The TRANSBey Prototype: An Online Collaborative Wiki-Based CAT Environment for Volunteer Translators

Youcef Bey[1,2], Christian Boitet[2], and Kyo Kageura[3]

[1]Graduate School of Advanced Studies, NII.
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo,
101-8430, Japan
youcefb@grad.nii.ac.jp

[2]Laboratoire CLIPS-GETA-IMAG
Université Joseph Fourier
385, rue de la Bibliothèque.
Grenoble, France
Christian.boitet@imag.fr

[3]Graduate School of Education
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku.
Tokyo, 113-0033, Japan
kyo@p.u-tokyo.ac.jp

Abstract

The aim of our research is to design and develop a new online collaborative translation environment suitable for the way in which the online volunteer translators work. In this paper, we discuss thus how to exploit collaborative Wiki-based technology for the design of the online collaborative computer-aided translation (CAT) environment TRANSBey, which is currently under development. The system maximizes the facilitation of managing and using existing translation resources and fills the gap between the requirements of online volunteer translator communities and existing CAT systems/tools.

1. Introduction

In accordance with the current global exchange of information in various languages, we are witnessing a rapid growth in the activities of online volunteer translators, who individually or collectively make important documents available online in various languages. Two major types of online volunteer translator communities can be identified (Bey, 2005):

(i) *Mission-oriented translator communities:* mission-oriented, strongly-coordinated groups of volunteers are involved in translating clearly defined sets of documents. Many such communities translate technical documentation of project like Linux documentation (Traduct, 2005), W3C (W3C, 2005), and Mozilla (Mozilla, 2005).

(ii) *Subject-oriented translator network communities:* individual translators who translate online documents such as news, analyses, and reports and make translations available on personal or group web pages. These groups of translators do not have any orientation in advance, but they share similar opinions about events (anti-war humanitarian communities, report translation, news translation, humanitarian help, etc.) (TeaNotWar, 2005).

The aim of our research is to design and develop a new online collaborative translation environment suitable for the way in which these online translators work, with a special focus on mission-oriented translator communities, but also taking into account the needs of subject-oriented individual translators. To achieve this aim, we decided to use a Wiki-system as a base technology for developing an online collaborative translation environment that facilitates the management and use of documents and linguistic reference resources. This paper discusses the basic requirements of online translators working in a collaborative environment and reports the system functions developed for satisfying these needs within the TRANSBey system that we are currently developing.

We introduce general volunteer translators' needs by describing and analyzing various existing communities. In the second section, we attempt to outline the basic functionalities of online Wiki technology and its advantages for constructing of an online computer-aided translation environment (CAT). In the last section, the main modules of TRANSBey are described.

2. Current stat of online volunteer translators

Many translator communities are currently involved in translating various types of documents in different formats. In W3C, 301 volunteer translators translate specification documents (XML, HTML, Web service, etc.) into 44 languages. Paxhumana (PaxHumana, 2006) is another community of volunteer translators who translate report documents into four languages (English, Spanish, French, German). The twos groups show basically the same behavior during the translation process. In general, translation is done using a stand-alone personal environments. In the process, translators do not use linguistic tools on the server from which they disseminate translated documents (Bey, 2005). They communicate with each other to avoid duplicate translations.

The function of this translation processes currently not only falls short of what can be achieved using current technology but also does not satisfy translators' potential needs. The major insufficiencies, among others, of existing collaborative translation environments are that (i) different file formats (e.g., DOC, HTML, PDF, XML) cannot be automatically dealt with in the translation environment, (ii) existing translated document pairs cannot be efficiently and systematically looked up within the overall community environment in the process of translating new related documents, and (iii) linguistic tools are not sufficiently provided. Existing CAT systems, on the other hand, do not address fully the functions required by collaborative environments. As stated, our aim is to develop a system that can fill the gap between volunteer translators requirements and the existing community-based translation environment as well as CAT systems.

The insufficiencies above can be filled by functions/modules that (i) unify and consistently manage document formats and versions so that they can not only be consistently administered but also can be processed into recyclable units for future reference as translation memory (TM), (ii) integrate the rich online wysiwyg editing environment for direct document creation on the server with various linguistic reference lookup functions, and (iii) support multilingual content. If these functions/modules are integrated into (iv) basic online community management mechanisms, we would be able to further promote the activities of mission-oriented translator communities. Let us elaborate these points in the following paragraphs:

(i) Combining TM with document management

Most existing TM systems provide little or no support for document management and versioning (Bowker, 2002). However, keeping information or traces from original documents translated sentences is useful when translators look for the context of translation and could at least allow them to reconstruct documents from translated segments in TM. Translators would be able to use TM to search general information related to the context of documents (e.g., to find the latest text translated for a particular organization). We have adopted translation memory exchange (TMX) to support document structure and TM exchange. For unit detection, we have exploited the efficiency of LingPipe tools (LingPipe, 2006), witch deal with sentence-boundary detection and linguistic unit detection (e.g, named-entity detection). This tool can be extended to support additional languages and trained resources for more precision.

(ii) In-browser wysiwyg editor

Translators have shown interest in developing online translation editors that would allow multiple translators to share TMs and documents for translation. This is particularly appealing to freelance translators and useful for sharing translation.

(iii) Multilingual content support

Another improvements that is underway is extending CAT tools to support a wider variety of languages by using encoding methods such as Unicode (UTF-8) and designing new standards and filters to support a wider variety of file formats, including formats using tags (e.g., HTML and XML).

(iv) Collaboration for enhancing translation

In terms of more general developments, the current movement away from stand-alone systems and toward online environments which facilitate networking is likely to continue (Bowker, 2002), thus making it possible for multiple users to share the same TM, translation and various type of linguistic resources.

We explained the principal need of volunteer translators in the above sections, witch leads us to underline the basic and relevant functionalities of Wiki technology and its advantages and facilities for the overall design of TRANSBey in the next section.

3. Basic Wiki functionalities for online CAT environment

A Wiki environments allow users to freely create and edit web page content using any web browser. On the one hand, they have simple syntax for creating new pages and links between internal pages, and on the other hand, they allow the organization of contributions to be edited in addition to the content itself. Augar stated (Augar, 2004):

"A Wiki is a freely expandable collection of interlinked web pages, a hypertext system for storing and modifying information– a database, where each page is easily edited by any user with a forms-capable web browser client".

Browser-based access means that neither special software nor a third-party webmaster is needed to post content. Content is posted immediately, eliminating the need for distribution. Participants can be notified about new content, and they review only new content. Access is flexible. In fact, all that is needed is a computer with a browser and Internet connection (Schwartz, 2004).

The most important feature of a Wiki technology is the open editing, which means that content is open for direct editing and direct dissemination of information. Among existing open Wiki environments, we have chosen XWiki, a Java-based environment with the following features (XWiki, 2006):

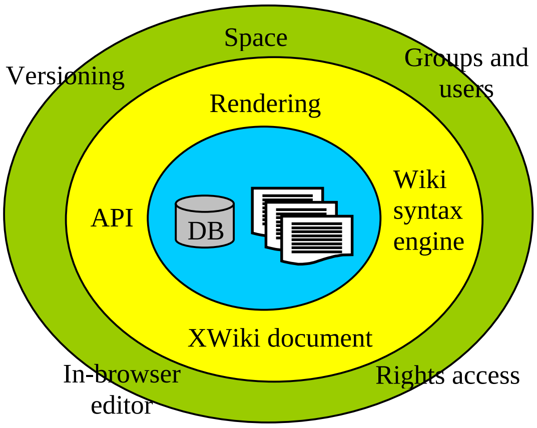


Figure 1: Main layers in XWiki

(i) Document management

Documents are managed in the core of an internal database where each entry is a document, which in the overall function of the environment is considered as the principal component or entries in the central databases. Documents are managed under the control of space and have a set of access rights for users. Documents can contain not only information to display (in HTML) but also Wiki syntax and programming codes for extending the system functionalities (Schwartz, 2004).

(ii) Versioning

Documents in the XWiki environment are supposed to be modified by direct editing. For any modification of the content, the system stores a new version in a new document and saves the old version for possible comparison or reuse.

(iii) Multilingual support

Translators are supposed to deal with content in several languages. Allowing the dissemination of translations on the web in several languages will motivate translator communities to use easily XWiki environment.

(iv) Space concept for volunteer communities

Organizing documents in space is very useful for communities and communication. Indeed, volunteers often work in an identified space, which they use it for sharing documents.

(v) Group (users) and access Rights

Access to XWiki content can be controlled. Users are identified via their IP addresses, and access can be limited for specific documents and functionalities.

Note that these features provide us with basic environment for collaborative translator communities. Within this overall environment, we have developed functions and modules specifically for translators, to which we now turn.

4. The TRANSBey prototype: integration of documents management and TM

4.1. Importing and processing source documents

Under the control of the uploader module that we have added to XWiki, documents (from source documents) can be uploaded directly into a unified format from various format types (PDF, DOC, RTF, HTML, etc.) or copied to source text areas in the in-browser editor (Figure 2). They are then stored in a unified format for document management with proper segmentation for recycling useful reference units.

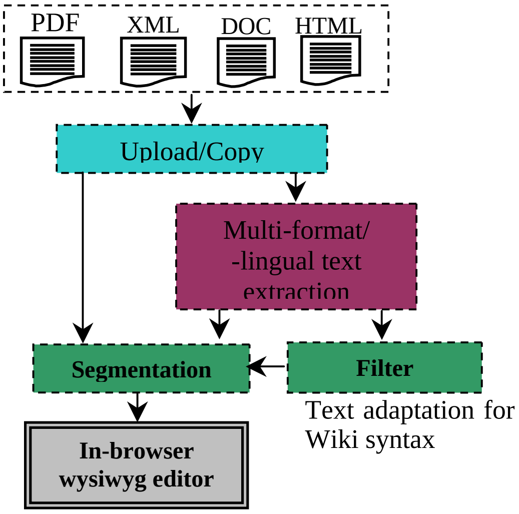


Figure 2: Monolingual document importation.

The extracted texts are segmented to logical translatable and linguistic units. The segmentation process is done semi-automatically[1], and translatable units are defined in the cores of textual sources.

4.2. TMX-C format for TM management

For the integrated management of documents and TM recycled from the documents, the document data structure should satisfy two requirements: (a) maximal facilitation of providing recyclable units and (b) unified management of translated documents. The first requirement come from individual translators, who strongly look for relevant linguistic units (especially collocations and quotations) in existing translations. The second requirement comes from the manager of the community in which translators take part or from the community itself. For this aim, we found that the translation memory exchange (TMX) standard is suitable (LISA, 2006). This standard was developed to simplify the storage/exchange of TM and to facilitate source/target sentences to be stored in a multilingual format in XML format(Bey, 2005)(Boitet, 2005).

Annotation is done in our environment in accordance with the TMX standard[2] and the 3-tiers level model proposed by Saha (Saha, 2005). The 3-tiers model for dealing with various information (Metadata annotation, sentences and linguistic unit annotations) is illustrated in Figure 3. The result of segmentation is an annotated document, which used to auto-construct TM and linguistic resources in the Wiki store.

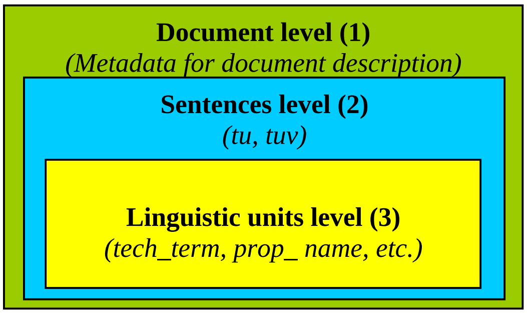
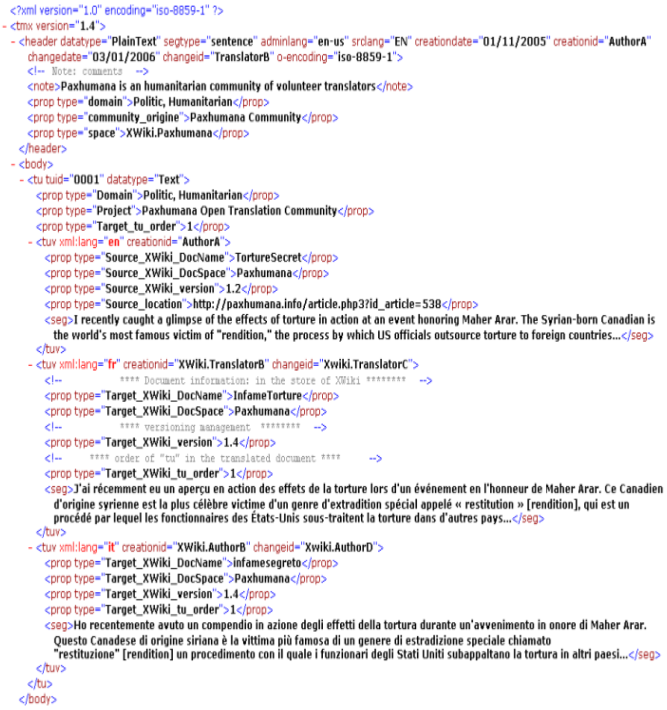

Document level (1)
(Metadata for document description)
Sentences level (2)
(tu, tuv)
Linguistic units level (3)
(tech_term, prop_ name, etc.)

Figure 3: 3-tiers level for document segmentation.

Taking into consideration the advantages of the 3-tiers level model and the TMX standard capabilities, we have proposed TMX for Collaboration (TMX-C), witch is adapted for dealing with three levels during segmentation, for constructing the TM format, and for supporting collaborative Wiki information (Figure 4).



```
  <?xml version="1.0" encoding="iso-8859-1" ?>
- <tmx version="1.4">
 - <header datatype="PlainText" segtype="sentence" adminlang="en-us" srclang="EN" creationdate="01/11/2005" creationid="AuthorA"
     changedate="03/01/2006" changeid="TranslatorB" o-encoding="iso-8859-1">
     <!-- Note: comments   -->
     <note>Paxhumana is an humanitarian community of volunteer translators</note>
     <prop type="domain">Politic, Humanitarian</prop>
     <prop type="community_origine">Paxhumana Community</prop>
     <prop type="space">XWiki.Paxhumana</prop>
   </header>
 - <body>
  - <tu tuid="0001" datatype="Text">
      <prop type="Domain">Politic, Humanitarian</prop>
      <prop type="Project">Paxhumana Open Translation Community</prop>
      <prop type="Target_tu_order">1</prop>
   - <tuv xml:lang="en" creationid="AuthorA">
       <prop type="Source_XWiki_DocName">TortureSecret</prop>
       <prop type="Source_XWiki_DocSpace">Paxhumana</prop>
       <prop type="Source_XWiki_version">1.2</prop>
       <prop type="Source_location">http://paxhumana.info/article.php3?id_article=538</prop>
       <seg>I recently caught a glimpse of the effects of torture in action at an event honoring Maher Arar. The Syrian-born Canadian is
         the world's most famous victim of "rendition," the process by which US officials outsource torture to foreign countries...</seg>
     </tuv>
   - <tuv xml:lang="fr" creationid="XWiki.TranslatorB" changeid="Xwiki.TranslatorC">
       <!--          **** Document information: in the store of XWiki ********   -->
       <prop type="Target_XWiki_DocName">InfameTorture</prop>
       <prop type="Target_XWiki_DocSpace">Paxhumana</prop>
       <!--          **** versioning management  ********   -->
       <prop type="Target_XWiki_version">1.4</prop>
       <!--     **** order of "tu" in the translated document ****      -->
       <prop type="Target_XWiki_tu_order">1</prop>
       <seg>J'ai récemment eu un aperçu en action des effets de la torture lors d'un événement en l'honneur de Maher Arar. Ce Canadien
         d'origine syrienne est la plus célèbre victime d'un genre d'extradition spécial appelé « restitution » [rendition], qui est un
         procédé par lequel les fonctionnaires des États-Unis sous-traitent la torture dans d'autres pays...</seg>
     </tuv>
   - <tuv xml:lang="it" creationid="XWiki.AuthorB" changeid="Xwiki.AuthorD">
       <prop type="Target_XWiki_DocName">infamesegreto</prop>
       <prop type="Target_XWiki_DocSpace">Paxhumana</prop>
       <prop type="Target_XWiki_version">1.4</prop>
       <prop type="Target_XWiki_tu_order">1</prop>
       <seg>Ho recentemente avuto un compendio in azione degli effetti della tortura durante un'avvenimento in onore di Maher Arar.
         Questo Canadese di origine siriana è la vittima più famosa di un genere di estradizione speciale chiamato
         "restituzione" [rendition] un procedimento con il quale i funzionari degli Stati Uniti subappaltano la tortura in altri paesi...</seg>
     </tuv>
    </tu>
   </body>
```

Figure 4: TMX-C format: Collaborative TMX-based for managing documents and TMs (PaxHumana, 2006).

At the top level, document information is provided, which is essential for document management but also useful for translators for checking the context and/or

[1] Translators have the ability in TRANSBey to annotate text in both source and target documents. The process is done (i) automatically by direct detection of translatable units before starting translation and (ii) by translators who delimitate translatable and linguistic units during translation.

[2] A standard proposed by the Localization Industry Standards Association (LISA) communities for TM support, exchange between humans specialists (or software) for more consistency, and decreased data loss.

domain to which documents belong (Table 1). The second and third levels are concerned with language units, i.e., sentences in the second level and various linguistic units (quotations, collocations, technical terms, proper names, idioms, etc.) in the third level (Table2). These units can be automatically detected using sentence-boundary tools (LingPipe, 2006) and other basic language processing tools, but translators can manually control these units in the process of editing and translation. The segments and metadata XML tags are defined as follows:

Metadata	Description
Domain	Domain of document: technical information, medical, personal, sports, humanitarian, etc.
Original_Community	Original community name.
Space	Community space name in the XWiki store.
S/T_XWiki_DocName	Document name in Xwiki.
S/T_XWiki_DocSpace	Space containing the document in XWiki.
S/T_XWiki_version	The version generated by XWiki.
S/T_XWiki_TU_Order	Order of “TUV” in the document XWiki.
Etc.	Etc.

Table 1: Metadata annotation tags.

TU/LU	Description	Format
tech_term	Technical term	XLD
prop_name	Proper name	XLD
Ord_word	Ordinary word	XLD
Quot	Quotation	XLD
Colloc	Collocation	XLD
TU	Translatable unit	TMX
TUV	Translatable unit version[3]	TMX
Etc.	Etc.	Etc.

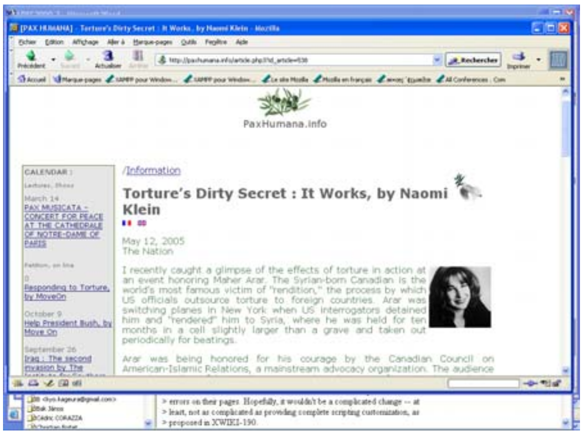
Table 2: Translatable/linguistic unit annotation tags.

5. Online in-browser wysiwyg editor in TRANSBey environment

Editing source and target documents in an enhanced editor is the most important module that translators look for. Offering online editing in TRANSBey means also leading translators to edit in a rich environment that, among other functions, efficiently manage document formats, includes linguistic tools for accelerating translation and increasing quality, and avoids making translators become web developers, which is in general a hard task (which includes editing html code).

Among existing in-browser editors, we have chosen HTMLArea to integrate our environment for its many advantages (HTMLArea, 2006): (i) compatibility with almost all web browser (IE, Mozilla, Firefox) (ii) production of a well-formed HTML code (iii) ease of integration with XWiki for managing Wiki syntax (iv) several wysiwyg editing features (table, images, headings, etc.).

Figure 5 and Figure 6 illustrate how without any effort web documents in English can be imported from their original web sites to the TRANSBey environment for collaborative translation without losing their format and style presentations.



[PAX HUMANA] - Torture's Dirty Secret : It Works, by Naomi Klein - Mozilla

http://paxhumana.info/article.php3?id_article=538

PaxHumana.info

CALENDAR :

Lectures, Shows

March 14

PAX MUSICATA - CONCERT FOR PEACE AT THE CATHEDRALE OF NOTRE-DAME OF PARIS

Petition, on line

Responding to Torture, by MoveOn

October 9

Help President Bush, by Move On

September 26

Iraq : The second invasion by The

/Information

Torture’s Dirty Secret : It Works, by Naomi Klein

May 12, 2005

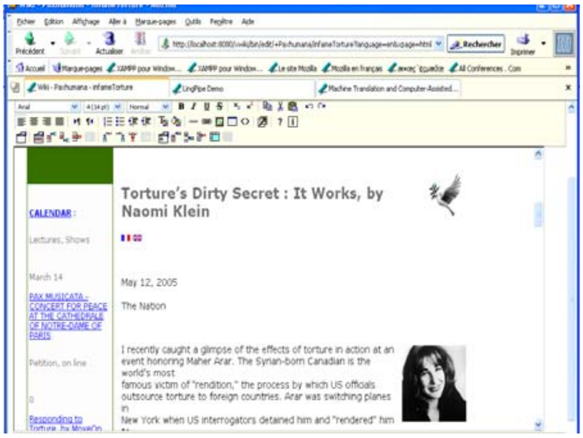
The Nation

I recently caught a glimpse of the effects of torture in action at an event honoring Maher Arar. The Syrian-born Canadian is the world's most famous victim of "rendition," the process by which US officials outsource torture to foreign countries. Arar was switching planes in New York when US interrogators detained him and "rendered" him to Syria, where he was held for ten months in a cell slightly larger than a grave and taken out periodically for beatings.

Arar was being honored for his courage by the Canadian Council on American-Islamic Relations, a mainstream advocacy organization. The audience

> errors on their pages. Hopefully, it wouldn't be a complicated change -- at
> least, not as complicated as providing complete scripting customization, as
> proposed in XWIKI-190.

Figure 5: Source document in its original web site[4].



http://localhost:8080/xwiki/bin/edit/+Paxhumana/infameTorture?language=en&page=html

Wiki - Paxhumana - infameTorture

LingPipe Demo

Machine Translation and Computer-Assisted...

CALENDAR :

Lectures, Shows

March 14

PAX MUSICATA - CONCERT FOR PEACE AT THE CATHEDRALE OF NOTRE-DAME OF PARIS

Petition, on line

Responding to Torture, by MoveOn

Torture’s Dirty Secret : It Works, by Naomi Klein

May 12, 2005

The Nation

I recently caught a glimpse of the effects of torture in action at an event honoring Maher Arar. The Syrian-born Canadian is the world's most
famous victim of "rendition," the process by which US officials outsource torture to foreign countries. Arar was switching planes in
New York when US interrogators detained him and "rendered" him

Figure 6: Imported document in TRANSBey environment[5].

Furthermore, using the integrated in-browser wysiwyg editor allows the same document to be produced in French without any modification of the source format during the translation (figure 7).

This example shows the feasibility of joining volunteer translators in comparable individual

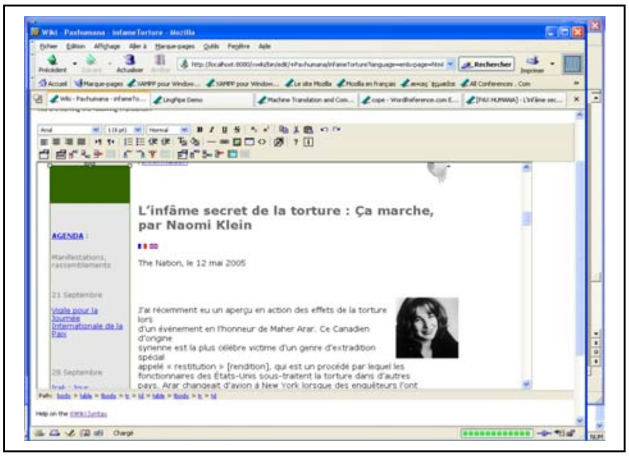
[3]For further information about translation unit (TU) , translation unit version (TUV) refer to TMX standard (LISA, 2006).

[4]http://paxhumana.info/article.php3?id_article=538

[5]http://localhost:8080/xwiki/bin/view/+Paxhumana/TortureDirty (Wiki path for the imported document on the local server)

environments. The environment allows users easy HTML link navigation and gives them enhanced multilingual research functions for easily finding source/target documents and switching directly to the editing wysiwyg environment.

The integrated in-browser editor in XWiki is an open source, witch was developed separately by a group of volunteers called HTMLArea (HTMLArea, 2006). It contain the principal functionalities for editing and visual HTML component design (forms, tables, images, buttons, etc.). Furthermore, it manages well HTML/Wiki tag conversion and is compatible with IE and all Gecko web browsers (MOZILLA, FIREFOX, etc.).


Wiki - Paxhumana - infameTorture - Mozilla
Fichier Édition Affichage Aller à Marque-pages Outils Fenêtre Aide
Précédent Actualiser Rechercher Imprimer
AGENDA :
Manifestations, rassemblements
21 Septembre
Vigile pour la Journée Internationale de la Paix
28 Septembre
Help on the XWiki Syntax
Chargé

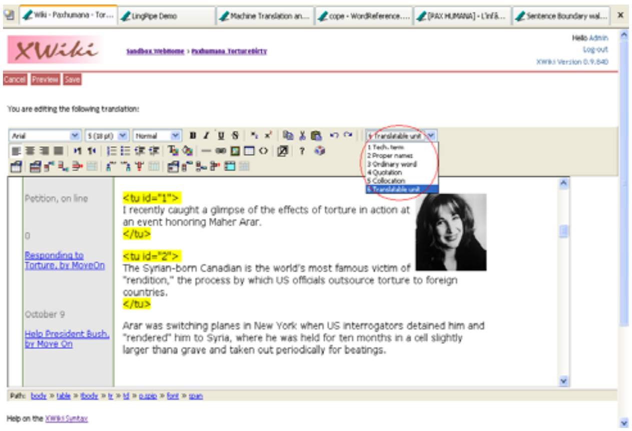
L'infâme secret de la torture : Ça marche, par Naomi Klein

The Nation, le 12 mai 2005

J'ai récemment eu un aperçu en action des effets de la torture lors
d'un événement en l'honneur de Maher Arar. Ce Canadien d'origine
syrienne est la plus célèbre victime d'un genre d'extradition spécial
appelé « restitution » [rendition], qui est un procédé par lequel les
fonctionnaires des États-Unis sous-traitent la torture dans d'autres
pays. Arar changeait d'avion à New York lorsque des enquêteurs l'ont

Figure 7: Target document after translation[6].

The editor, which is integrated into the collaborative environment and whose functions are currently under development, will be able to deal with different source texts in different formats in a unified framework while keeping the original format and can provide translators reference lookup and semi-automatic annotation based on TMX-C.



You are editing the following translation:

<tu id="1">
I recently caught a glimpse of the effects of torture in action at an event honoring Maher Arar.
</tu>

<tu id="2">
The Syrian-born Canadian is the world's most famous victim of "rendition," the process by which US officials outsource torture to foreign countries.
</tu>

Arar was switching planes in New York when US interrogators detained him and "rendered" him to Syria, where he was held for ten months in a cell slightly larger thana grave and taken out periodically for beatings.

Figure 8: Wysiwyg edition and direct source annotation.

After translation is finished, source and target documents are recycled and translated segments and linguistic units are stored in linguistic resources for possible reuse for translating other documents (Figure 8).

6. Conclusion

We have proposed the TRANSBey prototype, an environment for helping volunteer translators produce high quality translations of various types of documents. This environment, which we are developing, will open a way for gathering skills and enhancing quality for all communities involved in translation. On the one hand, we used Wiki technology to exploit collaborative and open editing functionalities on the web; on the other hand, we have integrated the management of translatable units and linguistic resources using annotation system. Our aims for our environment are to offer to online volunteer translators important components for producing a quick translation with high quality in several languages.

In the near future, we are interesting for the enhancement of the integrated online editor for supporting synchronization and semi-automatic alignment between source and target documents for automatic TM construction, and integration during the translation process.

7. Acknowledgements

Special thanks go to Prof. Akiko Aizawa (National Institute of Informatics, NII, Japan) for many most valuable suggestions, advices and her invitation for the international internship at the NII.

8. References

Augar, N. , Raitman, R. and Zhou, W. (2004). Teaching and Learning Online with Wikis School of Information Technology. *In Proceedings of the 21st Australasian Society of Computers In Learning In Tertiary Education Conference.* Western Australia, Deakin University, Australia.

Bey, Y., Kageura, K. and Boitet, C. (2005). A Framework for Data Management for the Online Volunteer Translators' Aid System QRLex. *In Proceedings of the 19th Pacific Asia Conference on Language, Information and Computation.* Taiwan, pp. 51-60.

Boitet, C., Bey, Y. and Kageura, K. (2005). Main Research Issues in Building Web Services for Mutualized, Non-Commercial Translation. *In Proceeding of the 6th Symposium on Natural Language Processing, Human and Computer Processing of Language and Speech*, Thailand.

Bowker, L. (2002). Computer-Aided Translation Technology: *A Practical Introduction. Didactics Of Translation Series.* University of Ottawa Press, Canada.

HTMLArea. In-browser wysiwyg editor for XWiki. http://www.htmlarea.com/ (last accessed 02/02/2006).

Hutchins, J. (2003). Machine Translation and Computer-Based Translation Tools: What's Available and How It's Used. *Edited Transcript of a Presentation.* University of Valladolid, Spain. http://ourworld.compuserve.com/homepages/WJHutchins/ (last accessed 26/01/2006).

[6] http://localhost:8080/xwiki/bin/view/+Paxhumana/InfameTorture (Link to the French translation)

LingPipe. Linguistic Tools (Sentence-Boundary Detection, Named-Entity Extraction, Language Modelling, Multi-Class Classification, etc.). http://alias-i.com/lingpipe/demo.html (last accessed 10/01/2006).

LISA: Localization Industry Standards Association. Translation Memory eXchange. http://www.lisa.com/ (last accessed 25/01/2006).

Mozilla: French Mozilla Project. Open Software Localization. http://frenchmozilla.online.fr/ (last accessed 11/11/2005).

PaxHumana: Translation of Various Humanitarian Reports in French, English, German, Spanish. http://paxhumana.info (last accessed 30/01/2006).

Queens, F. and Recker-Hamm, U. (2005). A Net-Based Toolkit for Collaborative Editing and Publishing of Dictionaries. *Literary and Linguistic Computing Advance Access*. Oxford Journal. Lit Linguist Computing, pp. 165-175.

Radeox Engine. http://radeox.org/space/start (last accessed 15/12/2005).

Saha, G.K.A. (2005). Novel 3-Tiers XML Schematic Approach for Web Page Translation. *In ACM IT Magazine and Forum*. http://www.acm.org /ubiquity/views/v6i43_saha.html (last accessed 26/01/2006).

Schwartz, L. (2004). Educational Wikis: Features and Selection Criteria. *International Review of Research in Open and Distance Learning*. Athabasca University, Canada's Open University. http://www.irrodl.org/content/v5.1/technote_xxvii.html (last accessed 27/01/2006).

TeaNotWar: Human Rights Documents Translation. English to Japanese News Translation. http://teanotwar.blogtribe.org/ (last accessed 20/12/2005).

Traduc Project. Linux Documentation Translation. http://wiki.traduc.org/ (last accessed 20/11/2005).

W3C Consortium. Specification Translation. http://www.w3.org/Consortium/Translation (last accessed 01/11/2005).

Walker, D.J., Clements D.E., Darwin, M. and Amtrup, W. (2001). Sentence Boundary Detection: A Comparison of Paradigms for Improving MT Quality. *In Proceedings of the 8th Machine Translation Summit.* Spain.

XWiki. http://www.xwiki.com/ (last accessed 30/01/2006).