



Published on *Multilingual Online Translation* (<http://www.molto-project.eu>)

1. Publishable summary

1.1 Project context and objectives

The project MOLTO - Multilingual Online Translation, started on March 1, 2010 and will run for 39 months. It promises 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 and statistical machine translation to simplify the production of multilingual documents without sacrificing the quality. The interlinguas are based on domain semantics and are equipped with reversible generation functions: namely translation is obtained as a composition of parsing the source language and generating the target language. An implementation of this technology is provided by GF, Grammatical Framework. GF technologies in MOLTO are complemented by the use of ontologies, such as those used in the semantic web, and by methods of statistical machine translation (SMT) for improving robustness and extracting grammars from data. GF has been applied in several small-to-medium size domains, typically targeting up to ten languages but MOLTO will scale this up in terms of productivity and applicability.

A part of the scale-up is to increase the size of domains and the number of languages. A more substantial part is to make the technology accessible to domain experts without GF expertise and minimize the effort needed for building a translator. Ideally, with the tools produced by MOLTO, this can be done by just extending a lexicon and writing a set of example sentences.

MOLTO is committed to dealing with 15 languages, which includes 12 official languages of the European Union - Bulgarian, Danish, Dutch, English, Finnish, French, German, Italian, Polish, Romanian, Spanish, and Swedish - and 3 other languages - Catalan, Norwegian, and Russian. In addition, there is on-going work on at least Arabic, Farsi, Hebrew, Hindi/Urdu, Icelandic, Japanese, Latvian, Maltese, Portuguese, Swahili, Tswana, and Turkish.

While tools like Systran (Babelfish) and Google Translate are designed for consumers of information, MOLTO will mainly target the producers of information. Hence, the quality of the MOLTO translations must be good enough for, say, an e-commerce site to use in translating their web pages automatically without the fear that the message will change. Third-party translation tools, possibly integrated in the browsers, let potential customers discover, in their preferred language, whether, for instance, an e-commerce page written

in French offers something of interest. Customers understand that these translations are approximate and will filter out imprecision. If, for instance, the system has translated a price of 100 Euros to 100 Swedish Crowns (which equals 10 Euros), they will not insist to buy the product for that price. But if a company had placed such a translation on its website, then it might be committed to it. There is a well-known trade-off in machine translation: one cannot at the same time reach full coverage and full precision. In this trade-off, Systran and Google have opted for coverage whereas MOLTO opts for precision in domains with a well-understood language.

The MOLTO Enlarged EU proposal adds two countries (Switzerland and The Netherlands) and two work packages. The Semantic Wiki work package builds a system that integrates the functionalities of MOLTO tools with a collaborative environment, where users can create content in different languages, and all edits become immediately visible in all languages, via automatic semantic-based translation. The Interactive Knowledge-Based System work package puts MOLTO technology to use in an enterprise environment, for the localization of end-user oriented systems to new languages and the generation of high-quality explanations in natural language. Noteworthy in this work package is the fact that translation grammars are constructed in house by Be Informed's non-expert staff without the intervention of grammar specialists.

MOLTO technology will be released as open-source libraries, which can be plugged into standard translation tools and web pages and thereby fit into standard workflows. It will be demonstrated in web-based demos and applied in three case studies: mathematical exercises in 15 languages, patent data in at least 3 languages, and museum object descriptions in 15 languages.

1.2 Main results achieved so far

The results achieved during the first 24 months of the projects have been demonstrated during the 4th Project Meeting. They include:

- GF plugin for the Eclipse IDE
- patent retrieval in NL in the pharmaceutical domain
- first results on hybrid models combining GF and SMT
- SageGF, a natural language input for the suite of computer algebra systems Sage with output rendered aurally
- GF grammars for the cultural heritage domain
- a number of translation tools, including the *Global sight* translation system, and the MOLTO editors for ontologies, translations and equivalence in TermFactory
- preliminary integration of GF-based multilingual editing in AceWiki.

A detailed list with short abstracts is available at <http://www.molto-project.eu/content/molto-4th-project-meeting-demos> [1].

In the past semester we reported:

- the first web services demonstrations highlighting the fundamental features of the MOLTO system tools: high-quality multilinguality and NLG multilingual query interface to semantic data;
- the mathematical grammar library that allows to express simple mathematical problems in more than 10 languages, although not all at the same level of quality.
- a online GF grammar editor for assisting authors of GF application grammars when editing in the cloud, also including the option to carry out example-based grammar generation;
- GF plugins in Python and C to call library functions;
- APIs for GF Grammar compilation and for the MOLTO translator's tools;
- a prototype demonstrating GF grammar-ontology interoperability;
- a description of the cultural heritage ontology for museum artifacts and use case scenarios for the application grammar.

1.3 Expected final results and their potential impact and use

The expected final product of MOLTO is a software toolkit made available via the MOLTO website. It will consist in a family of open-source software flagship products:

- Cloud-IDE for GF
- Eclipse-IDE for GF
- a professional translator tool with integrated terminology extension, predictive parsing, syntax editing, robust parsing
- an ontology query system, based on a generic reusable grammar (carefully built manually) with automatically derived domain-specific extensions; demoed with patents and museum
- a hybrid patent translation system that beats its competitors, if possible
- the Mathematical Grammar Library (MGL), supporting both math documentation (Wiki) and tutorial (Sage, word problems)
- a museum visitors' system generating descriptions and providing a query system, in 15 languages
- multilingual semantic wiki system integrating the technologies and resources developed within MOLTO
- a software product from Be Informed that is useful for their customers (to be specified)
- a set of evaluations, for each of the flagships listed before, showing that high quality can be reached with reasonable effort using MOLTO tools

These tools will be portable to different platforms as well as generally portable to new domains and languages. By the end of the project, MOLTO expects to have grammar resource libraries for 18 languages, whereas MOLTO use cases will target between 3 and 15 languages.

The main societal impact of MOLTO will be on contributing to a new perception for the

possibilities of machine translation, moving away from the idea that domain-specific high-quality translation is expensive and cumbersome. MOLTO tools will change this view by radically lowering the effort needed to provide high-quality scoped translation for applications where the content has enough semantic structure.

Source URL: <http://www.molto-project.eu/wiki/living-deliverables/d15-periodic-management-report-t24/1-publishable-summary>

Links:

[1] <http://www.molto-project.eu/content/molto-4th-project-meeting-demos>