

Work package 2: Definition of Pre- Editing and Post-Editing rules

Description of work

Description of work

Task 2.1: Pre-editing rules for MT (acrolinx, UNIGE, SYMANTEC, month 1-18)

The main objective of this package is to define a set of lightweight pre-editing rules that can be used in an environment where content is generated by users (for example, forum, MS Word). A standard corpus analysis will first be conducted to identify areas where input normalisation is required (spelling, terminology, grammar, style) to ensure that input text is as understandable and consistent and matches as much as possible the coverage of the translation system. A simplification of long sentences and complex syntactic structures to lower the burden on MT will also be envisaged.

Task 2.2: Post-editing rules for MT (acrolinx, UNIGE, month 19-36)

A secondary objective is to define a set of lightweight post-editing rules that can be used in an environment where content is edited by users (for example, forum, MS Word). Machine Translation errors are different in nature to errors made by human writers or translators. Therefore, the postediting rules have to focus on these specific errors, set up on the base of an error analysis. These rules will then be tested for these MT post-editing tasks:

- Re-ranking: Statistical MT systems give ranked results. Rules that focus on typical errors of these systems can be use to re-rank these results.
- Reformulation: The rules can be used to reformulate erroneous sentences, containing for example spelling, grammar, or terminology errors.
- Pre- and post-editing rules can be used on the training data for SMT. Often, this data is inconsistent and erroneous, which leads to quality loss in MT. Spotting the problematic sentences and either deleting or reformulating could lead to better results.
The rules can be formulated with the acrolinx IQ rule formalism and applied using the acrolinx IQ checking and annotation tools.

Deliverables **Delivery date** **Description**

D 2.1	PM 12	Manage project meetings and reports. Definition of pre-editing rules for English and French (draft version)
D 2.2	PM 18	Definition of pre-editing rules for English and French (final version)
D 2.3	PM 20	Definition of post-editing rules for English, French, German and Japanese (draft version)
D 2.4	PM 30	Definition of post-editing rules for English, French, German and Japanese (final version)