlinguas.

FP7-ICT-247914, Strep, [www.molto-project.eu](http://www.molto-project.eu)

U Gothenburg, U Helsinki, UPC Barcelona, Ontotext (Sofia), U Zurich,  
Be Informed (Apeldoorn)

March 2010 - May 2013

EC contribution 2,975,000 EUR

## What's different?

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

# Producer's quality

## **Responsibility** for the translation

Cannot afford translating French

- *prix 99 euros*

to Swedish

- *pris 99 kronor*

Typical SMT error due to parallel corpus containing localized texts.  
(N.B. 99 kronor = 11 euros)

# Linguistic knowledge

(From Google Translate 1 September 2011)

Finnish: *yö, yön, yötä, yönä, yöksi, yössä, yöstä, yöhön, yöllä, yöltä, yölle, yöttä, öineen, öin, yöt, öitä, öiden, öinä, öiksi, öissä, öistä, öihin, öillä, öiltä, öille, öittä, öin*

English: *Night, night, night, night, night, night, night, night, night, night, night, nights, yöttä, öineen, night, night, nights, nights, nights States by quotas, domestic insurance companies, nights, nights, öillä, against loss, States, öittä, night*

# Predictability

German to English

- *er bringt mich um -> he is killing me*

correct, but

- *er bringt meinen besten Freund um -> he brings my best friend for*

should be *he kills my best friend*. (Typical error due to **long distance dependencies**, causes **unpredictability**)

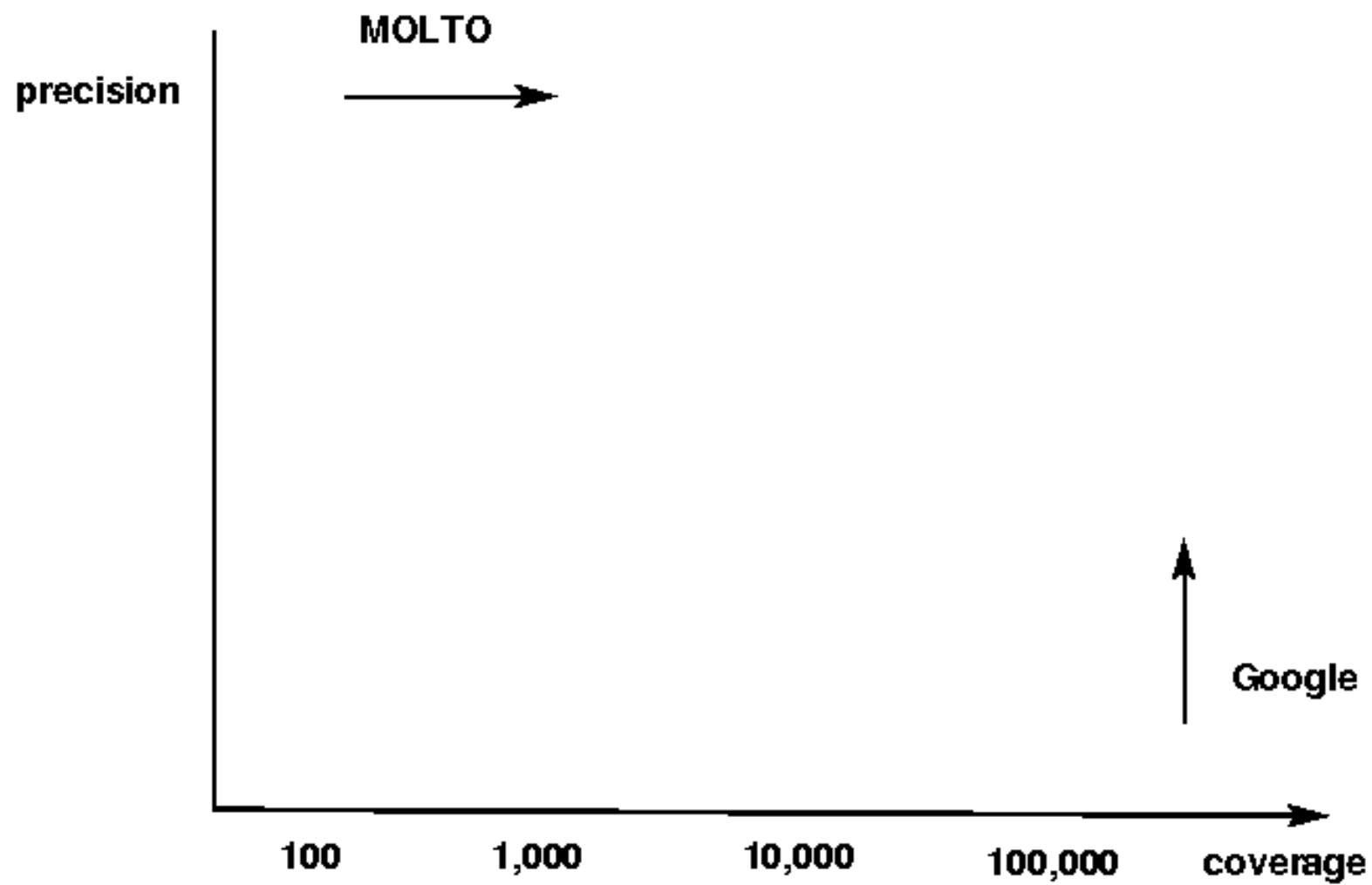
## **Aspects of reliability**

Linguistic knowledge

Predictability (vs. randomness)

Programmability (vs. holism)

Coverage/precision trade-off: we cannot deal with millions of concepts





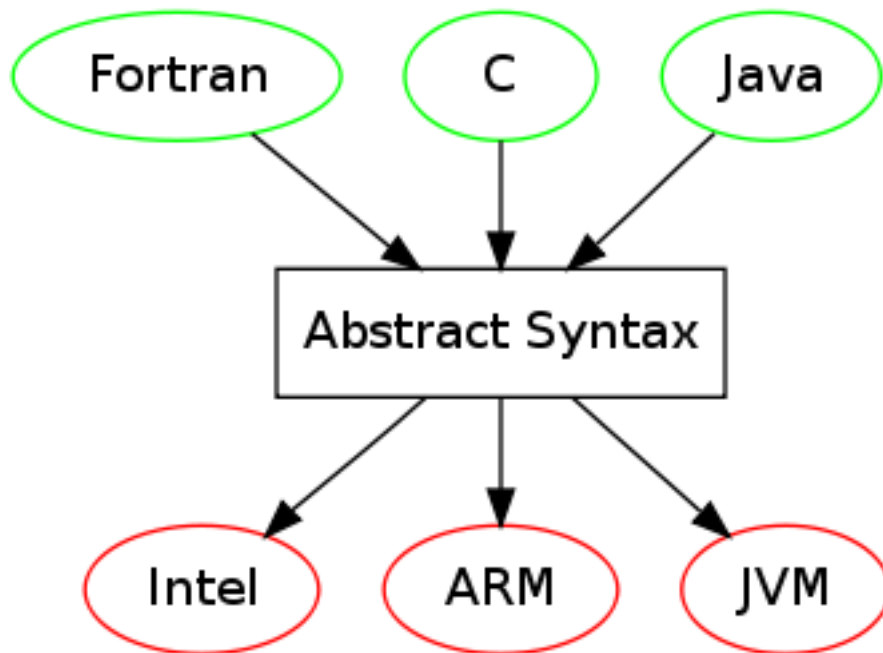
## **Main technologies**

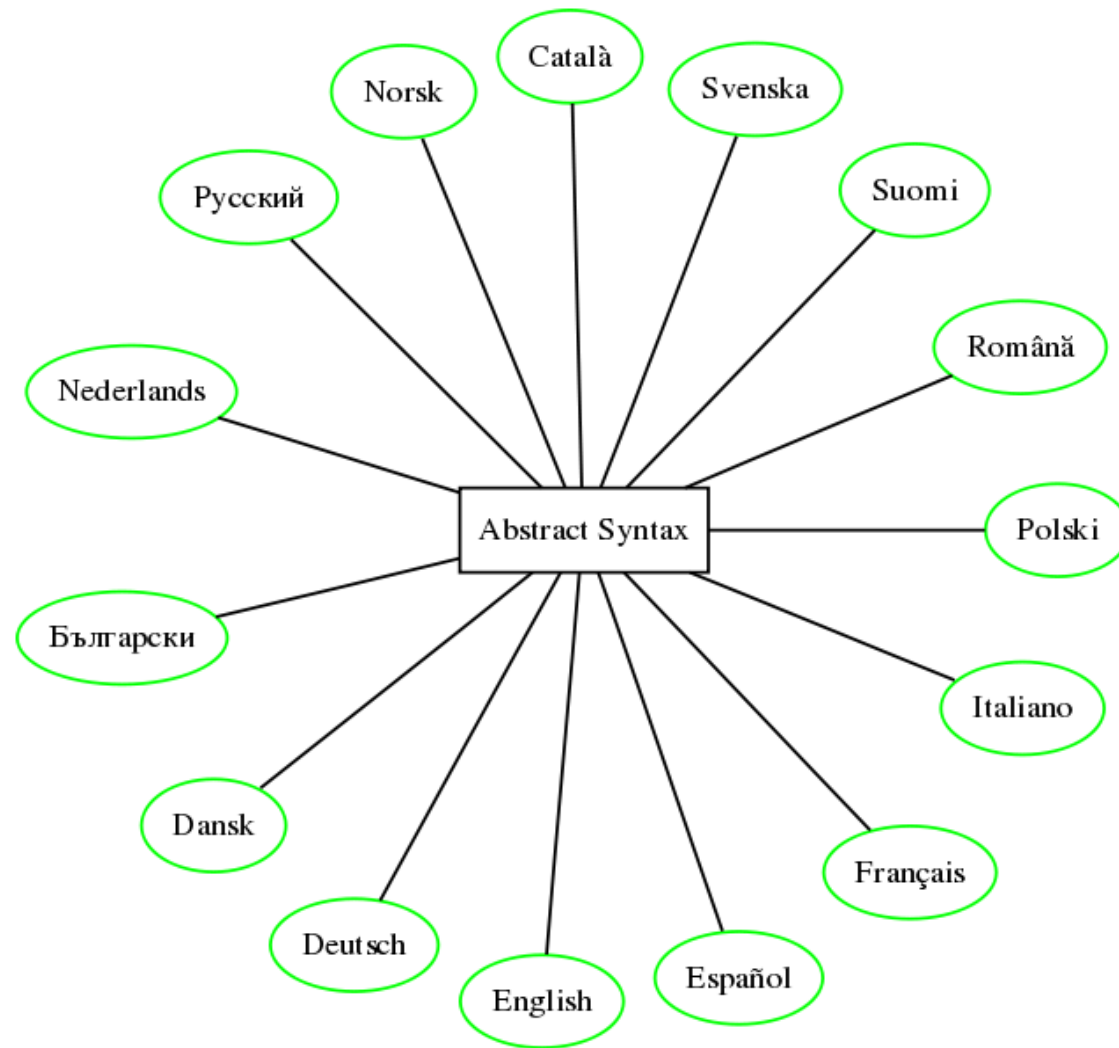
GF = Grammatical Framework

OWL Ontologies

Statistical Machine Translation

## The GF model: multi-source multi-target compilers





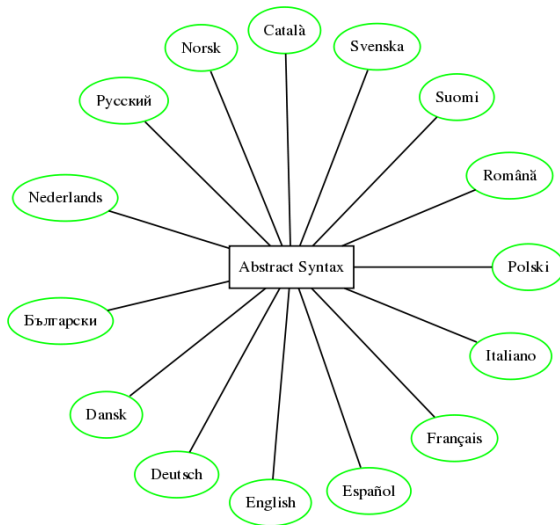
**MOLTO languages**

# The multilingual document

**Master document:** semantic representation (abstract syntax)

**Updates:** from any language that has a concrete syntax

**Rendering:** to all languages that have a concrete syntax



## Two things we do better than before

**No** universal interlingua:

- a framework for domain-specific interlinguas: **type theory**

**Yes** universal concrete syntax:

- a general-purpose **Resource Grammar Library**
- no hand-crafted *ad hoc* grammars

## Example: social network

Abstract syntax: **functions**,

```
fun Like : Person -> Item -> Fact
```

Concrete syntax: **linearizations**,

```
lin Like x y = x ++ "likes" ++ y      -- Eng  
lin Like x y = x ++ "tycker om" ++ y  -- Swe  
lin Like x y = y ++ "piace a" ++ x    -- Ita
```

## Complexity of concrete syntax

Italian: agreement, rection, clitics (*il vino piace a Maria* vs. *il vino mi piace* ; *tu mi piaci*)

```
lin Like x y = y.s ! nominative ++ case x.isPron of {  
  True  => x.s ! dative ++ piacere_V ! y.agr ;  
  False => piacere_V ! y.agr ++ "a" ++ x.s ! accusative  
}  
oper piacere_V = verbForms "piaccio" "piaci" "piace" ...
```

Moreover: contractions (*tu piaci ai bambini*), tenses, mood, ...

# The GF Resource Grammar Library

Currently for 24 languages; 3-6 months for a new language.

Complete morphology, comprehensive syntax, some lexicon

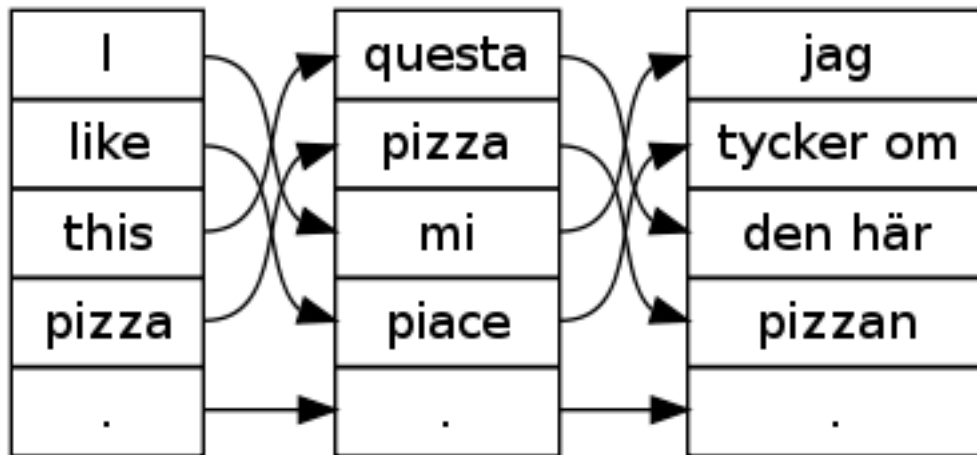
Common syntax API:

```
lin Like x y = mkC1 x (mkV2 (mkV "like")) y      -- Eng
lin Like x y = mkC1 x (mkV2 (mkV "tycker") "om") y -- Swe
lin Like x y = mkC1 y (mkV2 piacere_V dative) x   -- Ita
```



mkC1	<u>NP</u> -> <u>V2</u> -> <u>NP</u> -> <u>Cl</u>	<i>she loves him</i>	
mkC1	<u>NP</u> -> <u>V3</u> -> <u>NP</u> -> <u>NP</u> -> <u>Cl</u>	<i>she sends it to him</i>	
mkC1	<u>NP</u> -> <u>VV</u> -> <u>VP</u> -> <u>Cl</u>	<i>she wants to sleep</i>	
mkC1	<u>NP</u> -> <u>VS</u> -> <u>S</u> -> <u>Cl</u>	<i>she says</i>	<ul style="list-style-type: none"> <li>• API: mkC1 she_NP want_VV (mkVP sleep_V)</li> <li>• Afr: sy wil te slaap</li> <li>• Bul: тя иска да спи</li> <li>• Cat: ella vol dormir</li> <li>• Dan: hun vil sove</li> <li>• Dut: ze wil slapen</li> <li>• Eng: she wants to sleep</li> <li>• Fin: hän tahtoo nukkua</li> <li>• Fre: elle veut dormir</li> <li>• Ger: sie will schlafen</li> <li>• Hin: वह सोना चाहती है</li> <li>• Ita: lei vuole dormire</li> <li>• Lav: viņa grib gulēt</li> <li>• Nep: उनी सुत्न चाहन्छिन्</li> <li>• Nor: hun vil sove</li> <li>• Pes: او می خواهد بخوابد</li> <li>• Pnb: او سونا چاندی اے</li> <li>• Pol: ona chce spać</li> <li>• Ron: ea vrea să doarmă</li> <li>• Rus: она хочет спать</li> <li>• Snd: هوع سمن چاهمی ای</li> <li>• Spa: ella quiere dormir</li> <li>• Swe: hon vill sova</li> <li>• Tha: หล่อนอยากนอนหลับ</li> <li>• Urd: وہ سونا چاہتی ہے</li> </ul>
mkC1	<u>NP</u> -> <u>VQ</u> -> <u>QS</u> -> <u>Cl</u>	<i>she works</i>	
mkC1	<u>NP</u> -> <u>VA</u> -> <u>A</u> -> <u>Cl</u>	<i>she becomes</i>	
mkC1	<u>NP</u> -> <u>VA</u> -> <u>AP</u> -> <u>Cl</u>	<i>she becomes</i>	
mkC1	<u>NP</u> -> <u>V2A</u> -> <u>NP</u> -> <u>A</u> -> <u>Cl</u>	<i>she paid</i>	
mkC1	<u>NP</u> -> <u>V2A</u> -> <u>NP</u> -> <u>AP</u> -> <u>Cl</u>	<i>she paid</i>	
mkC1	<u>NP</u> -> <u>V2S</u> -> <u>NP</u> -> <u>S</u> -> <u>Cl</u>	<i>she answers</i>	
mkC1	<u>NP</u> -> <u>V2Q</u> -> <u>NP</u> -> <u>QS</u> -> <u>Cl</u>	<i>she asks</i>	
mkC1	<u>NP</u> -> <u>V2V</u> -> <u>NP</u> -> <u>VP</u> -> <u>Cl</u>	<i>she begins</i>	
mkC1	<u>NP</u> -> <u>A</u> -> <u>Cl</u>	<i>she is</i>	
mkC1	<u>NP</u> -> <u>A</u> -> <u>NP</u> -> <u>Cl</u>	<i>she is</i>	
mkC1	<u>NP</u> -> <u>A2</u> -> <u>NP</u> -> <u>Cl</u>	<i>she is not</i>	
mkC1	<u>NP</u> -> <u>AP</u> -> <u>Cl</u>	<i>she is very</i>	
mkC1	<u>NP</u> -> <u>NP</u> -> <u>Cl</u>	<i>she is the</i>	
mkC1	<u>NP</u> -> <u>N</u> -> <u>Cl</u>	<i>she is a</i>	
mkC1	<u>NP</u> -> <u>CN</u> -> <u>Cl</u>	<i>she is a</i>	
mkC1	<u>NP</u> -> <u>Adv</u> -> <u>Cl</u>	<i>she is here</i>	
mkC1	<u>NP</u> -> <u>VP</u> -> <u>Cl</u>	<i>she always</i>	

## Word/phrase alignments via abstract syntax



# Controlled language

Almost what MOLTO is, except that we

- generalize this to **multilingual controlled language systems**
- support ambiguous language (and **disambiguation**)

Prime example: Attempto Controlled English (U Zurich)

- generalized to 5 languages in GF (CNL 2009)
- extended to 15 in MOLTO

**Work packages**

**WP1: management (UGOT)**

## WP2: grammar tools (UGOT)

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

**New:**

- IDE's: Eclipse (John Camilleri) and cloud-based (Thomas Hallgren)
- support for on-the-fly extension
- resource grammars: Hindi, Latvian, Nepali, Persian, Punjabi, Sindhi, Thai (Shafqat Virk & al., Normunds Gruzitis)

GF - Demo/HelloAbs.gf - Eclipse

File Edit Navigate Search Project Run Window Help

Project Explorer

- Demo
  - AbsCat.gf
  - HelloAbs.gf
  - HelloEng.gf
  - ResEng.gf

Outline

- HelloAbs
  - Extends
    - AbsCat
  - flags
    - startcat = Greeting
  - cat
    - Farewell
  - fun
    - Hello
    - Goodbye
    - World
    - Parent
    - Friends

\*AbsCat.gf

```
abstract AbsCat = {  
  cat Greet ; Recipient ;  
};
```

\*HelloAbs.gf

```
abstract HelloAbs = AbsCat ** {  
  flags startcat = Greeting ;  
  cat Farewell ;  
  fun  
    Hello : Recipient -> Greeting ;  
    Goodbye : Recipient -> Farewell ;  
    World, Parent, Friends : Recipient ;  
}
```

ResEng.gf

```
--# -path=.../abstract.../common.../prelude  
resource ResEng = {  
  param  
    Gender = Masc | Fem ;  
};
```

HelloEng.gf

```
concrete HelloEng of HelloAbs = open ResEng in {  
  lincat  
    Greeting, Farewell = {s : Str} ;  
    Recipient = {s : Gender => Str} ;  
  lin  
    Hello recip = {s = "hello" ++ recip.s ! ResEng.Masc} ;  
    Goodbye recip = {s = "goodbye" ++ recip.s ! Fem} ;  
    World = {s = \"_\" => "world"} ;  
    Parent = { s = table {  
      Masc => "dad" ; Fem => "mum"  
    } } ;  
    Friends = superate "friends" ;  
  oper  
    superate : Str -> Recipient = \s ->  
      lin Recipient { s = \"_\" => "super" ++ s } ;  
}
```

Console

GF Compiler

Writable Insert 3:32

Start category "Greeting" not found

## GF online editor for simple multilingual grammars

Hello%

Show plain

Upload

X

Abstract

× Finnish

× Romanian

× Swedish

× English

+

```
concrete HelloFin of Hello =  
  open  
  +  
  lincat  
    Greeting = Str%  
    Friend = Str%  
  lin  
    Hello friend = ("terve" | "hei" | "moro") ++ friend%  
    World = "maailma"%  
  param  
  +  
  oper  
  +
```

☒ Enable editing on touch devices. +=Add an item, ×=Delete item, %=Edit item.

[About](#)



## WP3: translator's tools (UHEL)

### Transparent use

- text input + prediction, syntax editing
- disambiguation
- on the fly extension

### New:

- terminology tools (Lauri Carlson, Inari Listenmaa, Seppo Nyrkkö)
- translator user interface (Lauri Carlson, Inari Listenmaa)
- fast large-scale parsing: a C runtime for GF (Lauri Alanko, Krasimir Angelov)

## The term page

terms

KeyWords: Mayonnaise

☐ case sensitivity

query

class	uri
French cuisine	http://dbpedia.org/resource/Category:French_cuisine
Spanish cuisine	http://dbpedia.org/resource/Category:Spanish_cuisine
Sauces	http://dbpedia.org/resource/Category:Sauces
Condiments	http://dbpedia.org/resource/Category:Condiments
Sauces of the m...	http://dbpedia.org/resource/Category:Sauces_of_the_mayonnaise_fa...
French loanwords	http://dbpedia.org/resource/Category:French_loanwords
2000s music gro...	http://dbpedia.org/resource/Category:2000s_music_groups
Philippine rock m...	http://dbpedia.org/resource/Category:Philippine_rock_music_groups

items

Add Record

Delete Record

name	uri	en	de
Agrodolce	http://dbpedia.org/resource...	Agrodolce	
Aioli	http://dbpedia.org/resource...		Aioli
Aji (food)	http://dbpedia.org/resource...	Aji (food)	
Ajvar	http://dbpedia.org/resource...		Ajvar
Albert sauce	http://dbpedia.org/resource...	Albert sauce	
Allemande sauce	http://dbpedia.org/resource...	Allemande sauce Parisienne sauce	
Apple sauce	http://dbpedia.org/resource...		Apfelmus
Avgolemono	http://dbpedia.org/resource...	Avgolemono	
Babi panggang sauce	http://dbpedia.org/resource...	Babi panggang sauce	
Beurre blanc	http://dbpedia.org/resource...		Beurre blanc
Bow Wow Sauce	http://dbpedia.org/resource...	Bow Wow Sauce	
Brandy butter	http://dbpedia.org/resource...		Brandy Butter
Bread sauce	http://dbpedia.org/resource...	Bread sauce Bread sauces	
Café de Paris sauce	http://dbpedia.org/resource...		Café de Paris

Undo

Submit

Page 1 of 5

Displaying 1 - 20 of 85

## **WP4: knowledge engineering (Ontotext)**

Grammar + ontology

- OWL interoperability
- transform web ontologies to interlinguas
- natural language search and inference

**New:**

- natural language queries (Milen Chechev, Borislav Popov)
- ontology verbalization (Milen Chechev)



An application for viewing datasets of the project [MOLTO](#)



MOLTO is funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement FP7-ICT-247914.

## Search

Results for "**Näytä kaikki organisaatiot ja näiden sijainnit**" (100 of 14857) (Sparql: [construct WHERE {...}](#))

Abu Dhabi National Oil Company is a company.

YPF, S\_A\_ is located In Argentine Republic.

YPF, S\_A\_ is a public company.

AAPT Limited is a company.

Tattersall's Holding Pty\_ Ltd\_ is a company.

James Hardie Industries N\_V\_ is located In Commonwealth of Australia.

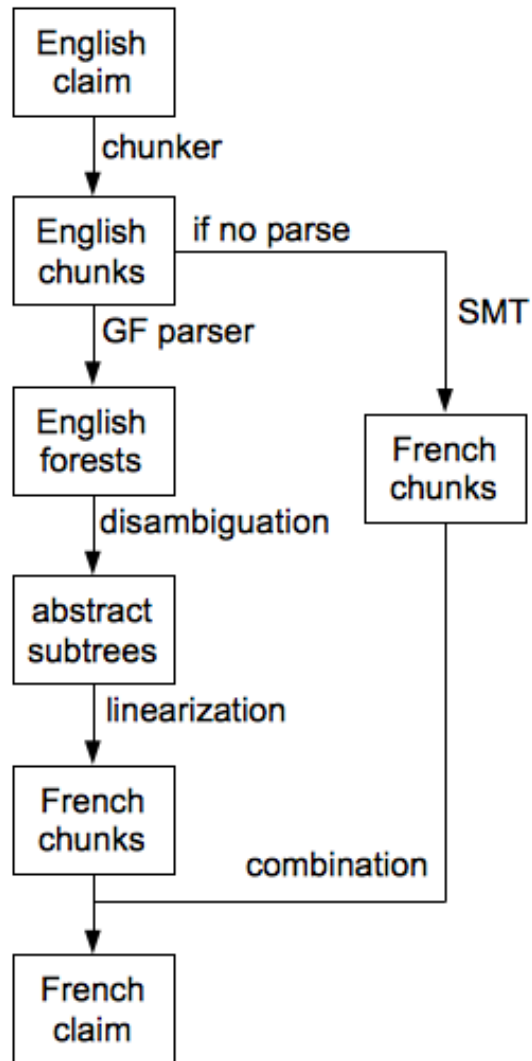
## **WP5: statistical and robust translation (UPC)**

### Hybrid systems

- statistical Machine Translation (SMT) as fall-back
- hard/soft integration
- learning of GF grammars by statistics
- improving SMT by grammars

### **New:**

- hybrid architecture with soft/hard integration
- robust parsing in GF



## **WP6: case study: mathematics (UPC)**

Multilingual rendering and translation of teaching material

- grammar and lexicon for the OpenMath standard, 12 languages
- high school and undergraduate level

**New:**

- a dialogue system for computer algebra Sage (Jordi Saludes)

**Abstract:**  $\text{if } x \text{ implies } (\text{mkProp } (\text{gt\_num } (\text{Var2Num } x) (\text{Var2Num } y))) (\text{mkProp } (\text{neq\_num } (\text{Var2Num } y) (\text{Var2Num } x)))$

**Bul:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Cat:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Eng:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Fin:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Fre:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Ger:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Ita:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**LaTeX:**  $(x > y) \Rightarrow (y \neq x)$

**Pol:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Ron:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Rus:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Spa:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Swe:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

**Urd:**  $\text{if } x \text{ is greater than } y \text{ then } y \text{ is not equal to } x$

$$(x > y) \Rightarrow (y \neq x)$$



## Untitled

last edited on February 13, 2012 12:52 PM by admin

[Save](#) [Save & quit](#) [Discard & quit](#)

File... [Action](#) [Data...](#) [sage](#) ☐ Typeset

[Print](#) [Worksheet](#) [Edit](#) [Text](#) [Undo](#) [Share](#) [Publish](#)

```
factor(1001)
```

```
7 * 11 * 13
```

```
%english
compute the summation of 1 over the factorial of x when x ranges from 0 to 6.
approximate it
```

```
1957/720
2.7180555555555554
```

```
%english
compute the absolute value of the difference of e and it.
approximate it
```

```
abs(e - 2.7180555555555554)
0.0002262729034896438
```

[evaluate](#)

## **WP7: case study: patents (UPC)**

Translation of pharmaceutical patents

- English, French, German
- SMT-GF hybrid

**New:**

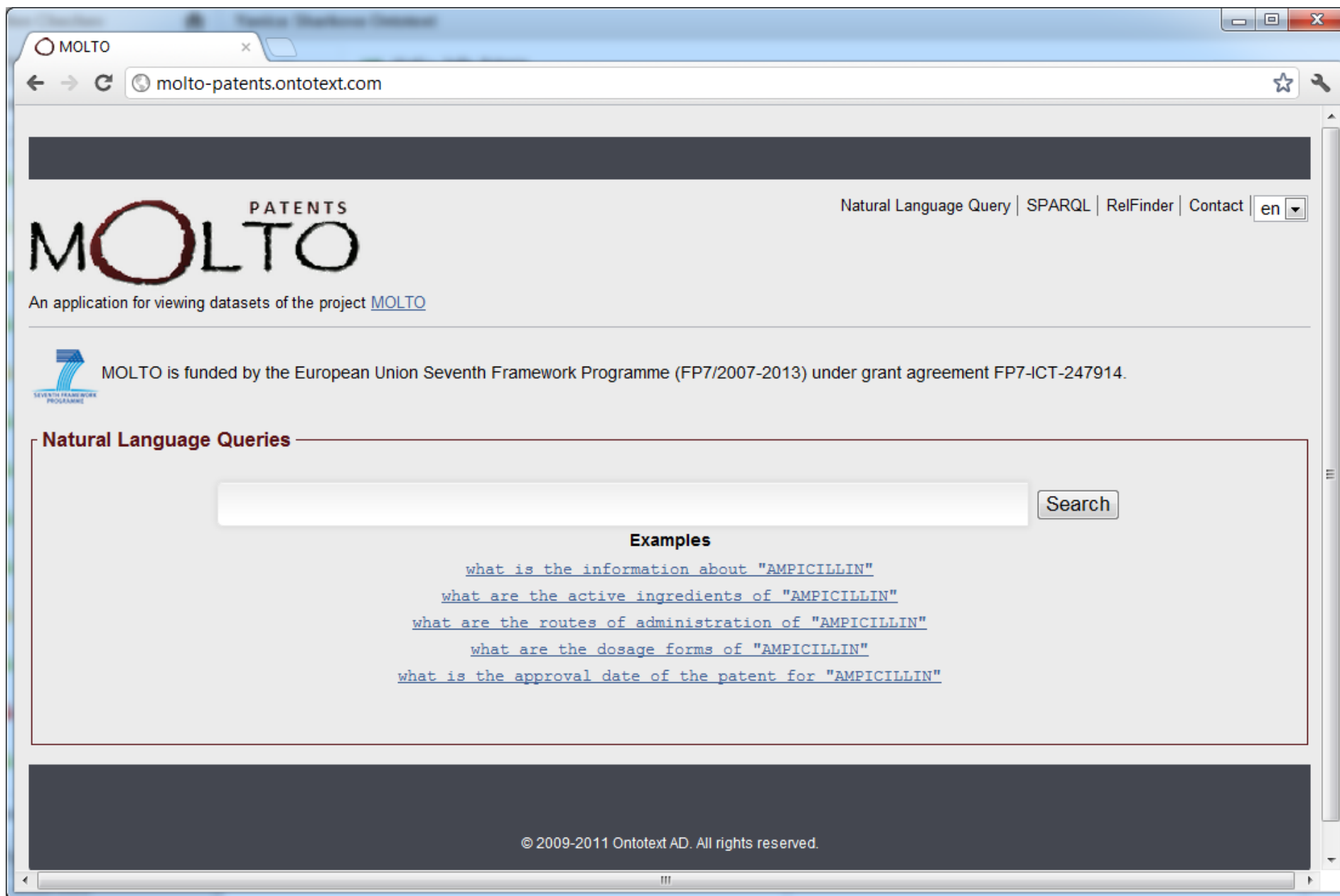
- SMT baseline + GF improvements (Cristina España, Lluís Màrquez, Ramona Enache)
- natural-language information retrieval from patents (Meritxell Gonzalez, Milen Chechev)

	WER	PER	TER	BLEU	NIST	GTM-2	MTR-pa	RG-S*	ULC
GF	60.96	50.08	58.90	26.56	5.57	22.74	38.76	29.00	16.17
SMT	27.03	17.50	25.32	63.18	9.99	44.58	71.64	72.65	67.14
HI	33.56	21.95	31.24	55.88	9.24	38.81	67.30	67.80	58.84
SI1.0	26.76	17.39	25.10	63.56	10.02	<b>44.86</b>	<b>71.96</b>	72.89	67.56
SI0.5	<b>26.63</b>	<b>17.32</b>	<b>25.02</b>	<b>63.60</b>	<b>10.03</b>	44.84	71.94	<b>72.93</b>	<b>67.60</b>
SI0.0	27.08	17.48	25.36	63.15	9.99	44.54	71.60	72.66	67.11

Table 3: Automatic evaluation of the baselines and hybrid systems.

<b>GF</b>	Une utilisation selon la revendication 3, dans laquelle le médicament séparé est administré at the same time as...
<b>SMT</b>	Utilisation selon la revendication 3, dans laquelle le médicament séparée est administré en même temps que...
<b>HI</b>	Une utilisation selon la revendication 3, dans laquelle le médicament séparé est administré en même temps que...
<b>SI0.5</b>	Utilisation selon la revendication 3, dans laquelle le médicament séparé est administré en même temps que...
<b>Ref.</b>	Utilisation selon la revendication 3, dans laquelle le médicament <b>séparé</b> est administré <b>en même temps</b> que...

Figure 2: Example where GF translates with the correct gender of the adjective and the SMT completes the untraslated words.



[Natural Language Query](#) | [SPARQL](#) | [RelFinder](#) | [Contact](#) | en ▾

An application for viewing datasets of the project [MOLTO](#)



MOLTO is funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement FP7-ICT-247914.

### Natural Language Queries

Search

#### Examples

[what is the information about "AMPICILLIN"](#)

[what are the active ingredients of "AMPICILLIN"](#)

[what are the routes of administration of "AMPICILLIN"](#)

[what are the dosage forms of "AMPICILLIN"](#)

[what is the approval date of the patent for "AMPICILLIN"](#)

## **WP8: case study: cultural heritage (UGOT)**

Translation of museum object descriptions

- based on CRM ontology
- applied to Gothenburg City Museum collections

**New:**

- prototype with natural language generation (Dana Dannélls, Ramona Enache, Milen Chechev)

## **WP9: user requirements and evaluation (UHEL)**

How good is MOLTO translation?

- comparative evaluations
- development of metrics

**New:**

- hybrid evaluation (Cristina España, Lluís Màrquez)
- software testing methods (QuickCheck) applied to grammars (Ramona Enache, Koen Claessen)

## **WP10: dissemination and exploitation (UGOT)**

Guide new users to MOLTO tools, find new applications, create a network.

### **New:**

- FreeRBMT12 in Gothenburg, 13-15 June (submission deadline 7 April)

## **WP11: multilingual semantic wiki (UZH)**

The ultimate user interface

- combine translation and grammar extension
- reasoning based on abstract syntax

**New:**

- ACE-Wiki ported to GF (Kaarel Kaljurand, Tobias Kuhn, Norbert Fuchs)



[Main Page](#)

- The image shows a 'Sentence Editor' dialog box. The title bar is 'Sentence Editor'. The main text area contains the sentence 'compute the imaginary part of the derivative of the ...'. Below the text area is a list of words for selection, including 'arccosecant', 'arccosine', 'arccotangent', 'arcsecant', 'arcsine', 'arctangent', 'cosecant', 'cosine', 'cotangent', 'derivative', 'divergence', 'eighth', 'exponential', 'fifth', 'first', and 'fourth'. The word 'derivative' is currently selected. An 'OK' button is visible in the bottom right corner.

Sentence Editor

compute the imaginary part of the derivative of the ...

< Delete

text

word

- arcsecant
- arccosine
- arccotangent
- arcsecant
- arcsine
- arctangent
- cosecant
- cosine
- cotangent
- derivative
- divergence
- eighth
- exponential
- fifth
- first
- fourth

OK Cancel

## **WP12: interactive knowledge-based systems (BI)**

Multilingua questionnaires and decision making

- user input + reasoning
- explanations generated in the users' languages

**New:**

- a new category of grammarians: software engineers with minimal GF training

## Availability of MOLTO tools

Open source, LGPL (*except* parts of the patent case study, parts of Be Informed applications)

Web demos

Mobile applications (Android)

# Conclusion

You shouldn't expect

- general-purpose translation ("Google competitor")

You should expect

- high quality multilingual translation
- portability to new domains (up to 1000's of words)
- productivity (days, weeks, months)
- ease of use (no training for authoring, a few days for grammarians)