

# ITS 2.0 Automated Translation of Natural Language Content

## MultilingualWeb Linguistic Technology

Pedro L. Díez Orzas

Linguaserve I.S. S.A.

Seminario de Nobles, 4

28015 Madrid (Spain)

[pedro.diez@linguaserve.com](mailto:pedro.diez@linguaserve.com)

Tel.: +34 917616460

**Abstract**— We<sup>1</sup> present in this paper how ITS 2.0 constitutes a powerful metadata set which can be represented in different formats and “is used to annotate web content (referred to henceforth just as content) to facilitate its interaction with multilingual technologies and localization processes with the aim of publishing that content on the Web in multiple languages.” As a standardization initiative, it provides real life technical and business advantages.

Twenty use case applications have been developed within the MultilingualWeb-LT working group using different ITS 2.0 data categories. We will show two main implementations and their technical and business benefits:

**ITS2.0 CMS to TMS implementation: Drupal to GBC Server/PLINT Showcase with VDMA:** The large volume of information and web content justifies the use of CMS systems for medium to big size companies and organizations. They provide benefits such as content control, multiple user profiles, abstraction and workflows. When introducing multilingualism to the picture, a translation workflow is highly recommended.

**ITS 2.0 Real Time Multilingual Publishing and Translation System: ATLAS RT, Lucy LT, MaTrex for the Spanish Tax Office.:** The vast volume of content in Web 2.0 and 3.0, or e-commerce, which is growing daily, requires real time multilingual publishing and translation systems that provide proper quality and performance.

**Keywords**— *Multilingualweb, Internationalization Tag Set 2.0, ITS 2.0, web localization, interoperability, Content Management System, Translation Management System, metadata, web translation.*

---

<sup>1</sup> Pedro L. Díez Orzas is founder and CEO at Linguaserve I.S. S.A., a company providing cutting-edge multilingual solutions. PhD in Computational Linguistics, is also Professor at the Univ. Complutense de Madrid. Expert in Language Engineering & Translation, with 23 years of professional experience in R&D in language industries, he developed his career at Novell as head of multilingual semantics; at WordPerfect as Head of Spanish Lexicography; and as European Project Coordinator (INTERLEX), Technical Manager (EuroWordNet), and also in several universities. Currently representing Linguaserve in the EU Project MLW-LT.

### I. INTRODUCTION

ITS 2.0 identifies and selects a number of concepts and defines implementations of these concepts, called ITS data categories, as a set of elements and attributes. ITS is not merely a tagging or labeling standard, but a conceptual system of elements and attributes for web content internationalization, translation and localization, which can be represented in different formats. ITS 2.0 provides 19 data categories and supports XML-based formats and HTML5, and it is useful for XHTML, and CMS-based ‘deep web’, DITA, DocBook, and mapped to RDF/NIF and XLIFF.

Its success is expected to materialize in real life implementations for different Use Cases, demonstrating that the development of this metadata standard is done through broad consensus across communities, involving content producers, localization workers, language technology experts, browser vendors and users. Currently, there are some 20 different implementations. Here we will show two of them.

### II. STANDARDS, THEY ARE GREAT. EVERYONE SHOULD HAVE THEIR OWN

Standards are sometimes produced in excess, making them compete with one another for the same purpose. This fact causes a situation which is exactly the opposite of the objective pursued by standardization initiatives: the proliferation of standards for the same purpose makes them become useless or restricted to specific communities that become isolated. This is especially true in those cases in which new standards are sought for existing or old needs.



By contrast, new technologies and paradigm shifts that occur in all disciplines require new rules for new needs. It is in this context that we

feel the need for a metadata reference standard for the multilingual web, like a new requirement for linguistic technology on the Internet. So the viability of the Web's multilingualism depends on the mechanization level of the methods and processes, and undoubtedly a certain level of metadata standards is crucial and can help to make this happen.

The multilingual information and knowledge society demands the development, dissemination and adoption of new standards. The problem is that the speed of this society does not allow this to take as long as the 'Space Shuttle and the horse's Rear End' did.



Fig. 1. Space Shuttle and the Horse's Rear End<sup>2</sup>

In addition, standards help both SMES and large companies. It helps SMES because they can better compete by using existing open standards to fulfill certain requirements faster, and also making them more compatible, thus avoiding customer reluctance on small businesses' proprietary software.

Large companies would also benefit in at least two ways: first, in cases in which they are heavily involved in the development of a standard, it can make them lead the market by leading the creation and exploitation of the standard; second, because it can help boost the communities of developers and users that appear around certain technologies, helping to facilitate new extensions and features by using standards.

Lastly, open source communities could certainly become open-open, i.e. open source based on open standards.

### III. MULTILINGUAL WEB LT AND ITS 2.0

In the last 10 years it has become increasingly clear that one of internet's new needs is to set the foundations for the integration of Language Technologies (LT) into core Web technologies (especially its standards family).

There is no magic button for multilingualism in the Internet. Human language and translation are extremely complex. Web content annotation greatly helps to improve results in Multilingual Web Linguistic Technology.

The new requirement has therefore been setting a standard that defines metadata on:

- Information in Web content that is relevant for language technology processing.
- Processes for creating Web content via localization and content management workflow.

<sup>2</sup> See <http://www.astrodigital.org/space/stshorse.html>.

- Language technology applications, tools and resources used in applications that employ or support this standard.

In order to achieve these objectives, the W3C MultilingualWeb-LT Working Group was created in 2012, funded by the European Commission (project name LT-Web) through the Seventh Framework Programme (FP7) in the area of Language Technologies (Grant Agreement No. 287815). MultilingualWeb-LT Working Group (MLW-LT WG) has continued the work carried out by MultilingualWeb community ([www.multilingualweb.eu](http://www.multilingualweb.eu)) and Internationalization Tag Set 1.0 to accelerate the process and the efforts already made.

Their mission was to enhance, improve and define new metadata for Web content (mainly HTML5, but not exclusively) and 'deep Web' content, such as CMS or XML files, from which HTML pages are generated, that facilitate Web content interaction with multilingual technologies and localization processes: Internationalization Tag Set 2.0 (<http://www.w3.org/TR/its20/>).

The dissemination and adoption of this new standard involved the consensus of different communities: content producers, localization workers, language technology experts, browser vendors and users. It is also necessary to join existing networks between stakeholders in these communities to ensure long-term adoption.

After almost two years, the MLW-LT Working Group will continue and coordinate their works in the ITS Interest Group (see ITS Interest Group, 2013) with the stakeholders who have developed ITS 1.0 and 2.0 and other that are now active.

#### A. Internationalization Tag Set 2.0

ITS 2.0 provides the following data categories<sup>3</sup>: Translate, Localization Note, Terminology, Directionality, Language Information, Elements Within Text, Domain, Text Analysis, Locale Filter, Provenance, External Resource, Target Pointer, Id Value, Preserve Space, Localization Quality Issue, Localization Quality Rating, MT Confidence, Allowed Characters, and Storage Size. ITS 2.0 also introduces or modifies important mechanisms for ITS 1.0. Some of the most important are local and global explicit selection rules. As we can read in the specification document, both local and global approaches can interact with each other, and with additional ITS dimensions such as inheritance and defaults. ITS 2.0 can be used with XML documents (e.g. a DocBook article), HTML documents, document schemas (e.g. an XML Schema document for a proprietary document format), or data models in RDF<sup>4</sup>.

#### B. ITS 2.0 implementations

Along with the standard specification itself, the MLW-LT WG has created interoperable implementations<sup>5</sup>. 'Interoperable' meaning that at least the metadata must be available in various parts of Web-related technologies like

<sup>3</sup> See <http://www.w3.org/TR/its20/#basic-concepts-datacategories>.

<sup>4</sup> See <http://www.w3.org/TR/its20/#basic-concepts-selection>.

<sup>5</sup> See [http://www.w3.org/International/its/wiki/Use\\_cases\\_-\\_high\\_level\\_summary](http://www.w3.org/International/its/wiki/Use_cases_-_high_level_summary).

CMS systems, localization chains, etc., or in profiles of related formats like HTML5, XML, XHTML and XLIFF, as well as Natural Language Processing Interchange Format (NIF). As part of their activities, project members and members of the Working Group compiled a list of usage scenarios that exemplify how ITS 2.0 integrates automated processing of human language into core Web technologies.

In this paper we present some important aspects and benefits of ITS 2.0 via two of the implementations of leading edge web localization technologies:

- Interchange between Content Management System and Translation Management System
- Content Internationalization and Advanced Machine Translation

MultilingualWeb-LT has also laid the technical foundations for new business opportunities for content creators and vendors who provide them with language and content services and tools. This will enable content creators and distributors to reach out to a growing linguistic and cultural variety of Web users

worldwide, and to respond to their specific needs in a timely and cost-effective manner. We will try to also show some business implications of those two implementations.

#### IV. ITS 2.0 INTEROPERABILITY CMS/TMS

The large volume of information and web content justifies the use of Content Management Systems (CMS) and Document Management Systems (DMS). In the web context, CMS provide benefits such as content control, several user profiles, abstraction and workflows. When we include the multilingual variable to the CMS picture, a translation workflow is highly recommended. The advantages of using an external Language Service Provider (LSP), Machine Translation (MT), professional Human Translation (HT) and Computer Assisted Translation tools (CAT tools) gives added value such as the use of translation memories, glossaries and the experience with translation management for the web localization process.

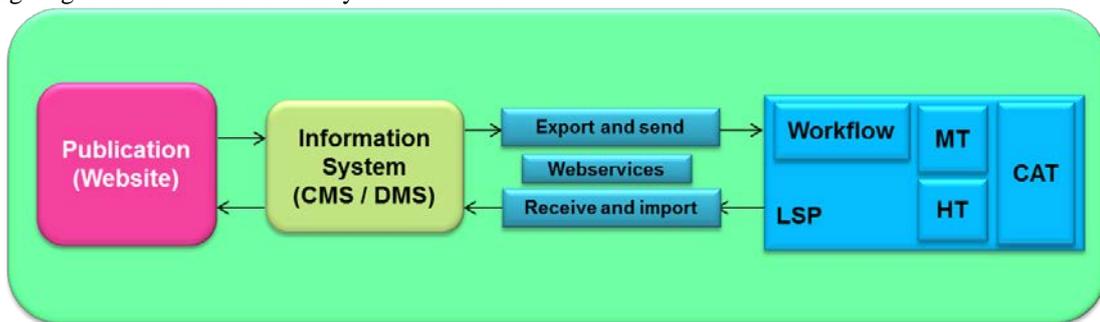


Fig. 2. Automated off-line translation system (Interoperability)

This use case will exemplify how ITS 2.0 allows more integration between both sides and how the contents' localization workflow benefits from each data category implemented. The content originates in a CMS, and gets exposed/serialized as XHTML + ITS 2.0. Then it's sent to a TMS and processed in a workflow. Upon completion, the TMS exposes/serializes localized/translated XHTML + ITS 2.0 to the CMS. For this showcase, Cocomore and Linguaserve had the Verband Deutscher Maschinen- und Anlagenbau (VDMA) as final user. VDMA, German machinery and plant manufacturers' association, is the largest industrial association in the capital goods industry in Europe (3170 industrial members), highly export-oriented. The technology involved has been extended to be ITS 2.0 compliant.

In the CMS side (Cocomore): DRUPAL is a free and open-source content management framework (CMF) written in PHP. Drupal version 7 was used for this implementation.

In the TMS side (Linguaserve):

- Global Business Connector Server (GBC Server) is a mature Globalization Management System. The interoperability applies Webservices and XML, and it can be connected to any CMS, DMS, or ERP, or proprietary systems.

- Platform for Localization, Interoperability and Normalization of Translation (PLINT) is a workflow control platform for translation and localization processes.
- Computer Assisted Translation tool.

Interoperability is achieved by the offline translation system interface between the "Client Application" (Drupal) and Linguaserve's translation server (Global Business Connector Server). GBC Server provides a number of webservices for the translation and localization of content, as well as for the exchange of other information necessary for the production and monitoring of orders. It provides a service with which the Client Application can check the correct functioning of the connections (Proxy, client's Firewall, etc.) without the CA having to simulate sending files. GBC Server uses Simple Object Access Protocol (SOAP).

This integrated approach affects almost all areas of the traditional translation workflow. Accordingly, it requires modifications and extensions throughout the tool chain. The following graphic shows a vision of the architectural entities that are involved in ITS 2.0-aware content and translation handling:

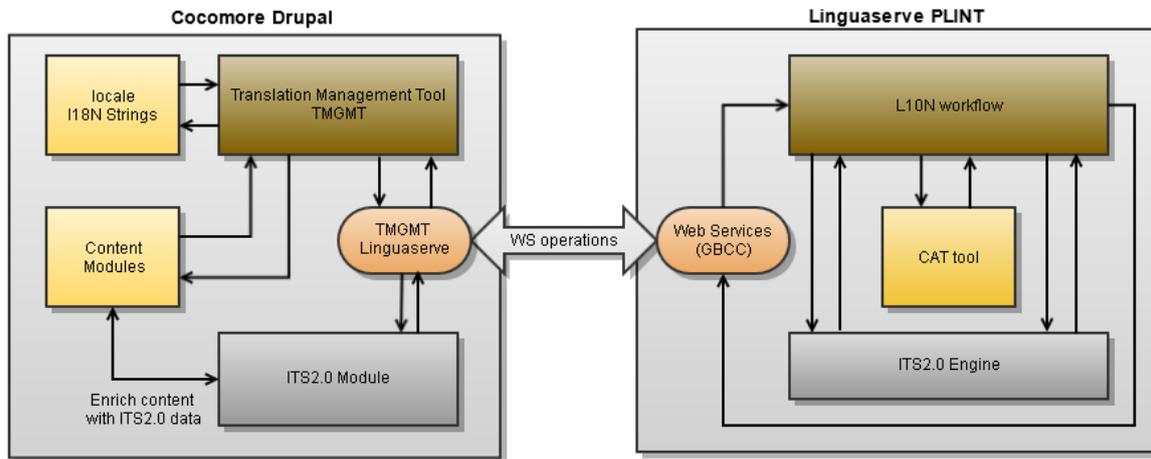


Fig. 3. ITS 2.0 CMS-TMS use case

Features of the solution created based on this architecture can be assigned to either the content provider's or the LSP side of the picture.

#### A. ITS 2.0 metadata aware workflow

1) On the content provider's side the creation of the ITS 2.0 metadata aware workflow involves the following areas:

a) Annotation of source language content with ITS 2.0 metadata within the Drupal CMS. Manual (toolbar buttons, context menu, keyboard shortcuts), automated (through a standardized interface), both as part of the content creation process or as a separate step.

b) Transparent data round-tripping via export/import of XHTML+ITS 2.0 markup files from Drupal, to be sent to/received from GBC Server. The process is based on an extended version of the Drupal Translation Management (TMGMT)-module.

c) Translation review for QA purposes.

2) On the LSP side, the creation of the ITS 2.0 metadata aware workflow encompasses three areas:

a) Pre-production/post-production engine for processing content files annotated with ITS 2.0.

b) LSP internal localization workflow to provide support to project management and production processes

c) Computer Assisted Translation (CAT) tool usage for translation, revision and postediting with ITS 2.0-annotated content.

#### B. Test case statistics

In the source texts, a total of 141 documents with more than 75 thousands words were annotated with the following data categories: Translate, Allowed Characters, Localization Note, Storage Size, and Language Information. The total annotation of source texts was 5.544 tags, 4.700 tags manually annotated and the rest automatically, and the density of tagging in the sources texts was 39, 3 tags per document.

In the localization processing and translation (from German into French and Chinese), other two data categories were annotated: Provenance and Readiness. "Readiness" is not part of ITS 2.0 since, in the given time frame, the W3C MultilingualWeb-LT Working Group could not reach consensus on all aspects of Readiness. The implementation of Readiness as an extension to ITS 2.0 allows to gather experience and to consider this data category for a future version of ITS.

The resulting distribution of data categories used was Translate (with value: no) 69,3%; Allowed Characters 11,3%; Provenance 5,4%; Language information 4,3%; Localization Note 3,8%, Storage Size 2,3%; and Readiness 2,3%.

#### C. Business case

ITS 2.0 improve CMT-TMS interoperable localization solutions in different aspects:

##### 1) Management

Specific information (notes, issues, etc.) can be coded in the content, allowing its automatic processing and fine grain communication between different actors, like webmaster and project manager, and this with translators and reviewers, with a delivery time reduction,

"Readiness" makes the process more efficient and significantly decreases management costs.

##### 2) Quality and control.

More efficient control over the content, ITS 2.0 allows a more dynamic and precise control from inside the content, increasing translation control.

Metadata independency from CMS technologies, localization platforms and formats.

Quality Assurance: it allows new and better mechanisms for quality assurance.

##### 3) Cost

On the other hand, although there is not a direct impact in translation cost, it enormously increases the capacities to

influence and improve quality and allows externalizing certain tasks, which has a final impact in the translation costs as well.

Finally, saving and reducing further corrections and interventions highly reduces non-quality costs.

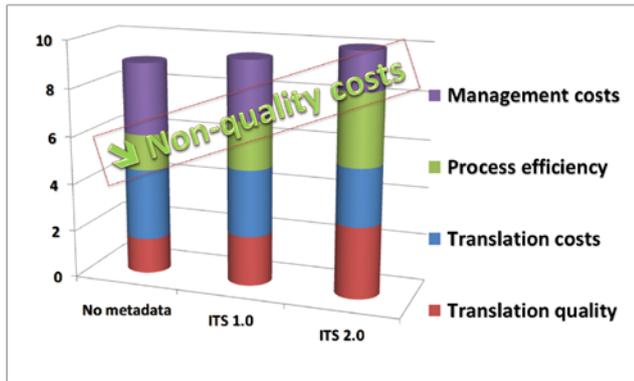


Fig. 4. ITS 2.0 impact

Some opportunities: ITS 2.0 can become an accelerator for the adoption of interoperability localization chains, improving existing clients by, for example, measuring the non-quality costs and client’s project management reduction, and increasing fully automatic processes and localization, like expert systems.

ITS 2.0 can also help SMEs by reducing the effort they have to make for complex proprietary developments and analysis and format specifications. Normalization also means availability and technological simplification, and it simplifies interoperability implementations.

New methodologies, methods and Language Technologies in CAT tools, post editing tools, Machine Translation, and Translation Management Systems can be developed applying ITS 2.0. Also new standards such as interoperability or Quality Assurance can be supported by ITS 2.0. Up to now, there are no CMS-TMS interoperability standards. There are also new threats created by ITS 2.0:

- Adoption by content creators: Annotation of source content. This is probably the greatest threat, convincing

content creators about the benefits of annotating the source content in the CMS.

- Adoption by CMS developers: providing the content creators with suitable tools for automatic, semiautomatic, and manual annotation.
- SEO applications: developing strategies to use ITS 2.0 metadata to improve content coherence for SEO and SEM.

## V. ONLINE MT INTERNATIONALIZATION

Recent advances in machine translation (MT) are emphasizing the importance of language technology for the Web. An ever-increasing need for a proper and useful metadata for content processing is in the spotlight nowadays. The metadata set offered by ITS 2.0 intends to address the “black box” problem of the three gaps in the chain of content processing and consumption on the Web, which are:

1. An online machine translation service might make mistakes like translation of fixed terminology or named entities. Language technology does not know about metadata in the source content, e.g. “What parts of the input should be translated?”

2. In the database from which the translated text has been generated, the information about translatability might have been available. However, the machine translation system does not acknowledge that kind of “hidden Web” information. There is no available description of the processes, which was the basis for generating “surface Web” pages.

3. The last gap is about clearly defined identification of metadata: Information like gaps 1 and 2 is not described in a standardized manner.

These gaps are especially important when the localization system only has the html content as input. ITS 2.0 specification has been implemented in an Online MT System to take advantage of its properties and features in order to improve the process of real time publishing and translation.

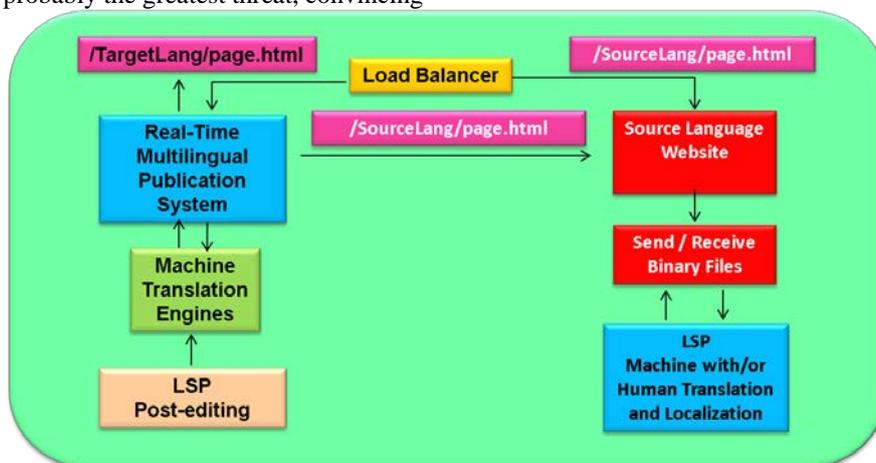


Fig. 5. Real Time Multilingual publication system

For the use case, the final user was The Spanish Tax Agency ([www.agenciatributaria.es](http://www.agenciatributaria.es)). The implementers: Linguaserve, Lucy Software and Dublin City University.

The following technology was involved:

- ITS 2.0 in HTML5
- Linguaserve's ATLAS Real Time: Real Time Multilingual Publishing System (RTMPS)
- Lucy LT (Lucy Software): Rule-Based Machine Translation
- MaTrEx (DCU): Statistical Machine Translation

ATLAS Real-time is a real-time solution for multilingual publications using the Internet. ATLAS Real-time allows the user to navigate a website in a completely transparent way in the selected languages. The system is fully scalable and configurable, allowing multi-server redundant configurations, and can be integrated in the client's Web site.

When ATLAS Real-time receives a translation request, it downloads the original document, sends it to the MT System

and finally delivers the target language translated document to the user's browser.

The ITS 2.0 data categories have been used are Translate, Localization Note, Language Information, Domain, Provenance, and Localization Quality Issue. Different prototypes and test suite engines have been developed, and the use case was built with the Spanish Tax Agency web ([www.agenciatributaria.es](http://www.agenciatributaria.es)) in HTML5 and contains 250 web pages to display from Spanish into English, and 30 web pages into French and German.

Migration from HTML4 to HTML5, and automatic, semiautomatic and manual content annotation of source HTML5 web pages and MT post-editing using EDI-TA methodology (see Díez Orzas, P.L., Rico Pérez, C. et al, 2013) was carried out.

Here is a diagram showing the basic workflow of the overall system:

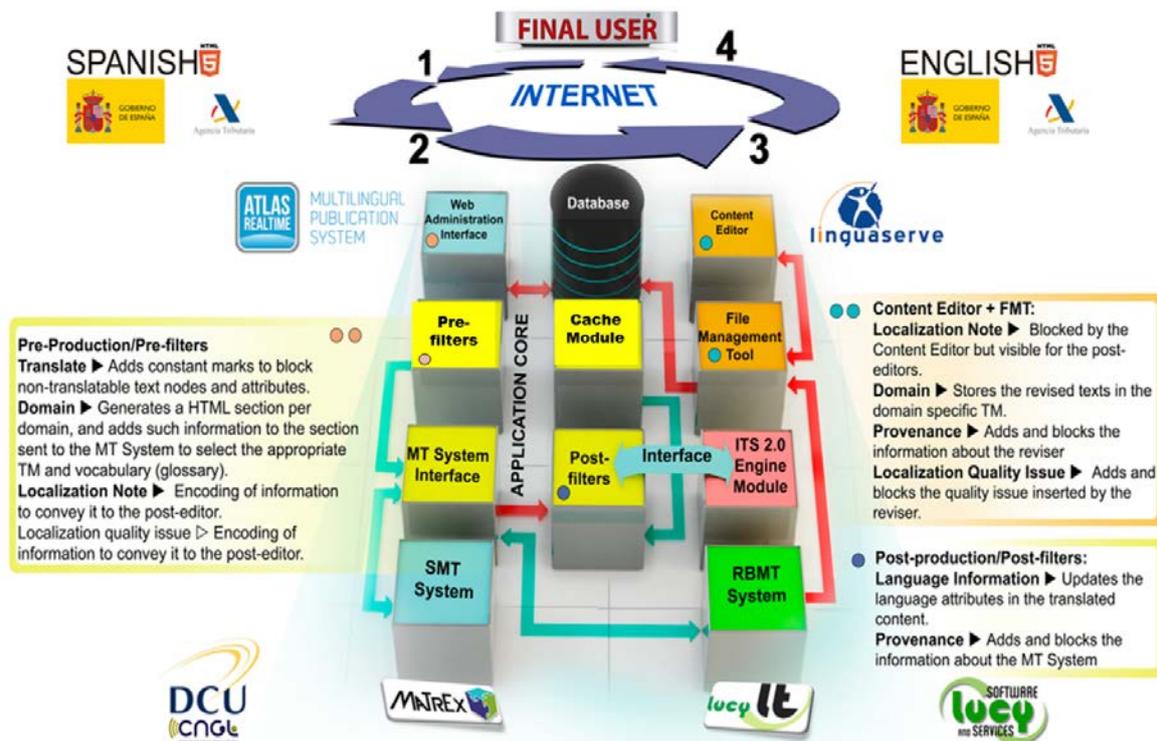


Fig. 6. ATLAS Real-time modules configuration

#### A. ITS 2.0 metadata aware workflow

This illustrates what happens from the moment the system receives a translation request until it delivers the translated content to the user's browser:

- The user makes a translation request and the client's Web Server redirects the request to ATLAS Real-time.

- ATLAS Real-time receives a translation request, downloads the original document, and checks whether it has been translated before or not to send it to the MT System.
- ATLAS Real-time also pre-processes the ITS 2.0 metadata of the file.

- If the content of the original document, or piece of the document, remains unchanged since the last translation request processed for that very document, the system retrieves the already translated content from the system's database.
- If the content of the original document is new or was not previously processed, then it's sent to the MT System.
- The MT System, after parsing the content to generate the list of translatable segments, checks if any of those segments is already stored in the Translation Memory (TM).
- For all of those who are not stored in the TM, the MT System creates a pair of input and output files, which are processed and conveyed to the Content Editor.
- Once in the Content Editor after the content of these input and output files is edited and validated to update the TM.
- The MT System sends the translated content back to ATLAS Real-time.
- ATLAS Real-time also post-processes the ITS 2.0 in the file.
- Finally the translated document is displayed in the user's browser.

#### B. Business case

As we said, new needs require new approaches. This use case demonstration illustrates how ITS, via a Real-time Translation System connected to different MT Service Providers, allows to:

- Translate HTML5 documents from an ITS-conformant Web CMS.
- Control a process depending on its progress, convey information to editors and indicate the state of the content to the user.
- Communicate instructions about language, domain and translation, and convey information about the translation to a content editor.
- Communicate instructions and information to a posteditor regarding the state of a process, and also to inform the user of its status.
- Communicate the identity of agents that have been involved in the revision and the translation of the content, and to allow translation quality reviewers, to evaluate how the performance of these agents affects the quality of the translation.

This reduction mainly affects the translation tasks, which need other tasks and responsibilities to be increased due to the nature of this kind of solutions. For instance, the effort in project preparation and setting is higher. The cost of MT-Post-editing is higher than the revision of human translation, as well as the technical work and Q&A and Testing. All these

increases are absorbed by the translation cost reduction, with a net result of overall project cost reduction.

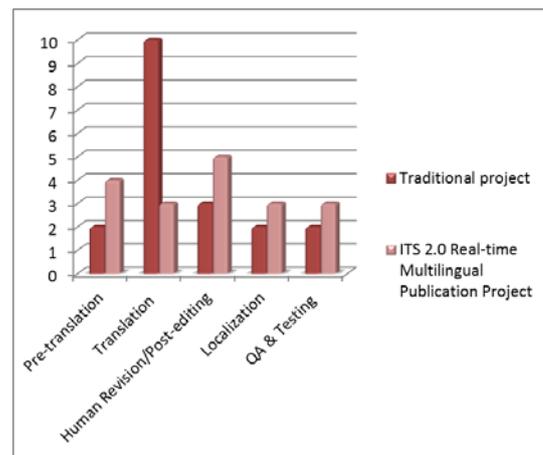


Fig. 7. Cost structure

Other tasks like project management vary from implantation to maintenance phases. In the initial phase, which includes implantation, the project management effort is higher, while later on, in the long term, maintenance efforts are lower.

In conclusion, using RTMPS with ITS 2.0 can reduce a multilingual web project at least in 40%:

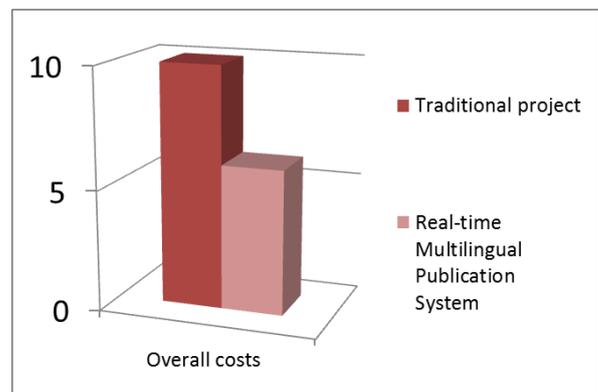


Fig. 8. Cost reduction

But of course, a full dedicated implementation of a RTMPS for a client has some constraints concerning size and volume for the project to be profitable. RTMPS platforms shared between several clients makes SMEs with lower project size and volume viable and profitable.

#### 1) Strengths

a) *Translation Cost Reduction:* In Real-Time Multilingual Publishing Systems, using Machine Translation and post-editing reduces costs drastically. The post-editing percentage depends on the quality of the MT output and the nature of the texts. Depending on post-editing percentage, cost reduction increases. For example: MT + 100% post-edited might have a Translation cost reduction of circa -30%. It

allows developing the concept of “Quality on-demand”. ITS 2.0 reduces efforts in tasks dependent on the MT system used. For example: instead of codifying in each different MT system the lexicon excluded from translation, this information is coded with ITS 2.0 and used by all the different MT Systems.

b) *Management cost reduction*: Although the implantation and setup have higher costs, they are much lower during the subsequent maintenance periods (-60%-80%). ITS 2.0 allows a more effective fine-grained coordination and management thanks to having the client’s requirements and indications reach post-editors by using metadata.

c) *Increase of Translation Control*: Control by the client over the quality or the customization of the translation is external to the HTML5 content (glossaries, MT rules, RTMPS parameters...). ITS 2.0 allows a more dynamic and precise control from inside the HTML5 content.

d) *Non-invasive technology*: clients do not need to install an application.

e) *Delivery time*: Real-time or fast post-edition.

2) *Weaknesses*: Differences on MT system outputs. Highly dependent on the language combination. E.g. Post-editing difficulty differs between language pairs: those more typologically close (ES↔FR, ES↔PT, ES↔CA, ES↔GL) are easier, those farther (ES↔EN, ES↔EU, ES↔DE, or FR↔RU) are slightly difficult, and the farthest (ES↔ZH, ES↔JA) are certainly more difficult. Recent MT approaches (Hybrids, vertical sectors, user customization...) will help to overcome this weakness, and quick Multilingual Publishing Systems can be used with human translation, becoming just “Quick” instead of being “Real-time”.

3) *Opportunities*

a) *Volume*: Affordable for high or very high volume web sites. E.g. sites with more than 1 million words/language.

b) *Update*: Affordable for high or very high frequency and/or volume update web sites. E.g. sites with several updates per day - Profitability Websites with a very high update frequency.

c) *Technological complexity*: Easier for Complex back-office web sites to operate in the final HTML5, and not with the systems that produce it. E.g. virtual offices that “paint” the HTML content from different information systems.

d) *SMEs*: In-house installations are better for sites with over 1 million words and frequently updated, but with shared RTMPS infrastructures, Small and Medium companies and organizations benefit from the low price and low technical effort and complexity.

e) *Recruitment of young translators for new jobs*. E.g. set up and project preparation, post-editing, terminology, MT customization...

f) *New methodologies and methods*. E.g. EDI-TA methodology and training.

g) *New Language Technologies*: Pre-editing and post-editing tools.

h) *New standards*: Interoperability and readiness information.

4) *Threats*

a) *Real-time performance*: The performance of the Multilingual Publishing System and the online Machine Translation all together must stay within the parameters for Internet publication (less than 500 milliseconds). Any additional feature or MT system needs to take this into account.

b) *Security levels*: RTMPS must consider security features, especially in the shared application service provider mode for SMEs.

c) *Robustness*: In-house hosting needs stable and robust systems that provide a service level agreement >99.9%.

d) *Control*: The client might lose control of the translation on the HTML. This can be solved using ITS 2.0 adequately.

e) *Methodologies and tools*: ITS 2.0 and HTML5 compliant pre-editing and post-editing methodologies and tools are needed.

f) *Annotation of source content*: This is probably the greatest threat, convincing content creators about the benefits of annotating the source content and providing them with suitable tools for automatic, semiautomatic, and manual annotation in HTML5.

## VI. CONCLUSIONS

The most evident ITS 2.0 benefit is that it increases both user’s control and automatic decision processes, allowing more “intelligence “ in the systems involved.

1) *Translate*: Translatability control from data. E.g. it allows to add “non-translatable” terms to be used by several specific glossaries or MT systems.

2) *Localization Note*: Direct communication between webmasters, PMs and translators. When *alert* type, it can be used for triggering certain processes in the Translation Workflow. Activation rules for MT post-editing.

3) *Domain*: Automatic selection of CAT/MT terminology, dictionaries, and translation memories.

4) *Language Information*: Quality checks to ensure the content’s source language or a part.

5) *Allowed Characters*: Quality check for the target content.

6) *Storage Size*: Quality check for both original content and target content. Can also be used for translators’ visual control.

7) *Provenance*: Identification of agents, possibility to reassign the same translator/reviewer in new versions, and inform the Project Manager. Tracking control in the CMS.

8) *Localization Quality Issue*: Quality aspects reported to translation consumer or post-editor.

9) *MT Confidence*: Post-editors judge quality of translation.

10) *Readiness (ITS 2.0 extension)*: Control of processes to be done, date control for availability, delivery, and priority.

We can conclude that ITS 2.0 opens up ways of win-win business. It can accelerate the adoption of interoperability localization chains for new users, and resolve certain problem and generate quality and control improvements to existing users:

- 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.
- Reduction of time by increasing the efficiency of the process.
- Cost savings in management and translation.

Business opportunities exist in very frequently updated web sites which need efficient multilingual updates and maximum control, such as corporate and industrial information, e-Government, e-Commerce or Educational web sites. Also, ITS 2.0 can benefit those environments with highly distributed content creation through the CMS, the Web 2.0 and user content created (applying MT systems for immediacy) and using ITS 2.0 to contribute for multilingual SEO.

## VII. ACKNOWLEDGMENT

Many thanks to Olaf-Michael Stefanov for encouraging me to write this article, and to all MLW-LT WG members for their valuable and excellent work to produce ITS 2.0 in 18 months. Special thanks to VDMA and the Spanish Tax Office, and to the colleagues from Cocomore Hans-Ulrich von Freyberg, Karl Fritsche, Clemens, Stephan Walter, and Moritz Hellwig, from Lucy Software Daniel Grasmick and Thomas Ruedesheim, from DCU Ankit Srivastava and Declan Declan Groves and from Linguaserve Giuseppe Deriard, Pablo Badía, Mauricio del Olmo, Pablo Nieto, Laura Guerrero, Consuelo Aldana, Félix Fernández, and many other collaborators who made these use cases possible.

## VIII. REFERENCES

[1] Díez Orzas, P.L., Rico Pérez, C. et al. (2012) EDI-TA: POST-EDITING METHODOLOGY FOR MACHINE TRANSLATION Deliverable D4.1.4 - ANNEX I, public report [online], available: <http://www.w3.org/International/multilingualweb/lt/wiki/Deliverables> [accessed 22 Oct 2013].

[2] Díez Orzas, P.L., Rico Pérez, C. et al. (2012) TRAINING METHODOLOGY FOR MACHINE TRANSLATION POST-EDITING, Deliverable D4.1.4 - ANNEX II, public report [online], available: <http://www.w3.org/International/multilingualweb/lt/wiki/Deliverables> [accessed 22 Oct 2013].

[3] Drupal Community (2013) 'Drupal Official Web' [online ], available: <http://drupal.org/> [accessed 22 Oct 2013]

[4] ITS Interest Group (2013) 'Access to many different presentations and videos on ITS 2.0 and implementations' [online], available: [http://www.w3.org/International/its/wiki/Outreach\\_material](http://www.w3.org/International/its/wiki/Outreach_material) [accessed 22 Oct 2013]

[5] MultilingualWeb-LT Working Group (2013) 'ITS 2.0. 2013. Internationalization Tag Set (ITS) Version 2.0. W3C Last Call Working Draft 24 September 2013' [online], available: <http://www.w3.org/TR/its20/> [accessed 22 Oct 2013]

[6] Nieto Caride, P., Deriard G., Díez Orzas, P.L., Fernández, F., Aldana, C., Badía, P., Srivastava, A., Groves, D., Ruedesheim, T., Díez, R., Crespo, A. (2013) Report on Online MT system. Multilingualweb-LT Deliverable D4.2.2, public report [online], available: <http://www.w3.org/International/multilingualweb/lt/wiki/Deliverables> [accessed 22 Oct 2013].

[7] del Olmo, M., Guerrero, L., Díez Orzas, P.L., Deriard, G., Badía, P., Fritsche, K., Weins, C., Walter S. (2013) B2B integration showcase. Multilingualweb-LT Deliverable D3.2.2, public report [online], available: <http://www.w3.org/International/multilingualweb/lt/wiki/Deliverables> [accessed 22 Oct 2013].

[8] del Olmo, M., Guerrero, L., Díez Orzas, P.L., Deriard, G., Badía, P., Fritsche, K., Weins, C., Walter S. (2013) B2B integration showcase. Multilingualweb-LT Deliverable D3.2.2, public report [online], available: <http://www.w3.org/International/multilingualweb/lt/wiki/Deliverables> [accessed 22 Oct 2013].

[9] Ritchie, P., del Olmo, M., Guerrero, L., Díez Orzas, P.L., Deriard, G., Badía, P., Fritsche, K., Weins, C., Walter S. (2013) Report on Showcases. Multilingualweb-LT Deliverable D3.2.3, public report [online], available: <http://www.w3.org/International/multilingualweb/lt/wiki/Deliverables> [accessed 22 Oct 2013].