Master in Artificial Intelligence (UPC-URV-UB)

Master's Thesis proposal

General Information

Master's Thesis Title:	**Linguistically-driven Error Analysis in Machine Translation**
Orientation:	☒ professional ☒ research
M.Sc. Th. Advisor's Dept. & University:	LSI,UPC
M.Sc. Th. Advisor:	Lluís Màrquez and Jesús Giménez
M.Sc. Th. Advisor e-mail:	lluism@lsi.upc.edu, jgimenez@lsi.upc.edu
Observations:	The master thesis is involved in a project
Student's Name: (if already known)	

M.Sc. Thesis Description

Main issues / Brief Description:

The goal of the project is to collaborate in the design and implementation of a web interface for automatic Machine Translation (MT) evaluation, with particular focus on the development of smart mechanisms for the exploitation of linguistic information during error analysis.

Detailed Description:

In the last years, we have been developing Asiya (http://www.lsi.upc.edu/~nlp/Asiya), an open toolkit for automatic MT evaluation [1]. Asiya offers system and metric developers a text interface to a rich repository of evaluation measures. We have gathered existing measures and we have also developed new ones. Our core research line in automatic MT evaluation addresses the incorporation of linguistic knowledge into evaluation methods [2].

The aim of the project is to participate in the design and implementation of a visual interface for the Asiya Toolkit, in the shape of a new web application. This application will allow system and metric developers to upload their test suites and perform error analysis, automatic and manual evaluation, and meta-evaluation, using their Internet browser. The student will be responsible for the development of some of the central use cases. The emphasis of the research work will be on the definition of smart mechanisms for linguistically-driven error analysis.

The thesis work will involve:

- Study of automatic evaluation techniques [2]
- Study of the Asiya Toolkit [1]
- Design and Implementation of some of the central use cases based on the Catalyst web framework (http://www.catalystframework.org/).
- Proposal of mechanisms for the exploitation of linguistic information during error analysis
    - Study of techniques for the visualization of linguistic structures
    - Definition of mechanisms for identifying correctly/incorrectly translated parts

Other comments:

- Previous knowledge of the Perl programming language is desirable (not required)
- Experience in the development of web applications is desirable (not required)

[1] Jesús Giménez. and Lluís Màrquez. Asiya: An Open Toolkit for Automatic Machine Translation (Meta-)Evaluation. The Prague Bulletin of Mathematical Linguistics (94), 77–86, 2010.

[2] Jesús Giménez and Lluís Màrquez. Linguistic Features for Automatic MT Evaluation. To Appear in Machine Translation, 2010.

Barcelona, October 18th 2010