Publishable summary

The Monnet project is concerned mainly with ontology localisation, i.e., the translation of the lexico-terminological level of ontologies (often referred to as the 'ontology labels'). The project outcomes, as currently understood by the project members, can be described as a set of software components as follows, all of which can be used in combination as well as stand-alone:

– Ontology Lexicalization
– Ontology Localization
– Cross-lingual Ontology-based Information Extraction (CLOBIE)
– Cross-lingual Knowledge Access & Presentation

The following picture illustrates the envisioned connections between these components, with their inputs and outputs:

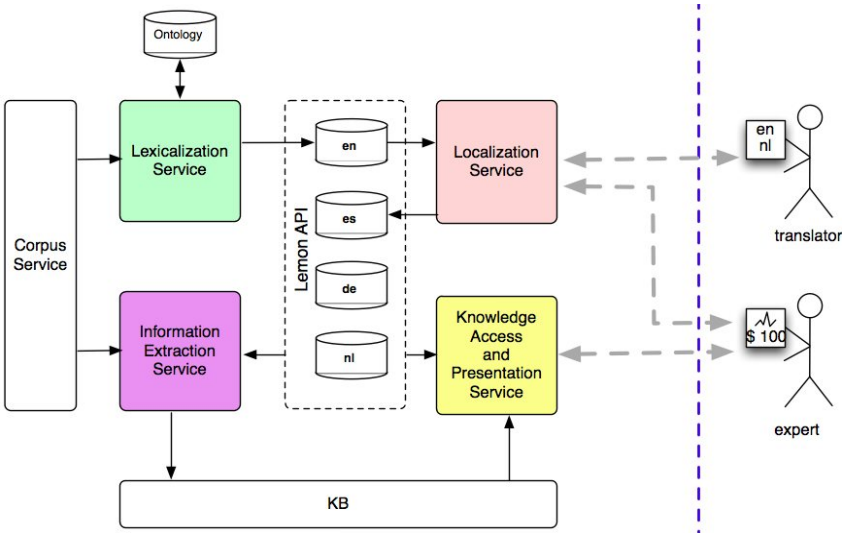

Figure 1: Monnet overview

The core objective of Monnet is the provision of advanced services for the translation of the lexico-terminological level of ontologies, which will be instantiated by the 'localization service'.

However, as ontologies often have only a very limited representation of lexico-terminological information, a first step will be to analyze a given ontology and enrich it with appropriate information on i) the terminological structure of ontology labels, ii) linguistic information on terminology items, and iii) analysis of implicit semantics where needed. Together we refer to these analysis and enrichment steps as "ontology lexicalisation", which will be instantiated by the 'lexicalisation service' that takes as input an ontology and outputs an 'ontology-lexicon' for at least one default language (depending on the language that was used in defining ontology labels). A 'corpus service' will enable access to external domain corpus evidence for modelling and analyzing language use in the ontology labels. The ontology-lexicon will be represented on the basis of the so-

called 'lemon' format[1], a lexicon model for ontologies that has been defined during this reporting period by the Monnet project for the appropriate integration of lexical/linguistic and terminological information in ontologies. The different lexicons will therefore be handled by use of the 'lemon API' as shown.

After the generation of an appropriate ontology-lexicon by the 'lexicalisation service', the 'localisation service' can now operate on this for the translation process, i.e., note that Monnet thus implements ontology localisation by translation of the ontology-lexicon. A hybrid machine translation (MT) approach will be explored that will build on state of the art statistical MT as provided by the open source Moses[2] system, and knowledge-based approaches using domain knowledge provided by ontologies and associated domain terminologies, exploited by the new version of LabelTranslator system implemented in Monnet. The ontology localisation service will be developed and tested in the context of two real-life use cases on the translation of i) financial services taxonomies in XBRL (eXtensible Business Reporting Language) developed by Monnet partner XBRL Europe, and ii) public services ontologies developed by Monnet partner Be Informed for the Dutch government.

To provide a further proof of concept in a real-life use case, the Monnet project implements also a cross-lingual ontology-based information extraction component that exploits ontology-lexicons for various languages for financial fact extraction with the XBRL-based xEBR vocabulary for European-wide harmonization of business register information. The 'information extraction service' instantiates this functionality by use of the lemon API (for accessing lemon-based ontology lexicons) and the corpus service (for accessing linguistically/semantically annotated domain corpora) to generate a knowledge base ('KB') of financial facts.

Finally, the 'knowledge access & presentation service' will instantiate a component for the multilingual querying and visualisation of this KB, which will provide a further proof of concept in the real-life use case of cross-lingual business intelligence. This component will be able to provide cross-lingual, ontology-based access to financial information by the use of XBRL semantics in combination with multilingual ontology-lexicons.

During the current reporting period the following achievements were met in implementing the objectives as outlined above (corresponding to Monnet milestone II Focus and Development):

- Use Case Definition completed
- 1st version of Translational Processing Component completed
- CLOBIE Architecture completed
- 1st version of CLOBIE Prototype completed
- 1st version of Knowledge Access and Presentation Framework
- 1st version of Knowledge Access and Presentation Prototype completed
- 2nd iteration of Prototype & Evaluation Integration completed

[1] http://lexinfo.net/
[2] http://www.statmt.org/moses/