Spanish Tax Agency

ITS 2.0 implementation experience in HTML5:

www.agenciatributaria.es

MultilingualWeb Workshop
Making the Multilingual Web Work
Rome, 12–13 March 2013

Román Díez González

Spanish Tax Agency

Pedro L. Díez-Orzas

Linguaserve

Linguaserve collaborators:

Giuseppe Deriard-Nolasco, Pablo Nieto Caride, Consuelo Aldana, Félix Fernández

What are we talking about?

1. Introducing the Spanish Tax Agency
2. www.agenciatributaria.es in the MLW-LT project
3. Shifting to HTML5
4. Experience in ITS2.0 annotation:
   a. Automatic annotation of new ITS2.0 metadata
   b. Reusing custom tags for ITS2.0 metadata annotation
   c. Manual ITS2.0 annotation
5. Next steps and some proposals

(1) Spanish Tax Agency

Spain: General Indicators 2011

ÅSpain is a country regionally structured into 17 autonomous communities and 2 autonomous cities with **5 co-official languages**

- Population : 47.190.493 inhabitants (**12,2 % foreign residents**)

- **Spanish Tax Agency mission**
  - Effective application of Spain’s tax and customs structure
  - Management of tax resources on behalf of other public administrations when ordered by Law or Agreement
- **Overall census of obliged taxpayers**

•Individual taxpayers:	46.509.231
• Companies:	2.674.547
• Other organisations:	2.293.939
Total taxpayers:	**51.477.717**

What are we talking about?

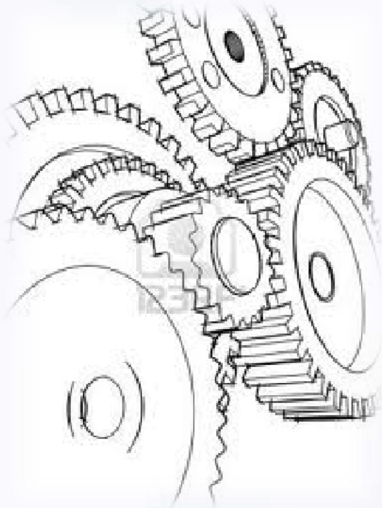
1. Introducing the Spanish Tax Agency
2. www.agenciatributaria.es in the MLW-LT project
3. Shifting to HTML5
4. Experience in ITS2.0 annotation:
   a. Automatic annotation of new ITS2.0 metadata
   b. Reusing custom tags for ITS2.0 metadata annotation
   c. Manual ITS2.0 annotation
5. Next steps and some proposals based on experience

(2) The Spanish Tax Agency in MLW-LT

- **www.agenciatributaria.es**, user in the “Online MT System” use case in the MultilingualWeb-LT (MLW-LT).

- The MLW-LT Working Group is administered by W3C and receives EC funding (LT-Web) through FP7 in the area of Language Technologies

(2) The Spanish Tax Agency in MLW-LT

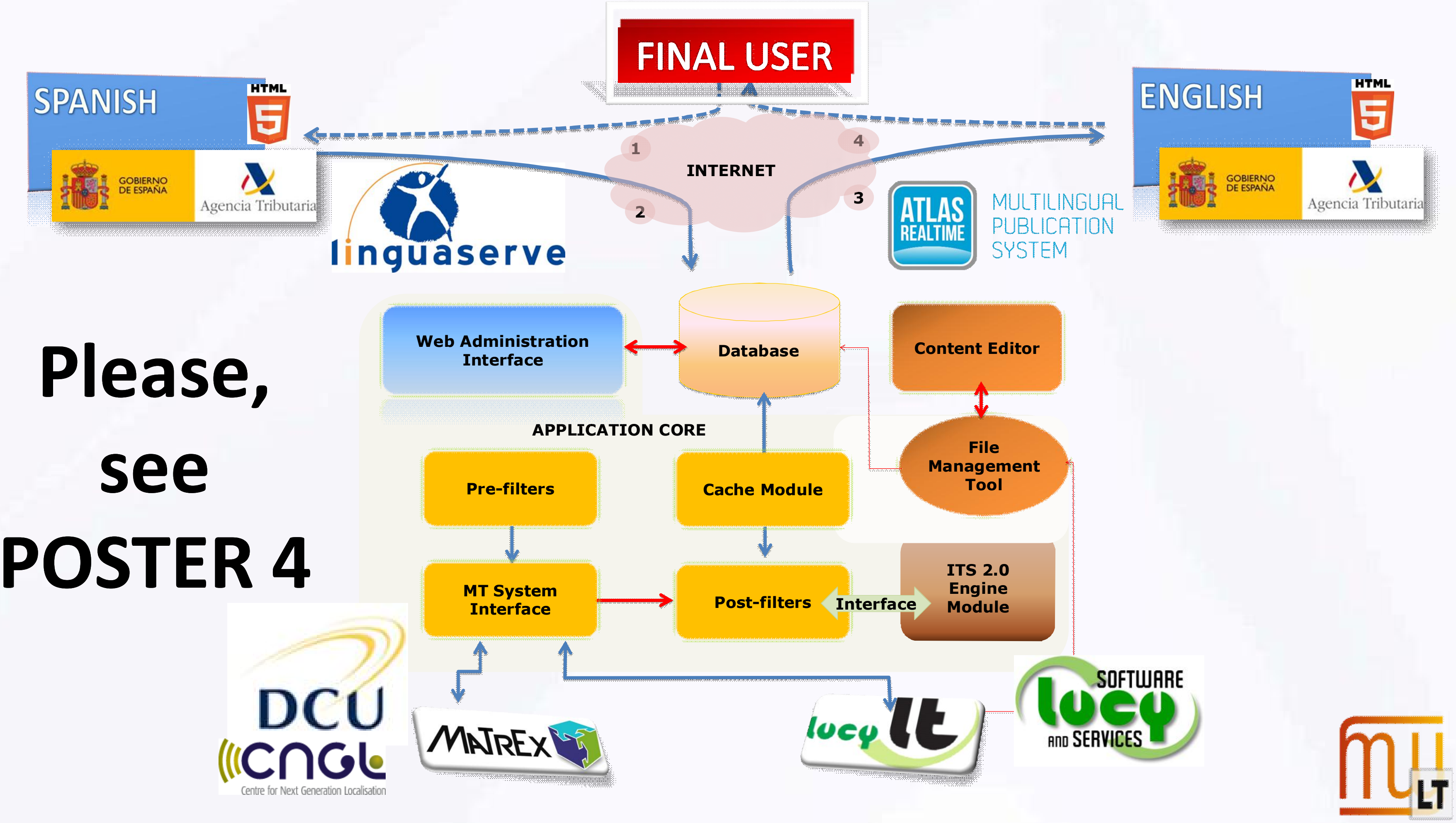


- Online MT System use case components:
  - Multilingual www.agenciatributaria.es (CMS: OpenText WEM)
  - HTML5
  - ITS 2.0
  - Real-time Multilingual Publication System
    - ATLAS (Linguaserve’s Real Time Translation System)
    - Lucy Software MT (Rule-based Machine Translation)
    - MaTrEx from Dublin City University (Statistical Machine Translation)

(2) Online MT System Use Case State

- RTMPS Implementation
  - Prototype 100% (ITS 2.0 definition from Dec 2012)
  - Showcase: preproduction demo (http://its2-aeat.linguaserve.net)
  - ITS 2.0 data categories: 6 (Translate, Localization Note, Language Information, Domain, Provenance, Localization Quality Issue)
- ES-EN total scope: 250 web pages. State:
  - Source language: 30% of target
  - Target language and Post-editing: 30% of target
- ES-FR, ES-DE total scope: 30 web pages. State:
  - Source language: 50% of target
  - Target language and Post-editing: 50% of target
- Testing: pending

(2) Online MT System I18N



Please, see POSTER 4

(2) MLW-LT Online MT SWOT

Strengths

RTMPS highly reduces:

ÅTranslation costs (Quality on-demand)

ï MT + depending on % of post-editing cost reduction increases.

ÅManagement costs

ÅDelivery time

ÅNon-invasive technology

Weaknesses

Viability dependent on :

ÅLanguage combination

ÅMT system output

ÅPre-editing and post-editing methodologies and tools (ITS 2.0 and HTML5 compliance)

Opportunities

Profitability:

ÅWebsites with more than half a million words

ÅWebsites with a very high update frequency

Threats

Control, performance and security:

ÅThe client might lose control of the translation → userôs control with ITS 2.0

ÅReal-time performance

ÅSecurity level

(2) ITS 2.0 benefits for the Spanish Tax Agency

- ITS 2.0 Increases user’s control and automatic decision processes:
  - Translatability and language pair selection (Translate, Language information)
  - Specific terminology to apply (Domain)
  - Activation rules for 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)

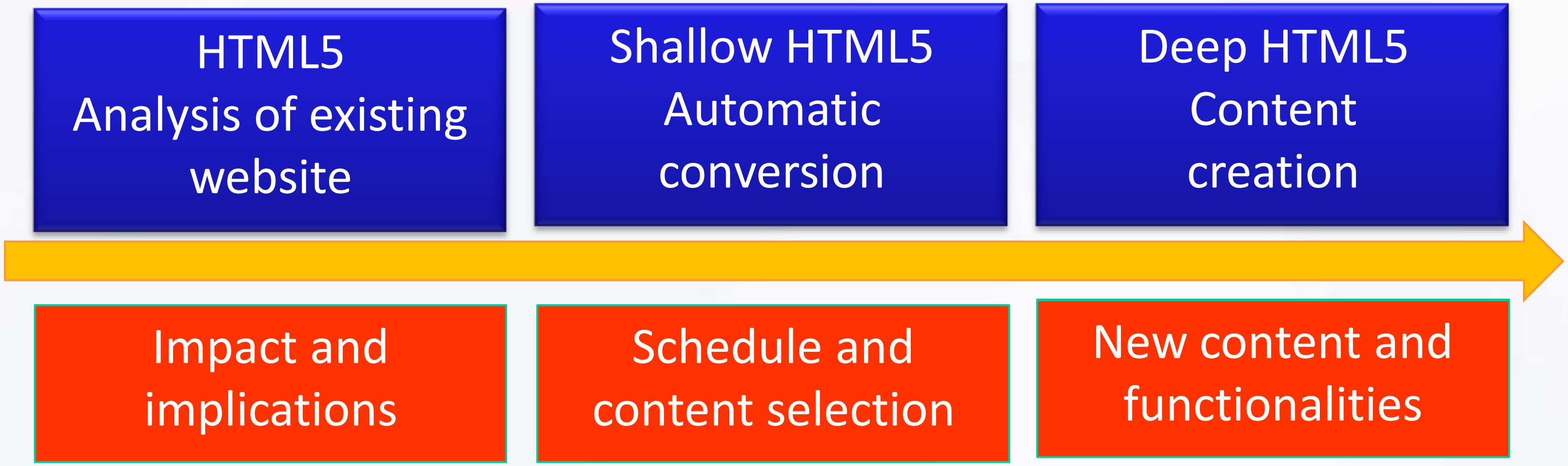
ITS 2.0

What are we talking about?

1. Introducing the Spanish Tax Agency
2. www.agenciatributaria.es in the MLW-LT project
3. Shifting to HTML5
4. Experience in ITS2.0 annotation:
   a. Automatic annotation of new ITS2.0 metadata
   b. Reusing custom tags for ITS2.0 metadata annotation
   c. Manual ITS2.0 annotation
5. Next steps and some proposals based on experience

(3) Shifting to HTML5: Strategy

- Using ITS 2.0 requires HTML version 5 according to the current W3C specification.



HTML5
Analysis of existing website

Shallow HTML5
Automatic conversion

Deep HTML5
Content creation

Impact and implications

Schedule and content selection

New content and functionalities

(3) Shifting to shallow HTML5: Modifications

- HTML5 DOCTYPE
- The language page (ISO 639-ISO 3166)
- Self-closed tags not allowed
- Head tags
- Erroneous nesting tags
- Attributes separated by spaces
- Non inclusion of presentation attributes in tags
- Header and body structure needed by tables
- HTML entities instead of special characters
- URLs cannot contain special characters
- ID attribute cannot contain spaces
- Required attributes (e.g. tag "object" must always have the attributes "data" and "type")
- Assessed attributes (e.g. "rel" attribute of tags "a" and "link" must be one from a closed list)

(3) Shifting to shallow HTML5: Obsolete attributes

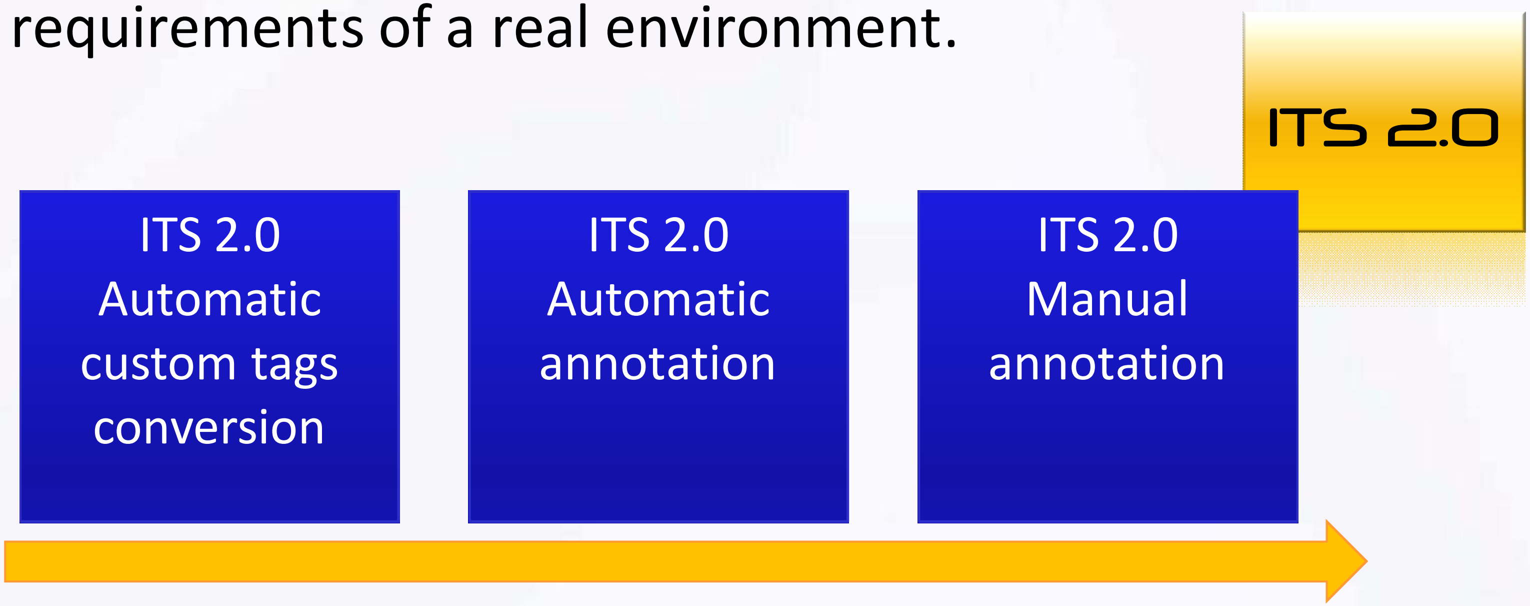
Tags	Impact
input	Removed the alt attribute from any input tag that does not contain the attribute "type = 'image'"
div	Cannot define a "name" attribute in a "DIV“ tag
a	Not allowed to define the attributes "name" and "title" in tag "a"
embed and object	Cannot define the attributes: •"Applet" in the "embed" and "object“ tags •"Name" in the "embed“ tag •"Code", "archive", "classid", "codebase", "codetype", "state“ and "standby" in the "object“ tag
table	Not allowed to define the attributes "summary" and "border" in the "table“ tag
img	Not allowed to define the attributes "name" and "border" in the "img“ tag
option	Cannot define the attribute "name" in the "option“ tag.
param	Not allowed to define the attributes "type“ and "valuetype" in the "param“ tag
script	Not allowed to define the attribute “lang” except in “JavaScript", it being case-insensitive in the tag "script"
br	Cannot define the attribute “clear” in the “br” tag
background attribute	No attribute is used to define the "background" in the tags "body", "table", "thead", "tbody", "tfoot", "tr", "td" and "th".

What are we talking about?

1. Introducing the Spanish Tax Agency
2. www.agenciatributaria.es in the MLW-LT project
3. Shifting to HTML5
4. Experience in ITS2.0 annotation:
   a. Automatic annotation of new ITS2.0 metadata
   b. Reusing custom tags for ITS2.0 metadata annotation
   c. Manual ITS2.0 annotation
5. Next steps and some proposals based on experience

(4) ITS2.0 annotation experience

- Strategy adopted in order to annotate the content with ITS2.0 in an efficient and pragmatic way, considering the pressure and requirements of a real environment.



ITS 2.0

ITS 2.0 Automatic custom tags conversion

ITS 2.0 Automatic annotation

ITS 2.0 Manual annotation

(4) Automatic ITS2.0 reuse of custom tags

Å Custom ñno translateò tag already exists in the content and is automatically annotated as ITS 2.0 *Translate* data category:

*<li>**<!--ATLASP1NOTRAD-->**<a target="_blank"
href="http://www.boe.es/diario_boe/txt.php?id=BOE-A-2011-20472">Orden
EHA/3552/2011, de 19 de diciembre [**é**] **<!--/ATLASP1NOTRAD-->**</li>*

*<li><a **translate=ònoò** target="_blank"
href="http://www.boe.es/diario_boe/txt.php?id=BOE-A-2011-20472">Orden
EHA/3552/2011, de 19 de diciembre [**é**] </li>*

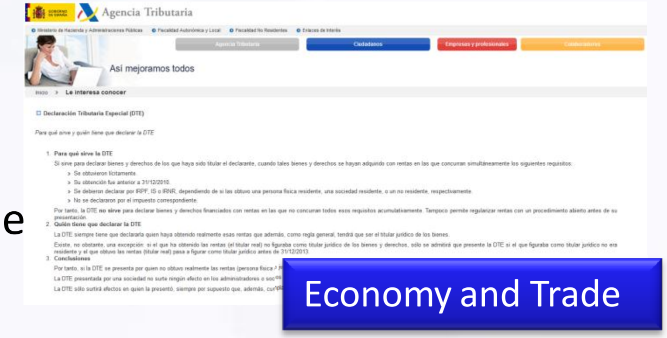
*Respecting the behaviour of the previous tag and the precedence rules of ITS:

ï Addition of ITS default rules for known translatable attributes:

ï <its:translateRule selector="//h:*/@title" translate="yes"/>

ï <its:translateRule selector="//h:*/@alt" translate="yes"/>

(4) Automatic ITS 2.0 annotation: Domain



Agencia Tributaria

Declaración Tributaria Especial (DTE)

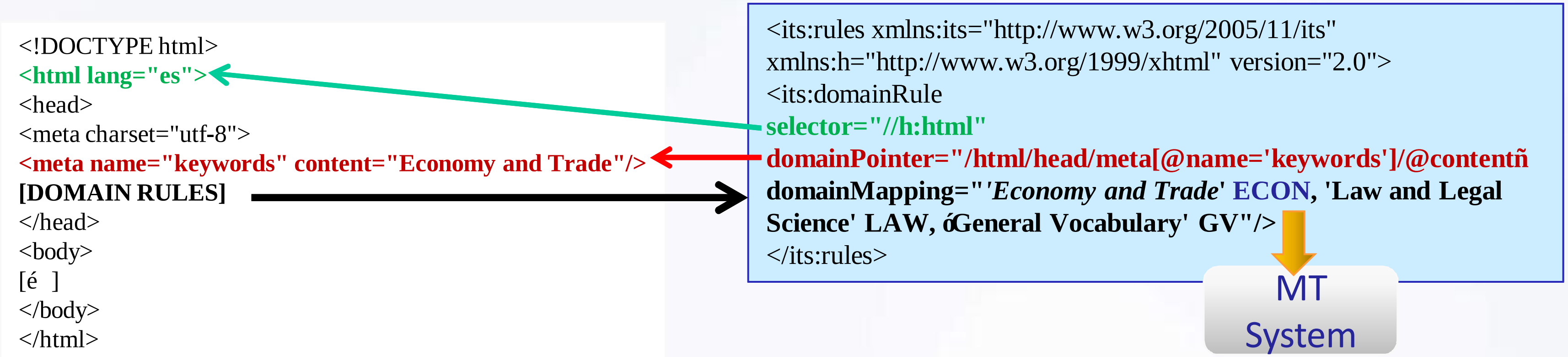
Para qué sirve y quién tiene que declarar la DTE

1. **Para qué sirve la DTE**
   Sí sirve para declarar bienes y derechos de los que haya sido titular el declarante, cuando tales bienes y derechos se hayan adquirido con rentas en las que concurran simultáneamente los siguientes requisitos:
   - Se obtuvieron lícitamente.
   - Su obtención fue anterior a 31/12/2010.
   - Se debieron declarar por IRPF, IS o IRNR, dependiendo de si las obtuvo una persona física residente, una sociedad residente, o un no residente, respectivamente.
   - No se declararon por el impuesto correspondiente.

   Por tanto, la DTE **no sirve** para declarar bienes y derechos financiados con rentas en las que no concurran todos esos requisitos acumulativamente. Tampoco permite regularizar rentas con un procedimiento abierto antes de su presentación.
2. **Quién tiene que declarar la DTE**
   La DTE siempre tiene que declararla quien haya obtenido realmente esas rentas que además, como regla general, tendrá que ser el titular jurídico de los bienes.
   Existe, no obstante, una excepción: si el que ha obtenido las rentas (el titular real) no figuraba como titular jurídico de los bienes y derechos, sólo se admitirá que presente la DTE si el que figuraba como titular jurídico no era residente y el que obtuvo las rentas (titular real) pasa a figurar como titular jurídico antes de 31/12/2013.
3. **Conclusiones**
   Por tanto, si la DTE se presenta por quien no obtuvo realmente las rentas (persona física
   La DTE presentada por una sociedad no surte ningún efecto en los administradores o soc
   La DTE sólo surtirá efectos en quien la presentó, siempre por supuesto que, además, cu

Economy and Trade

1. Extracting relevant domains based on the content.
2. Alignment of the domains with each web page.
3. Use of scripts and regular expressions to annotate the content.
4. Document processing:
   i. The selector points to the html root element, indicating that the domain applies to the whole **HTML** document (inheritance).
   ii. The **domainPointer** attribute indicates where the domain that applies to the selected content is ("**Economy and Trade**").
   iii. The **domainMapping** maps the domain "Economy and Trade" to "**ECON**", which will be sent as an understandable parameter to the MT System.



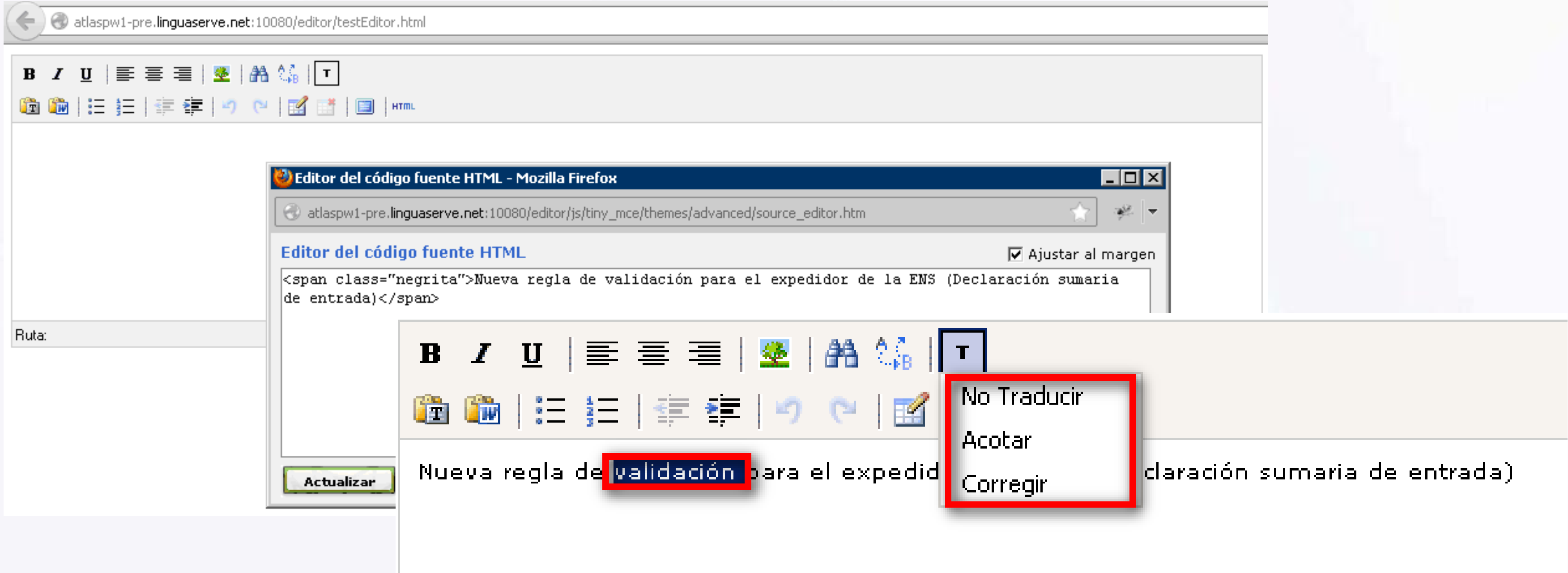
```
<!DOCTYPE html>
<html lang="es">
<head>
<meta charset="utf-8">
<meta name="keywords" content="Economy and Trade"/>
[DOMAIN RULES]
</head>
<body>
[é ]
</body>
</html>
```

```
<its:rules xmlns:its="http://www.w3.org/2005/11/its"
xmlns:h="http://www.w3.org/1999/xhtml" version="2.0">
<its:domainRule
selector="//h:html"
domainPointer="/html/head/meta[@name='keywords']/@contentñ
domainMapping="'Economy and Trade' ECON, 'Law and Legal
Science' LAW, óGeneral Vocabulary' GV"/>
</its:rules>
```

MT System

(4) Manual ITS2.0 annotation: Tool

- Quick and pragmatic approach:
  - New HTML Editor plugin created for the ITS 2.0 manual annotation for open source HTML Editor
  - User-friendly interface for the manual insertion of tags.



atlaspw1-pre.linguaserve.net:10080/editor/testEditor.html

Editor del código fuente HTML - Mozilla Firefox

atlaspw1-pre.linguaserve.net:10080/editor/js/tiny_mce/themes/advanced/source_editor.htm

Editor del código fuente HTML

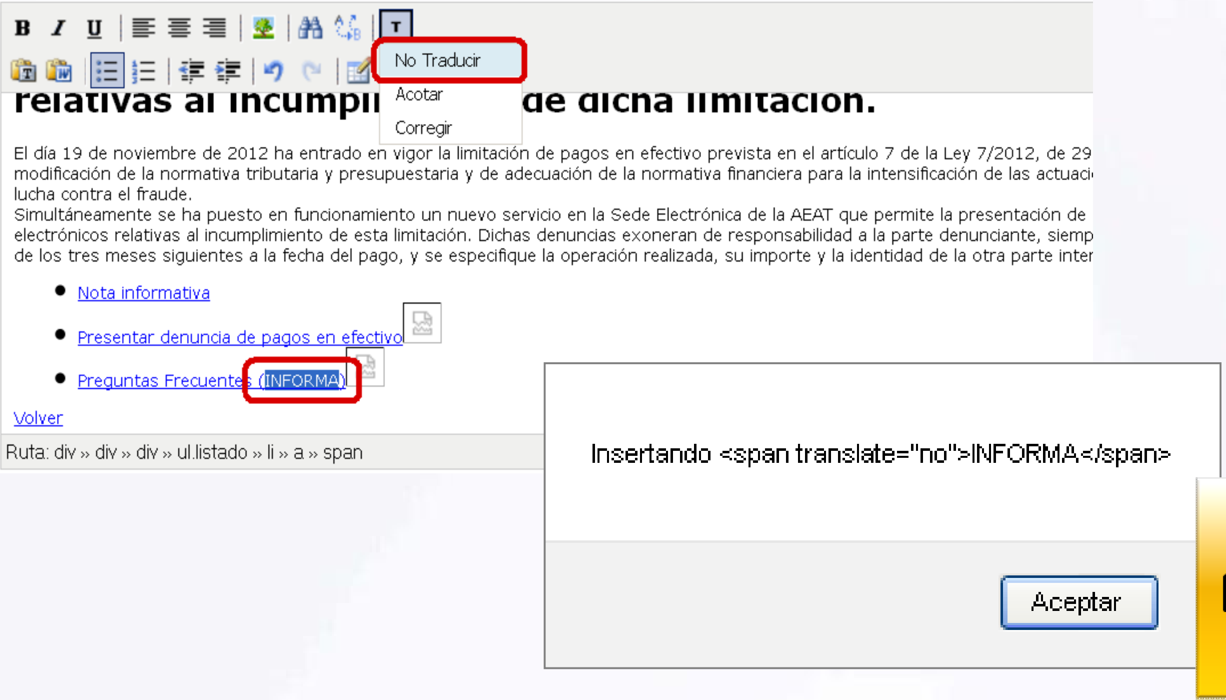
☑ Ajustar al margen

```
<span class="negrita">Nueva regla de validación para el expedidor de la ENS (Declaración sumaria
de entrada)</span>
```

Nueva regla de validación para el expedid... claración sumaria de entrada)

(4) ITS 2.0 Manual annotation: Translate

- The author must only select the non-translatable element, click on the insertion icon (T) and click on the annotation type: ***No Traducir***.



relativas al incumpli de dicha limitación.

El día 19 de noviembre de 2012 ha entrado en vigor la limitación de pagos en efectivo prevista en el artículo 7 de la Ley 7/2012, de 29 modificación de la normativa tributaria y presupuestaria y de adecuación de la normativa financiera para la intensificación de las actuaci lucha contra el fraude.
Simultáneamente se ha puesto en funcionamiento un nuevo servicio en la Sede Electrónica de la AEAT que permite la presentación de electrónicos relativas al incumplimiento de esta limitación. Dichas denuncias exoneran de responsabilidad a la parte denunciante, siemp de los tres meses siguientes a la fecha del pago, y se especifique la operación realizada, su importe y la identidad de la otra parte inter

- Nota informativa
- Presentar denuncia de pagos en efectivo
- Preguntas Frecuentes (INFORMA)

Volver

Ruta: div » div » div » ul.listado » li » a » span

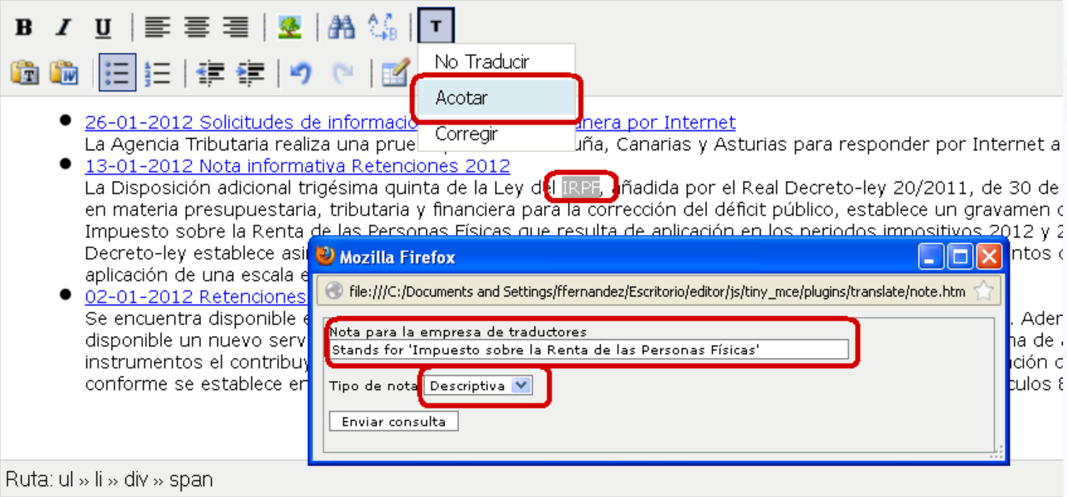
Insertando <span translate="no">INFORMA</span>

ITS 2.0

(4) ITS 2.0 Manual annotation: Localization Notes

- Use of the annotation type ***Acotar***: The author inserts the annotation text into the box and the software will automatically create the tag.

Å The pull-down menu is used to choose the type of localization note. It can either be <u>description</u> (*descriptiva*) or <u>alert</u> (*alerta*).



- 26-01-2012 Solicitudes de informació… …nera por Internet
  La Agencia Tributaria realiza una prue… …uña, Canarias y Asturias para responder por Internet a
- 13-01-2012 Nota informativa Retenciones 2012
  La Disposición adicional trigésima quinta de la Ley del IRPF, añadida por el Real Decreto-ley 20/2011, de 30 de en materia presupuestaria, tributaria y financiera para la corrección del déficit público, establece un gravamen c Impuesto sobre la Renta de las Personas Físicas que resulta de aplicación en los periodos impositivos 2012 y 2 Decreto-ley establece asi… …ntos c aplicación de una escala e…
- 02-01-2012 Retenciones…
  Se encuentra disponible e… …. Ader disponible un nuevo serv… …na de … instrumentos el contribuy… …ción c conforme se establece en… …culos 8

file:///C:/Documents and Settings/ffernandez/Escritorio/editor/js/tiny_mce/plugins/translate/note.htm

Nota para la empresa de traductores: Stands for 'Impuesto sobre la Renta de las Personas Físicas'

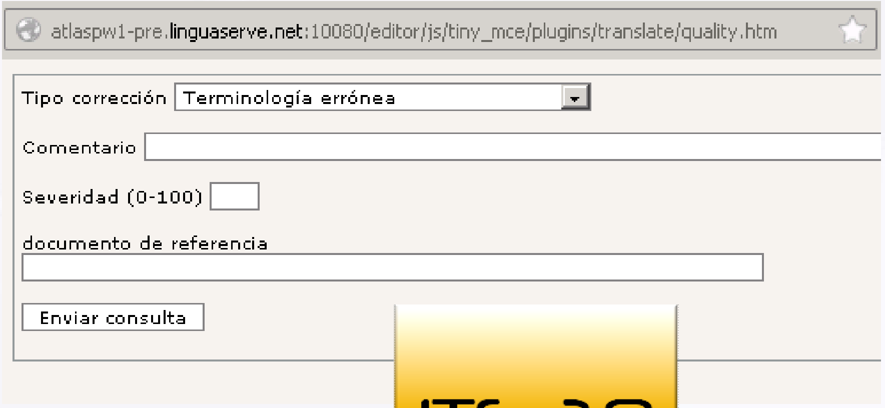
Tipo de nota: Descriptiva

Ruta: ul » li » div » span

```
<p>La disposición trigésima quinta de la Ley del <span its-loc-note="Stands for 'Impuesto sobre la Renta de las Personas Físicas ', use acronym in target language" its-loc-note-type="description">IRPF</span></p>
```

(4) ITS 2.0 Manual annotation: Localization Quality Issue

- Use of the annotation type ***Corregir***: The author chooses a type of issue from a pull-down menu, inserts a comment into the box (*Comentario*), chooses a severity level between 0 and 100 (*Severidad*) and an optional link to a reference document (*documento de referencia*), and the software will automatically create the tag.


atlaspw1-pre.linguaserve.net:10080/editor/js/tiny_mce/plugins/translate/quality.htm

Tipo corrección: Terminología errónea

Comentario:

Severidad (0-100):

documento de referencia:

Online filing can be done by the interested party or by someone representing them. In both cases, an electronic certificate X.509.V3 issued by the <span its-loc-quality-issue-comment="**Has previously been translated as 'Royal Mint'. Please be consistent.**" its-loc-quality-issue-type="inconsistency" its-loc-quality-issue-severity="70">**National Coin and Stamp Factory**</span>.

What are we talking about?

1. Introducing the Spanish Tax Agency
2. www.agenciatributaria.es in the MLW-LT project
3. Shifting to HTML5
4. Experience in ITS2.0 annotation:
   a. Automatic annotation of new ITS2.0 metadata
   b. Reusing custom tags for ITS2.0 metadata annotation
   c. Manual ITS2.0 annotation
5. Next steps and some proposals

(5) Next steps and some proposals

- End of Online Translation System MLW-LT use case – June 2013
- Exploring best practices using ITS 2.0 data categories
- Improving real-time translation and multilingual publishing processing by applying extensions, e.g. Readiness:
  - ITS 2.0 extension data category proposal.
  - Linguaserve is applying Readines in both use cases involved:
    - Applied in CMS-TMS showcase (WP3, poster 3)
    - Applicability in Online Translation system (WP4)
  - It indicates the readiness of a document for submission to L10n processes or provides an estimate of when it will be ready for a particular process.
  - It can be used in expert systems for automatic processing.

ITS 2.0

(5) Next steps and some proposals

- Training and methodologies
  - Pre-editing: ITS2.0 usage and training kits.
  - EDI-TA: Post-editing contextual, activation and identification rules.
- Specific tools
  - Pre-editing:
    - Full HTML5 compliance and ITS2.0 annotation facilities
    - Writing tools for content quality, and controlled language for post-editing output adaptation.
  - Post-editing:
    - Specific language-dependent and language-independent post-editing rules and functionalities.
    - ITS 2.0 assistance and viewing functions for post-editors.

Agencia Tributaria

MultilingualWeb Workshop
Making the Multilingual Web Work
Rome, 12–13 March 2013

www.agenciatributaria.es