Business Benefits of ITS 2.0-Aware Interoperability CMS-TMS and Online Multilingual Publishing System

Dr. Pedro L. Díez Orzas
Linguaserve specializes in multilingual web advanced solutions for 21st Century Challenges.

Experience in interoperability since 2002 and real time multilingual web publishing since 2008.

www.linguaserve.com

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.
MLW-LT and ITS 2.0

• Linguaserve is a member of MultilingualWeb-LT because:
  
  – **Standards** help both **SMEs** and large companies.
  – There is no magic button: human language and translation are extremely complex. Web content annotation **greatly helps** to improve results in Multilingual Web Linguistic Technology.
1. ITS 2.0-Aware Interoperability CMS-TMS

2. ITS 2.0-Aware Online Multilingual Publishing System

3. Horizons
Use case: VDMA

- VDMA: German machinery and plant manufacturers' association
- Largest industrial association in the capital goods industry in Europe (3170 industrial members)
- Highly export-oriented

Deutscher Maschinenaußenhandel/German machine trade

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.
Use case: Scope

- Annotated, processed and translated 150 press releases
- 75,000 words annotated and processed with ITS 2.0
- Using Drupal MLW-LT modules
- ITS 2.0-aware automatic content round tripping
- Languages: DE > FR, ZH
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.
The MultilingualWeb-LT Working Group receives funding by the European Commission Framework Programme (FP7) in the area of Language Technologies. Grant Agreement No. 287815.
Some ITS 2.0 benefits (I)

- ITS 2.0 increases both user’s control and automatic decision processes, allowing more “intelligence” in the systems involved:
  - **Translate**: can be automatically or manually annotated. E.g. it allows not to add “non-translatable” terms in several specific glossaries or MT systems.
  - **Localization Note**: direct communication between webmasters, PMs and translators. When *alert* type, it can be used for triggering certain processes in the Translation Workflow.
  - **Domain**: automatic selection of CAT/MT terminology and dictionaries. Selection of Translation Memories by domains.
Some ITS 2.0 benefits (II)

- **Language Information**: quality check to ensure the source language content is according to the Webservice parameter.
- **Allowed Characters**: quality check for the target content.
- **Storage Size**: quality check for both original content and the target content to be uploaded into the CMS. It can be used also for translators’ visual control.
- **Provenance**: possibility to reassign the same translator/reviewer in new versions of the same content and inform the PM. Tracking control in the CMS.
- **Readiness** (ITS 2.0 extension): control of processes to be done, date control for availability, delivery and priority.
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.
Win-win business

- Content creators and editors, CMS developers, the localization industry and users are benefited.
- More efficient control over the content and faster fine-grain communication between localization chain actors (e.g. webmaster/project manager)
- Localization platforms and format independent:
  - Better web and linguistic technology machine/machine interaction
  - Better web and localization human/machine interaction
- Increasing fully automatic processes and localization expert systems in CMS and TMS.
- Opens up ways for connectors, pre- and post-editing, and CAT tools.
Business: Opportunities rise from needs

- Very frequently updated web sites that need efficient multilingual updates and maximum control:
  - Corporate and industry information
  - e-Government
  - e-Commerce
  - Educational web sites
- Highly distributed content creation through the CMS
- Web 2.0 and user content created
  - Applying MT systems for immediacy
- Using ITS 2.0 to contribute for multilingual SEO
1. ITS 2.0-Aware Interoperability CMS-TMS

2. ITS 2.0-Aware Online Multilingual Publishing System

3. Horizons
Use case: the Spanish Tax Agency

- [www.agenciatributaria.es](http://www.agenciatributaria.es) is the user in the “Online MT System” showcase in MLW-LT
- **Spain: General Indicators 2011**
- Spain is a country that is regionally structured into 17 autonomous communities and 2 autonomous cities with 5 co-official languages
- Population: 47,190,493 inhabitants (12.2% foreign residents)
- **Mission of the Spanish Tax Agency**
- Effective application of Spain’s tax and custom system
- Management of tax resources on behalf of other public administrations when required by Law or Agreements
- **General taxpayer census**
  - Individual taxpayers: 46,509,231
  - Companies: 2,674,547
  - Other organisations: 2,293,939
  - **Total taxpayers:** 51,477,717
Use case: Scope

- Online MT System Internationalization showcase components:
  - ITS 2.0
  - HTML5
  - ATLAS RT (Linguaserve’s Real-time Multilingual Publishing System)
  - Lucy Software MT (Rule-based Machine Translation)
  - MaTrEx, from Dublin City University (Statistical Machine Translation)
  - www.agenciatributaria.es (CMS: OpenText WEM)
- RTMPS implementation and deployment in pre-production
  - ITS 2.0 data categories: 6 (Translate, Localization Note, Language Information, Domain, Provenance, Localization Quality Issue)
  - Prototypes, test suite engines, and use case
- 250 web pages ES-EN and 30 web pages ES-FR, ES-DE
  - Content annotation and MT post-editing (EDI-TA methodology)
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.
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.
ITS 2.0 immediate benefits

• ITS 2.0 increases user’s control and automatic decision processes:
  – Translatability and language pair selection (Translate, Language information). It allows not to add “non-translatable” terms in specific glossaries or MT.
  – Specific terminology to apply in the MT (Domain)
  – Activation rules for MT post-editing (Localization Note)
  – Quality aspects reported to translation consumer or post-editor (Localization Quality Issue)
  – Post-editors judge quality of translation (MT Confidence)
  – Identification of agents (provenance)
Shifting gears: New cost structure

--

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.
# MLW-LT Online MT Business Case

## Strengths

- Lower translation costs (MT + PE) depending on % of post-editing (E.g. 100% post-edited: -30%)
- Management costs: higher setup / lower maintenance (-60%-80%)
- Non-invasive technology
- Real-time or fast post-edition

## Weaknesses

- Viability dependent on:
- Language combination and MT system output

## Opportunities

- Web sites with daily high volume updates: Ecommerce, Administration, Corporate news and publications, user content generated (social media)
- In house installation for > 1 million words and frequently updated

## Threats

- Control, performance and security: The client might lose control of the translation: solved with ITS 2.0
- Real-time performance of MTs
- Security level in shared RTMPS
- Needs pre-editing and post-editing tools (ITS 2.0 and HTML5)
Business: Opportunities rise from needs

- **e-Commerce**
  - Very high volume and rotation
  - Short texts and repetitive descriptions
    - Better for MT
    - Quicker to post-edit
  - Very sensitive to ITS 2.0 benefits

- **e-Government**
  - Controlled language and content policies

- **HTML from several CMS and other applications (Content source independent)**

- **Web 2.0 and user content created**
  - GIST translation
  - Immediacy

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.
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.
1. ITS 2.0-Aware Interoperability CMS-TMS

2. ITS 2.0-Aware Online Multilingual Publishing System

3. Horizons
Transversal View of Multilingual Web

- **“Multilingual sensitive”** content creation and exploitation:
  - Automatic and assisted annotation and pre-editing tools to empower content creators and webmasters adoption.
  - Linked Open Data and enricher systems and tools.
  - CMS, Web publishing systems, browsers, SEO and web applications.
  - CMS-TMS interoperability standards (“Readiness”).

- **“Web sensitive”** translation systems
  - Management and Business Systems.
  - Expert translation and localization systems and workflows.
  - Performing and customizable Machine Translation (MT).
  - MT Post-editing, Computer Assisted Translation and Terminology.

**ITS 2.0: innovation opportunities for key multilingual web applications.**
The Multilingual Web Server

• ITS 2.0 improves a full equipped multilingual web server:
  – Interoperable Web Services.
  – Real-time multilingual publishing and translation.
  – Multilingual and interlingual Linked Open Data compliant.
  – Collaborative and cloud solutions between MT and human translation.
  – Maximizing re-use and productivity in pre and post-editing, lexical, terminological, and CAT Tools.
  – Automated Quality Assurance.
  – Operational and management virtual offices and interfaces.
Thank you!

pedro.diez@linguaserve.com