Online MT System Internationalization: ATLAS PW1 / Lucy LT / MaTrEx in www.agenciatributaria.es

Pre-Production/Pre-filters
Translate ▶ Adds constant marks to block non-translatable text nodes and attributes.
Domain ▶ Generates a HTML section per domain, and adds such information to the section sent to the MT System to select the appropriate TM and vocabulary (glossary).
Localization Note ▶ Encoding of information to convey it to the post-editor. Localization quality issue ▶ Encoding of information to convey it to the post-editor.

Content Editor + FMT:
Localization Note ▶ Blocked by the Content Editor but visible for the post-editors.
Domain ▶ Stores the revised texts in the domain specific TM.
Provenance ▶ Adds and blocks the information about the reviser
Localization Quality Issue ▶ Adds and blocks the quality issue inserted by the reviser.

Post-production/Post-filters:
Language Information ▶ Updates the language attributes in the translated content.
Provenance ▶ Adds and blocks the information about the MT System.

Text Handling – Pre/Trans Domain ▶ Set translation parameter "Subject Areas" according to global Domain mapping.
Text Handling - PreTranslate ▶ Hide nodes that should not be translated.

Domain ▶ Set internal MaTrEx parameter: selection of the appropriate SMT modules (TM / LM / RM / weights) for each segment to be translated
Language Information ▶ Set internal MaTrEx parameters: source language
Translate ▶ Mark nodes that should not be translated.

Translate ▶ Force translations of segments that should not be translated (i.e. target (output) ≠ source (input)).
MT Confidence ▶ Generate MT confidence scores for each translated segment

Locate Filter ▶ Pass information as-is (relevant for post-editors / localization)
Language Information ▶ Update language information after translation.
MT Confidence ▶ Update the MT Confidence info. after concatenating segments into one document.

Benefits, increases user’s control and automatic decision processes:
Translatability and language pair selection (Translate, Language information) ▶ Specific terminology (Domain) ▶ Activation rules for post-editing (Localization Note) ▶ Quality aspects report to translations consumer and post-editor (Localization Quality Issue) ▶ Post-editors judge translation’s quality (MT Confidence) ▶ Identification of agents (provenance).

ITS2.0: The “Dublin Core” of the Multilingual Web
The MultilingualWeb-LT Working Group receives funding by the European Commission (project name LT-Web) through the Seventh Framework Programme (FP7) in the area of Language Technologies. Grant Agreement No. 287815.