GF - Grammatical Framework

The core of a MOLTO translation system is a multilingual GF grammar, where meaning-preserving translation is automatically provided as a composition of parsing and generation via the abstract syntax, which works as an interlingua. More precisely, GF implements a framework for interlinguas and additionally provides RGL, the GF Resource Grammar Library, implementing the basic linguistic details of languages: inflectional morphology and syntactic combination functions. MOLTO will further improve grammar engineering in GF by:

- **Integrated Development Environment** to use the RGL and to manage large projects;
- **example-based grammar writing support** to bootstrap a grammar from a set of example translations.

Statistical Machine Translation

MOLTO will develop and evaluate combination approaches to integrate grammar-based and SMT models in a hybrid MT system. At least four variants will be studied:

- **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.

OWL Ontologies

MOLTO sees ontologies as a way to formalize interlinguas in specific domains. Based on this observation, it will carry out research to develop two-way grammar-ontology interoperability that will bridge natural language and formal knowledge. The resulting MOLTO infrastructure will allow knowledge modeling, semantic indexing and retrieval using natural language. The engine will perform semi-automatic creation of abstract grammars from ontologies; derive ontologies from grammars, and retrieve instance level knowledge from/in natural language by first transforming queries to semantic queries, and secondly by expressing the resulting knowledge in natural language.

MOLTO's goal is to develop tools for web content providers to translate 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.

MOLTO

non multa, sed multum

Multilingual On-line Translation

MOLTO's tools

are meant for the producer who has a limited corpus with predictable coverage captured by a well defined, understood semantics. In exchange for the domain limitation, the translation will be high quality, with fixed word senses and proper idioms in several languages.

MOLTO case study domains

- **Mathematical exercises**: enhance the multilingual mathematical GF library by adding a grammar for commanding a Computer Algebra System by natural language imperative sentences. Using ontologies to describe word problems, the system will be able to carry out a dialog with the student solving the problem.
- **Biomedical and pharmaceutical patents**: create a commercially viable prototype of a system for multilingual translation and cross-language retrieval of patent abstracts and claims in at least 3 languages.
- **Museum object descriptions**: build an ontology-based multilingual grammar starting from a CRM ontology for artifacts at Gothenburg City Museum. The prototype will be tested for cross-language retrieval and representation, and for automatic generation of Wikipedia-like articles for museum artifacts in 5 languages.

Translator's Tools

The standard working method in current translation tools is to work on the source and translation as a bilingual text. Translation suggestions, sought from Translation Memory based on similarity, or generated by a Machine Translation system, are presented for the user to choose from and edit manually.

The MOLTO translator tool will add two constrained-language authoring modes: a robust statistical machine translation mode, plus vocabulary and grammar extension tools. Features of the tool include also:

- authoring with abstract-syntax driven context-sensitive word completion;
- syntax text editing by manipulating abstract syntax trees;
- back-up by robust and statistical translation for out-of-grammar input;
- support of on-the-fly extension by the translator using multilingual ontology-based lexicon builder;
- example-based grammar writing.

The MOLTO's API and the Web-based translator tool are designed as plug-ins to professional translation memory tools such as SDL and WordFast and implement a collaborative translation workflow, in particular for multilingual terminology.

Mathematics Domain

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
```

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

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
```

Abstract Syntax

GÖTEBORGS UNIVERSITET

HELSINGIN YLIOPISTO

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
UPC BARCELONATECH

ontotext
Semantic Technology Lab

MATRIXWARE
INFORMATION SERVICES

SEVENTH FRAMEWORK PROGRAMME

FP7- 247914 molto-project.eu