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

FINAL USER

1 2 3 4

INTERNET

SPANISH HTML5
GOBIERNO DE ESPAÑA
Agencia Tributaria

ENGLISH HTML5
GOBIERNO DE ESPAÑA
Agencia Tributaria

ATLAS REALTIME
MULTILINGUAL PUBLICATION SYSTEM

linguaserve

Web Administration Interface

Database

Content Editor

Pres-filtrers

Cahe Module

File Management Tool

MT System Interface

APPLICATION CORE

Post-filters

Interface

ITS 2.0 Engine Module

SMT System

RBMT System

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.
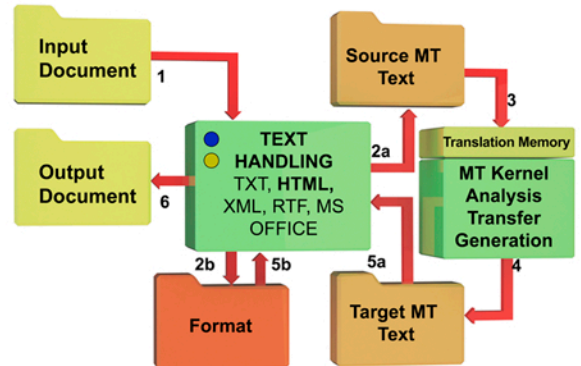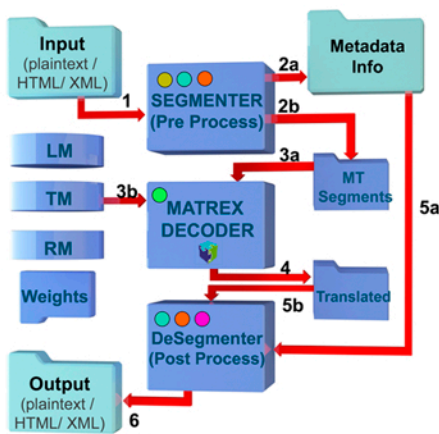**Provenance** ▶ Adds and blocks the information about the MT System

DCU CNGL

MATREX

lucy LT

SOFTWARE lucy AND SERVICES

● **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.

● **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.

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
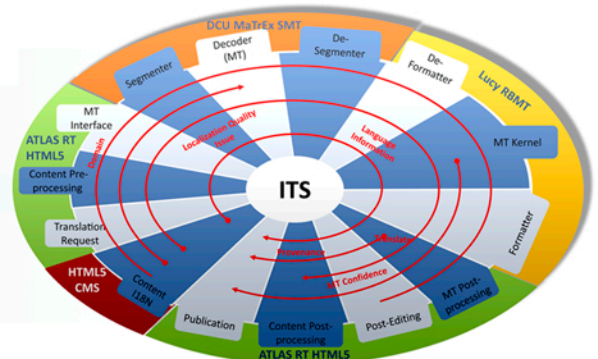
Input (plaintext / HTML/ XML)
LM
TM
RM
Weights

SEGMENTER (Pre Process)
1
2a
2b
Metadata Info

MATREX DECODER
3a
3b
4
5b
MT Segments

Translated
5a

DeSegmenter (Post Process)

Output (plaintext / HTML/ XML)
6

Input Document
1

Output Document
6

TEXT HANDLING TXT, HTML, XML, RTF, MS OFFICE
2a
2b 5b

Format

Source MT Text
3

Translation Memory
MT Kernel Analysis Transfer Generation
4

Target MT Text
5a

● **Text Handling – Pre/Post Language Info** ▶Hide nodes not matching the current translation direction (source). Update language information after translation (target).

● **Locale Filter** ▶ Pass information as-is (relevant for post-editors / localization)
● **Language Information** ▶ Update language information after translation.
MT Confidence ▶Update the MT Confidence inf. 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)

DCU MaTrEx SMT
Decoder (MT)
Segmenter
De-Segmenter
De-Formatter
Lucy RBMT
MT Interface
ATLAS RT HTML5
Content Pre-processing
Translation Request
HTML5 CMS
Content I18N
Publication
Content Post-processing
ATLAS RT HTML5
Post-Editing
MT Post-processing
Formatter
MT Kernel
ITS