Two Years of MOLTO

Aarne Ranta

MOLTO Meeting, Zurich, 7 March 2012
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.
What’s different?

<table>
<thead>
<tr>
<th>Tool</th>
<th>Google, Bing, Babelfish</th>
<th>MOLTO</th>
</tr>
</thead>
<tbody>
<tr>
<td>target</td>
<td>consumers</td>
<td>producers</td>
</tr>
<tr>
<td>input</td>
<td>unpredictable</td>
<td>predictable</td>
</tr>
<tr>
<td>coverage</td>
<td>unlimited</td>
<td>limited</td>
</tr>
<tr>
<td>quality</td>
<td>browsing</td>
<td>publishing</td>
</tr>
</tbody>
</table>
Producer’s quality

Responsibility for the translation

Cannot afford translating French

• *prix 99 euros*

• *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, yöitä, yönä, yöksi, yössä, yöstä, yöhön, yöllä, yöltä, yölle, yöttä, öineen, öin, yö, öitä, öiden, öinä, öiksi, öissä, öistä, öihin, öillä, öiltä, öille, öittä, öin

English: 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 = mkCl x (mkV2 (mkV "like")) y  -- Eng
lin Like x y = mkCl x (mkV2 (mkV "tycker") "om") y  -- Swe
lin Like x y = mkCl y (mkV2 piacere_V dative) x  -- Ita
```
<table>
<thead>
<tr>
<th>mKCl</th>
<th>NP -&gt; V2 -&gt; NP -&gt; Cl</th>
<th>she loves him</th>
</tr>
</thead>
<tbody>
<tr>
<td>mKCl</td>
<td>NP -&gt; V3 -&gt; NP -&gt; NP -&gt; Cl</td>
<td>she sends it to him</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; VV -&gt; VP -&gt; Cl</td>
<td>she wants to sleep</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; VS -&gt; S -&gt; Cl</td>
<td>she says</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; VQ -&gt; QS -&gt; Cl</td>
<td>she would</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; VA -&gt; A -&gt; Cl</td>
<td>she be</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; VA -&gt; AP -&gt; Cl</td>
<td>she became</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; V2A -&gt; NP -&gt; A -&gt; Cl</td>
<td>she paid</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; V2A -&gt; NP -&gt; AP -&gt; Cl</td>
<td>she paid</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; V2S -&gt; NP -&gt; S -&gt; Cl</td>
<td>she asks</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; V2Q -&gt; NP -&gt; QS -&gt; Cl</td>
<td>she asks</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; V2V -&gt; NP -&gt; VP -&gt; Cl</td>
<td>she begins</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; A -&gt; Cl</td>
<td>she is a</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; A -&gt; NP -&gt; Cl</td>
<td>she is a</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; A2 -&gt; NP -&gt; Cl</td>
<td>she is not a</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; AP -&gt; Cl</td>
<td>she is very</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; NP -&gt; Cl</td>
<td>she is the</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; N -&gt; Cl</td>
<td>she is an</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; CN -&gt; Cl</td>
<td>she is a</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; Adv -&gt; Cl</td>
<td>she is how</td>
</tr>
<tr>
<td>mKCl</td>
<td>NP -&gt; VP -&gt; Cl</td>
<td>she always</td>
</tr>
</tbody>
</table>

- API: mKCl she_NP want_VV (mKVP sleep_V)
- Afr: sy wil te slaap
- Bul: мa иска да сну
- Cat: ella vol dormir
- Dan: hun vil sove
- Dut: ze wil slapen
- Eng: she wants to sleep
- Fin: hän tahtoo nukkua
- Fre: elle veut dormir
- Ger: sie will schlafen
- Hin: वह सोना चाहती है
- Ita: lei vuole dormire
- Lav: viņa grib gultē
- Nep: उनी तुला चाहिन्छन्
- Nor: hun vil sove
- Pcs: او می خواهد بخوابد
- Pnb: اوساںا چاندی اے
- Pol: ona chce spać
- Ron: ea vrea să doară
- Rus: ona хочет спать
- Snd: हौ सेमा चाहिए
- Spa: ella quiere dormir
- Swe: hon vill sova
- Tha: หลับอย่างสงบ
- Urdu: وہ سونا حاشیہ ہے
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 online editor for simple multilingual grammars**

**Hello**

<table>
<thead>
<tr>
<th>Abstract</th>
<th>× Finnish</th>
<th>× Romanian</th>
<th>× Swedish</th>
<th>× English</th>
</tr>
</thead>
</table>

**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.
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)
<table>
<thead>
<tr>
<th>Class</th>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>French cuisine</td>
<td><a href="https://dbpedia.org/resource/Category:French_cuisine">https://dbpedia.org/resource/Category:French_cuisine</a></td>
</tr>
<tr>
<td>Spanish cuisine</td>
<td><a href="https://dbpedia.org/resource/Category:Spanish_cuisine">https://dbpedia.org/resource/Category:Spanish_cuisine</a></td>
</tr>
<tr>
<td>Sauces</td>
<td><a href="https://dbpedia.org/resource/Category:Sauces">https://dbpedia.org/resource/Category:Sauces</a></td>
</tr>
<tr>
<td>Condiments</td>
<td><a href="https://dbpedia.org/resource/Category:Condiments">https://dbpedia.org/resource/Category:Condiments</a></td>
</tr>
<tr>
<td>Sauces of the mayonnaise</td>
<td><a href="https://dbpedia.org/resource/Category:Sauces_of_the_mayonnaise">https://dbpedia.org/resource/Category:Sauces_of_the_mayonnaise</a></td>
</tr>
<tr>
<td>French loanwords</td>
<td><a href="https://dbpedia.org/resource/Category:French_loanwords">https://dbpedia.org/resource/Category:French_loanwords</a></td>
</tr>
<tr>
<td>2000s music groups</td>
<td><a href="https://dbpedia.org/resource/Category:2000s_music_groups">https://dbpedia.org/resource/Category:2000s_music_groups</a></td>
</tr>
<tr>
<td>Philippine rock music</td>
<td><a href="https://dbpedia.org/resource/Category:Philippine_rock_music">https://dbpedia.org/resource/Category:Philippine_rock_music</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Terms</th>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agrodolce</td>
<td><a href="http://dbpedia.org/resource:Agrodolce">http://dbpedia.org/resource:Agrodolce</a></td>
</tr>
<tr>
<td>Ají</td>
<td><a href="http://dbpedia.org/resource:Aj%C3%AD">http://dbpedia.org/resource:Ají</a></td>
</tr>
<tr>
<td>Ají (food)</td>
<td><a href="http://dbpedia.org/resource:Aj%C3%AD">http://dbpedia.org/resource:Ají</a> (food)</td>
</tr>
<tr>
<td>Ají (food)</td>
<td><a href="http://dbpedia.org/resource:Aj%C3%AD">http://dbpedia.org/resource:Ají</a> (food)</td>
</tr>
<tr>
<td>Ají</td>
<td><a href="http://dbpedia.org/resource:Aj%C3%AD">http://dbpedia.org/resource:Ají</a></td>
</tr>
<tr>
<td>Albert sauce</td>
<td><a href="http://dbpedia.org/resource:Albert_sauce">http://dbpedia.org/resource:Albert_sauce</a></td>
</tr>
<tr>
<td>Allemande sauce</td>
<td><a href="http://dbpedia.org/resource:Allemande_sauce">http://dbpedia.org/resource:Allemande_sauce</a></td>
</tr>
<tr>
<td>Allemande sauce</td>
<td><a href="http://dbpedia.org/resource:Allemande_sauce">http://dbpedia.org/resource:Allemande_sauce</a></td>
</tr>
<tr>
<td>Panzére sauce</td>
<td><a href="http://dbpedia.org/resource:Panz%C3%A9re_sauce">http://dbpedia.org/resource:Panzére_sauce</a></td>
</tr>
<tr>
<td>Apple sauce</td>
<td><a href="http://dbpedia.org/resource:Apple_sauce">http://dbpedia.org/resource:Apple_sauce</a></td>
</tr>
<tr>
<td>Avgolemono</td>
<td><a href="http://dbpedia.org/resource:Avgolemono">http://dbpedia.org/resource:Avgolemono</a></td>
</tr>
<tr>
<td>Babí panagpang sauce</td>
<td><a href="http://dbpedia.org/resource:Bab%C3%AD_panagpang_sauce">http://dbpedia.org/resource:Babí_panagpang_sauce</a></td>
</tr>
<tr>
<td>Beurre blanc</td>
<td><a href="http://dbpedia.org/resource:Beurre_blanc">http://dbpedia.org/resource:Beurre_blanc</a></td>
</tr>
<tr>
<td>Bow Wow Sauce</td>
<td><a href="http://dbpedia.org/resource:Bow_Wow_Sauce">http://dbpedia.org/resource:Bow_Wow_Sauce</a></td>
</tr>
<tr>
<td>Brandy butter</td>
<td><a href="http://dbpedia.org/resource:Brandy_butter">http://dbpedia.org/resource:Brandy_butter</a></td>
</tr>
<tr>
<td>Bread sauce</td>
<td><a href="http://dbpedia.org/resource:Bread_sauce">http://dbpedia.org/resource:Bread_sauce</a></td>
</tr>
<tr>
<td>Bread sauce</td>
<td><a href="http://dbpedia.org/resource:Bread_sauces">http://dbpedia.org/resource:Bread_sauces</a></td>
</tr>
<tr>
<td>Café de Paris sauce</td>
<td><a href="http://dbpedia.org/resource:Caf%C3%A9_de_Paris">http://dbpedia.org/resource:Café_de_Paris</a></td>
</tr>
</tbody>
</table>
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

Näytä kaikki organisaatiot ja näiden sijainnit

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{\text{\(\mathcal{A}\)}} \text{ implies } (\text{mkProp (gt_num (Var2Num x) (Var2Num y))) (mkProp (neq_num (Var2Num y) (Var2Num x))))} \)

Bul:  
\( \text{аako x e po - golyamo ot y tohava y ne e равно до x} \)

Cat:  
\( \text{si x es més gran que y llavors y no és igual a x} \)

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

Fin:  
\( \text{jos x on suurempi kuin y niin y ei ole yhtäsuuri kuin x} \)

Fre:  
\( \text{si x est plus grand qu' y alors y n' est pas égal à x} \)

Ger:  
\( \text{wenn x größer als y ist dann ist y nicht gleich x} \)

Ita:  
\( \text{se x è maggiore di y quindi y non è uguale a x} \)

LaTeX:  
\( \text{\(\mathcal{A}\) ( x > y ) \Rightarrow ( y \neq x )} \)

Pol:  
\( \text{jesli x jest wieksze ni\ż y to y nie jest równe x} \)

Ron:  
\( \text{dacă x este mai mare decât y atunci y nu este egal la x} \)

Rus:  
\( \text{если x большее у то у не равно x} \)

Spa:  
\( \text{si x es mayor que y entonces y no es igual a x} \)

Swe:  
\( \text{om x är större än y så är y inte lika med x} \)

Urd:  
\( \text{کی برابر نہیں ہے سب توہم ہے تب اگر} \)

\[ (x > y) \Rightarrow (y \neq x) \]
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
2.718055555555555

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

abs(e - 2.718055555555555)
0.0002262729034896438
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)
<table>
<thead>
<tr>
<th></th>
<th>WER</th>
<th>PER</th>
<th>TER</th>
<th>BLEU</th>
<th>NIST</th>
<th>GTM-2</th>
<th>MTR-pa</th>
<th>RG-S*</th>
<th>ULC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GF</td>
<td>60.96</td>
<td>50.08</td>
<td>58.90</td>
<td>26.56</td>
<td>5.57</td>
<td>22.74</td>
<td>38.76</td>
<td>29.00</td>
<td>16.17</td>
</tr>
<tr>
<td>SMT</td>
<td>27.03</td>
<td>17.50</td>
<td>25.32</td>
<td>63.18</td>
<td>9.99</td>
<td>44.58</td>
<td>71.64</td>
<td>72.65</td>
<td>67.14</td>
</tr>
<tr>
<td>HI</td>
<td>33.56</td>
<td>21.95</td>
<td>31.24</td>
<td>55.88</td>
<td>9.24</td>
<td>38.81</td>
<td>67.30</td>
<td>67.80</td>
<td>58.84</td>
</tr>
<tr>
<td>SI1.0</td>
<td>26.76</td>
<td>17.39</td>
<td>25.10</td>
<td>63.56</td>
<td>10.02</td>
<td>44.86</td>
<td>71.96</td>
<td>72.89</td>
<td>67.56</td>
</tr>
<tr>
<td>SI0.5</td>
<td>26.63</td>
<td>17.32</td>
<td>25.02</td>
<td>63.60</td>
<td>10.03</td>
<td>44.84</td>
<td>71.94</td>
<td>72.93</td>
<td>67.60</td>
</tr>
<tr>
<td>SI0.0</td>
<td>27.08</td>
<td>17.48</td>
<td>25.36</td>
<td>63.15</td>
<td>9.99</td>
<td>44.54</td>
<td>71.60</td>
<td>72.66</td>
<td>67.11</td>
</tr>
</tbody>
</table>

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

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ref.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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 (Ramonà 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)
approximate the truncation of the remainder of the minimal element of the integer interval from 3.0 to 6.0 divided by 2.0.
compute the product of the octal number 12 and the binary number 100.
compute the greatest common divisor of x and the product of x and y.
compute the greatest common divisor of x and the sum of x and y.
compute the real part of the derivative of the exponential at pi.
compute the imaginary part of the derivative of the exponential at pi.
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)