TOWARD A COLLABORATIVE WIKI-BASED TRANSLATION TOOLS FOR VOLUNTEER TRANSLATORS 'TRANSBEY'

Youcef Bey

Graduate School of Advanced Studies,
National Institute of Informatics (NII),
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo,
101-8430, Japan.
youcefb@grad.nii.ac.jp

Laboratoire CLIPS-GETA-IMAG,
Université Joseph Fourier,
385, rue de la Bibliothèque. Grenoble,
France.
Youcef.bey@imag.fr

Abstract : In this paper we discuss how to exploit the collaborative Wiki-based technology for the design of online Computer-Aided Translation environment TRANSBey, which is currently under development. The system pays maximal attention to the facilitation of the management and use of translation resources, and fills the existing gap between the requirements of online volunteer translator communities and existing CAT systems.

Keywords: Computer-Aided Translation (CAT), Collaborative Translation, Sentence Boundary, Translation Memory, Linguistic Resources, Online Wysiwyg Editor, Wiki Technology.

1. Introduction

We notice recently an important grow of online volunteer translators who are translating thousands of documents in different fields and, thereby, showing the true way to break the language barrier. After examining the current status and conditions of online volunteer translators and their translation environments are examined we have identified two main classes (BEY, 2005):

- Mission-oriented translators communities (FRENCH-MOZILLA, 2005), (TRADUCT, 2005).
- Subject-oriented translator networks communities (PAXHUMANA, 2006) (HUMAN-RIGHTS, 2005).

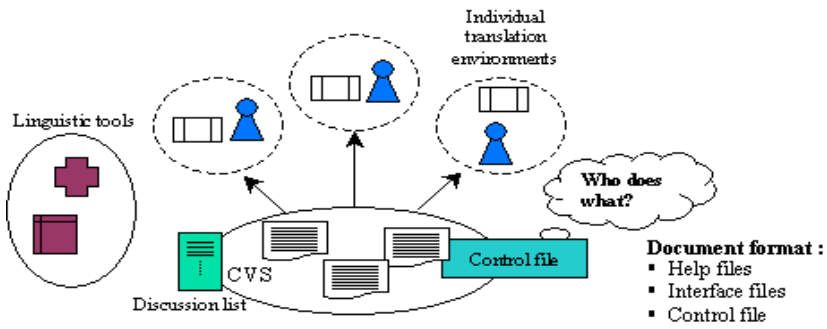

Figure 1 – Translation methods.

In general, translation is done on individual stand-alone environment using non-identified methods and linguistic resources (figure 1). In the process, they do not use linguistic tools and the communication between them is weak; these happen when they look for exchanging skills and reusing what other volunteer translators produce as good/high translation quality.

2. Translators' needs

By talking with volunteer translators and analyzing existing online/stand-alone translation aid environments, translator needs are:

- Translators look for integrated linguistic resources to the system functionality.
- They look for specific information (ordinary words, idioms, technical terms, etc.).
- They look for tools for collaboration and reusing existing translation.

QRLex framework (BEY, 2005) is the first attempts for helping volunteers translators need. In spite of its advantages for offering a rich content, it is not yet at the stage to solve all problems. Among its failure the following points:

- Lack of right access and group control.
- Recovering and recycling sentences from translated document is the main goal of QRLex, but this is not-sufficient when translators want to check the translation in its context or position in the source/target documents, especially for the reconstruction of documents (BOWKER, 2002).
- Collaboration is weak (QEEN, 2005).
- Lack for source/target edition of document.

In order to overcome these disadvantages, in next sections, I attempt to introduce existing technology and specifications for solving such stated problems.

3. Wiki-based technology

Wiki environment allows users to freely create and edit Web page content using any Web browser (XWIKI, 2005) (AUGAR, 2004). It can be exploited efficiently for the development of fully usable online CAT environments (HUTCHINS, 2003). But it remains to adopt another technology for structuring documents and TMs.
As the whole work is broad and depending on time, I present in the next two sections, the XML format for structuring documents/TMs as well as the wysiwyg online translators-oriented editor.

4. Specification and translators-oriented editor

4.1. TMX adoption for collaborative support management

The segmentation process is done semi-automatically. Firstly, a preliminary scan is done with LingPipe tools (LINGPIPE, 2005). Mainly, the segmentation serves for the auto-construction of linguistic resources like TMs, etc. In order to improve efficiently the collaboration, segmentation and multilingual aspects, content will be supported in unified XML structure. I proposed TMX-C (TMX for Collaboration) an extension of TMX standard (LISA, 2006) and 3-tiers level models (SAHA, 2005). TMX-C serves also for the importation/exportation of data between human experts (translators, managers, designer, etc.) and software (Figure 2).

```
<?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 2 – "TMX-C" format for multilingual support of documents/TMs.

4.2. Translators-oriented wysiwyg editor

There is an interest in developing online translators-oriented editors for sharing TMs and same documents for translation (BOITET, 2004). Our integrated editor has the following features (HTMLAREA, 2006):

- Compatibility with almost all web browsers (Mozilla, Firefox, IE).
- Produce a well formed HTML code.
- Easy to integrate it to Wiki for managing Wiki syntax.
- Contain several editing wysiwyg features (table, images, heading, etc.).

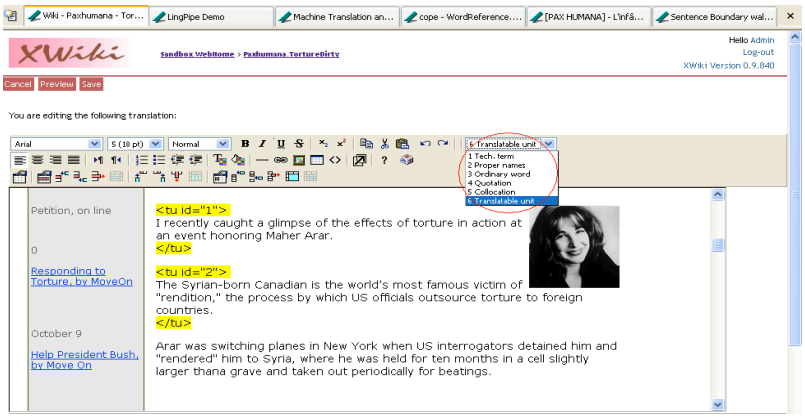

Figure 3 – Collaborative wysiwyg editor and direct source/target document segmentation.

Figure 3 illustrates the wysiwyg edition with the integration of the semi-automatic segmentation which serves for the construction of rich linguistic resources for translators.

5. Conclusion

We have outlined the general aspects for the design of fully open online CAT environment 'TRANSBey'. This environment is dedicated for the online volunteer translators for helping them during collaborative translation process. It is thus among the first attempts for designing of an open environment for collaborative edition, recycling translated data and integrating improved functionalities with the respect of translator's needs.

References

AUGAR, NAOMI , RAITMAN, RUTH & ZHOU, WANLEI (2004). Teaching and Learning Online with Wikis School of Information Technology, *Proceedings of the 21st Australasian Society of Computers In Learning In Tertiary Education Conference*, Western Australia, Deakin University, Australia.

BEY, YOUCEF, KAGEURA, KYO & Boitet, CHRISTIAN (2005). A Framework for Data Management for the Online Volunteer Translators' Aid System QRLex, *Proceedings of the 19th Pacific Asia Conference on Language, Information and Computation*, p. 51-60, Taiwan.

BOITET, CHRISTIAN, BEY, YOUCEF & KAGEURA, KYO (2005). Main Research Issues in Building Web Services for Mutualized, Non-Commercial Translation, *Proceeding of the 6th Symposium on Natural Language Processing, Human and Computer Processing of Language and Speech,* Thailand.

BOWKER, LYNNE (2002). Computer-Aided Translation Technology: A Practical Introduction, *Didactics Of Translation Series,* Otawa, Otawa Press.

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

QEEN, FRANK & RECKER-HAMM, UTE (2005). A Net-Based Toolkit for Collaborative Editing and Publishing of Dictionaries, *Literary and Linguistic Computing Advance Access,* 165-175, UK: Lit Linguist Computing*, Oxford Journal*.

SAHA, GOUTAM KUMAR (2005). Novel 3-Tiers XML Schematic Approach for Web Page Translation, *Ubiquity ACM IT Magazine*, 23-29, http://www.acm.org/ubiquity/views/v6i43_saha.html, (nd). (Last accessed 26/01/2006).

SCHWARTZ, LINDA, CLARK, SHARON, COSSARIN, MARY & RUDOLPH, JIM (2004). Educational Wikis: Features and Selection Criteria, *International Review of Research in Open and Distance Learnin,* vol 5, No 1, Canada.

WALKER, DANIEL J., CLEMENTS DAVID E., DARWIN, MAKI AND AMTRUP, JAN W. (2001). Sentence Boundary Detection: A Comparison of Paradigms for Improving MT Quality, *Proceedings of the 8th Machine Translation Summit,* Spain.

Web Sites

FRENCH-MOZILLA PROJECT (Open Software Localization), http://frenchmozilla.online.fr/, (nd). (Last accessed 11/11/2005).

HTMLAREA. In-browser WYSIWYG editor for XWiki. http://www.htmlarea.com/, (nd). (Last accessed 02/02/2006).

HUMAN-RIGHTS (Report Documents Translation, English to Japanese News Translation) http://teanotwar.blogtribe.org/, (nd). (Last accessed 20/12/2005).

LINGPIPE. Linguistic Tools (Sentence-Boundary Detection, Named-Entity Extraction, Language Modeling, Multi-Class Classification, etc.), (nd). 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).
PAXHUMANA. Translation of Various Humanitarian Reports in French, English, German, Spanish. http://paxhumana.info, (nd). (Last accessed 30/01/2006).
TRADUCT (Linux Document Translation), http://wiki.traduc.org/. (Last accessed 20/11/2005)
W3C (Specification Translation), http://www.w3.org/Consortium/Translation, (nd). (Last accessed 01/11/2005).
XWIKI, http://www.xwiki.com/, (nd). (Last accessed 30/01/2006).