

Translating the DEMGOL Etymological Dictionary of Greek Mythology with the BEYTrans Wiki

ABSTRACT

BEYTrans (Better Environment for Your TRANslation) is a generic Wiki tool designed to support communities of volunteer translators not only by offering them an online translation editor and helps to manage the translation progress, but a complete online computer-assisted translation (CAT) environment including a translation editor (BT-editor), translation memories, free dictionaries, automatic calls to MT systems, and support to collaborative volunteer translation. We present the basic concepts of BEYTrans and its experimentation on the translation from Italian to French of the DEMGOL project (OnLine Etymological Dictionary of the Greek Mythology).

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: User-centered design, Interaction styles, Natural language, Ergonomics; H.5.4 [**Hypertext/Hypermedia**]: Navigation, User issues; H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work, Web-based interaction.

General Terms

Design, Experimentation, Human Factors, Standardization, Languages.

Keywords

Collaborative Translation, Wiki, XML, Dictionary, Translation Memory, Fuzzy Matching, DEMGOL, Online Translation Editor, Computer-Assisted Translation, CAT, Multilingual Segmentation.

1. INTRODUCTION

Free online translation is becoming a social activity: thousands of volunteer translators are now involved in collaboratively translating into many languages and disseminating a wide range of open documents in multiple fields (e.g. W3C consortium, Traduct project, Mozilla localization projects...). Translation quality is usually quite high. Contrary to their professional counterpart, these communities did not yet take advantage of progress in NLP and translation technology. However, progresses in Web technology have recently led to the creation of two categories of free Computer Aided Translation (CAT) tools.

- (1) **Category A** includes tools that allow working in a standalone mode with linguistic helps such as translation memories and dictionaries.
- (2) **Category B** includes online tools that propose basic linguistic helps in the scope of web-based collaborative or individual contributions.

With A-type tools, volunteer translators are limited to work in “standalone” mode, which means that documents and linguistic resources are not visible and not shared with other translators. Omega-T is an example [21]. It integrates glossaries and TM suggestions depending on the translation project. Documents to be translated are segmented into translation units (TU). Translation is then done segment by segment. Omega-T can manage several document formats (OpenOffice, HTML...), and supports almost totally the TMX format for translation memories and the LISA standards for terminology [18], so that it can exchange data with other CAT systems (Trados, Similis...).

Type-B tools rely on the Wiki technology that makes it possible to develop collaborative online CAT tools. That technology has been invented by Cunningham Ward for managing efficiently communities’ contributions [15]. It is actually used for developing a variety of collaborative environments in several fields, in particular those oriented towards translation. For example, it was

exploited¹ for the construction of the collaborative Wiki-based environment “translationwiki.net” [28], which aimed at translating newspaper articles on a “volunteer” basis, and at disseminating the translated versions. It allowed easily switching between reading and editing and allowed users to upload plain text documents which were automatically segmented into a sequence of translation units (TU). The translation/edition was also done segment by segment. In fact, each TU was handled as a separate Wiki document, and translators could not check whole translated document, e.g. they could not perform global changes on a document. Contrary to Omega-T, there was no translation memory, so that previously translated segments could not be proposed as suggestions. Like Omega-T and almost all online CAT tools, it also did not offer any dictionary help [5].

Another interesting free environment which illustrates type-B tools is Pootle (PO-based Online Translation/Localization Engine) [23]. It is oriented for the translation of free software based on the PO (Mozilla, OpenOffice, etc). Volunteer translators are able to organize themselves around one or more projects and share the translation of PO files. The translation is again done segment by segments. However, it helps at attributing tasks as a set of “goals”. One goal can be a set of documents which can be attributed for one or more translators. The final translations could be exported and used as it is (in the PO format) for the compilation of multilingual software versions. It proposes neither open editing nor the historical flow. Translators, however, have to identify them self to be contributors. This is in contradiction with the open authoring concept proposed in Wikis. Data are also stored in respect of XLIFF LISA standards [18].

CAT tools of both categories offer of course some translation helps, but that remains too “basic”, and largely insufficient with respect to the needs expressed by the vast majority of volunteer translators we have interviewed or observed [1] [5] [14]. Our work focuses type-B tools, in particular, online collaborative Wiki CAT. However, we have attempted to combine advantages of both categories in our BEYTrans prototype.

For evaluating the usefulness of BEYTrans, we collaborated with the translation subproject of the DEMGOL project [9]. Feedbacks about the utilization of the environment helped us to enhance several components, especially the integrated online editor.

In the rest of this paper, we introduce the DEMGOL project and analyze the nature of the documents to be translated. The design and implementation of BEYTrans is then presented in detail, in particular the online translation editor BT-Edit. Finally, we describe an experiment and an evaluation done in a real translation situation, in the framework of the DEMGOL project. Results and discussions are also exposed at the end of the paper for showing the usefulness of our environment.

2. DEMGOL

The research group on Mythography and Myths at the University of Trieste (GRIMM, Department of Sciences of Antiquity "Leonardo Ferrero") has developed a project for the construction of an Online Etymological Dictionary of Greek Mythology²

¹ This web site is no longer operational, new ones have appeared.

² Dictionnaire Étymologique de la Mythologie Grecque OnLine.

(DEMGOL) [9]. The translation sub-project aims at translating it from Italian into French, Spanish and English.

Let us first fix our terminology.

- What is called a “document” to be translated in a usual translation project is an “entry”³ in the case of DEMGOL.
- As DEMGOL is itself a dictionary, we will use the term “translation lexicon” to refer to the specialized dictionary supported by BEYTrans (built by the translators, starting from some freely available bilingual terminological list if possible). Units of information are called “entries” in both cases.

2.1 Participants

Thanks to a grant awarded to Ms. Marzari Francesca (translator and specialist in Greek mythology, University of Sienna), as part of an “Émergence” project⁴ obtained at Grenoble by the group of Prof. Létoublon, specialist of ancient Greece, and leader of the “Homerica” project, a collaboration started between the Universities of Trieste (Italy), Sienna (Italy), and Grenoble (France). The two Italian groups coordinate the different contributions, the translation work, and manage the DEMGOL Web server at Trieste⁵. The Grenoble group takes care of translating and revising entries from Italian into French.

2.2 Need for Translation Helps

The DEMGOL translation project aims at translating about 1200 entries from Italian into French, Spanish and English. The number of segments in an entry varies between 5 and 20 (**Figure 1**).

Animale dell'India, feroce, antropofago, di colore fulvo o rossiccio. La grafia marticwvra prevale nelle fonti greche; in latino troviamo per lo più *manticora*, maschile, che diventerà femminile nelle fonti più tardive e medievali. È descritto dettagliatamente da Ctesia di Cnido (V-IV sec. a.e.v.) negli *Indikà* (F 45 14-15: riassunto nella Biblioteca di Fozio): vive in India, ha il viso, gli occhi e le orecchie simili all'uomo, le zampe di leone e una coda di scorpione in grado di scagliare come frecce gli aculei che vi crescono. Eliano (*Nat. an.* 4, 21) paragona curiosamente il suo modo di combattere a quello dei Saci, popolazione degli Sciti, noti come abilissimi arcieri a cavallo.

...

Figure 1. The original “MANTICORA” entry.

The idea of experimenting BEYTrans on DEMGOL was proposed by the third author, who previously cooperated with the Homerica projet, to put the Homerica dictionary (another one) on the Papillon Web-based contributive multilingual lexical database. Our hopes were to make progress on two important points: (i) acquire a realistic “data bank” for testing and evaluating our environment, (ii) enhance BEYTrans by making the best of feedbacks expected from the translators. These hopes were fulfilled, as we will see.

³ “Notice”, or better “entrée” in French.

⁴ Funded by the Rhône-Alpes region.

⁵ The technical aspects of the DEMGOL project are managed at Trieste (Italy) by Mr. Giovanni Zorzetti under the direction of Prof. Ezio Pellizer. Access to the web site is free for looking up and doesn't requires any authentication, except for administrators, for which, especially during the translation, authentication is required.

We will describe the experiment in more detail in section 4 below. In short, we have a BEYTrans instance, BT-DEMGOL, imported all DEMGOL entries into it, including those already manually translated, semi-automatically segmented them and aligned them when a translation was available, and put them in the translation memory (TM), with their translation. We built an initially empty translation lexicon, installed BT-DEMGOL on a server in the second author’s lab, and the experiment could begin.

3. BEYTRANS: TRANSLATION THE WIKI WAY

Before giving the details of the experiment, let us now describe how we designed and developed BEYTrans “the Wiki way”, and were helped by the emerging Wiki technology to produce relatively quickly and reliably an environment that integrates functionalities needed for both volunteer and professional translators communities, and offers the same “benefits” as of almost any good CAT tool currently available [1] [14].

We choose for our implementation the Java-based Wiki “XWiki” [33], after testing and comparing it with several other Wikis. As many other Wikis, XWiki includes communities support and collaboration features. Its historical flow module allows translators working on the same documents and controlling progression simultaneously.

XWiki was essential in developing the following modules: (1) multilingual document importation, (2) multilingual heuristic segmentation, (3) collaborative multilingual dictionaries and translation memories, and (4) in-browser online multilingual editor. Each module includes several functionalities [5] [6]. For example, the online editor includes “proactive” suggestions provided by translation memories using a fuzzy matching method [16] [29], MT (free online machine translation services), and dictionaries.

Using XWiki, we avoided developing the environment from scratch, and we could develop all modules corresponding to needed functionalities as additional layers. These modules interact through the internal API of XWiki (Figure 2).

3.1 Adding a Translation Layer on Wiki

Wiki environments allow users to freely create and edit web page content using any web browser. On one hand, they offer a simple syntax for creating new pages and links between internal pages. On the other hand, they allow 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 an Internet connection (Schwartz, 2004).

We can think of the XWiki’s modules as a set of layers. The central layer contains the API that offer functions for extending XWiki (access control, addition of new functionalities...), historical flow for checking progress, and documents with their

spaces. Users and communities’ access are controlled in the second layer⁶.

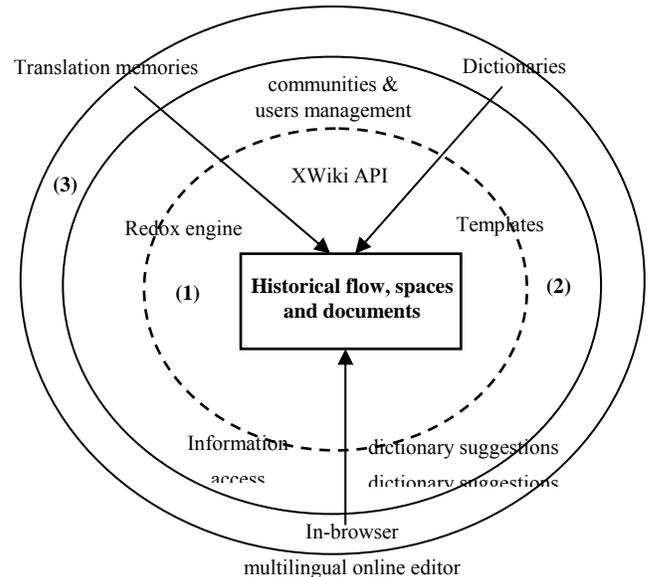


Figure 2. BEYTrans architecture, based on XWiki.

As for the translation layer, we developed and integrated dictionaries, translation memories and our online editor BT-editor. All of them communicate with other layers as if they were a part of the original XWiki. Any modification in any linguistic resources is saved for historical flow. The historical flow is done for the three levels following data: documents, dictionaries and translation memories.

One layer includes the most important component, which is the online translation editor BT-editor. The data flow and interaction between linguistic data and BT-editor are now explained.

3.2 Multilingual Online Editor “BT-editor”

BT-editor allows translators enhancing content and disseminating multilingual content. According to its configuration, it can propose various linguistic helps: TM, dictionaries, and MT.

All linguistic data are accessed (requested, modified) from this component. It has been developed to minimize the actions of translators. As a consequence, it calls MT servers, dictionary lookup, and TM search in the background, before the results are used, in order to offer suggestions in a “proactive” manner.

The translation is done in an Excel-like interface [11]. Columns sets contain TU of one language and lines contain cells of several languages.

The segmented textual content of a source document (to be translated) is stored in the TMX XML format [18]. This format is generated in three steps. Firstly, the textual content of the original document is extracted. Secondly, Lingpipe is used for generating segments (or TU) that serves for making the TMX documents [17]. For each imported document, the environment creates TMX

⁶ Access to XWiki content can be controlled. Users are identified via their IP addresses, and access can be limited for specific documents and functionalities.

document for supporting extracted segments which we call the *translation companion* (TC) and the original document.

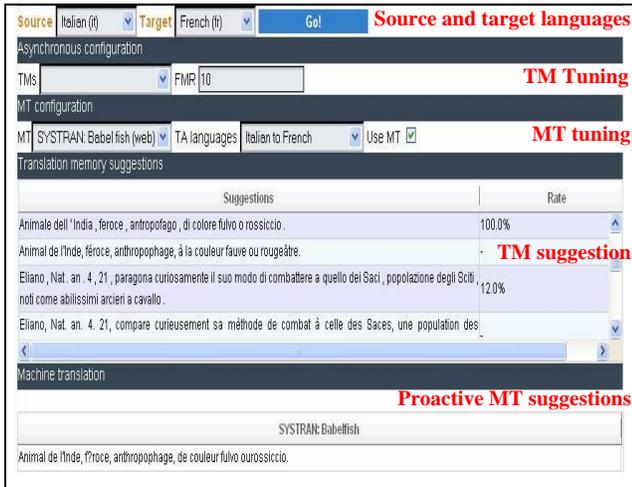


Figure 3. Tuning panel.

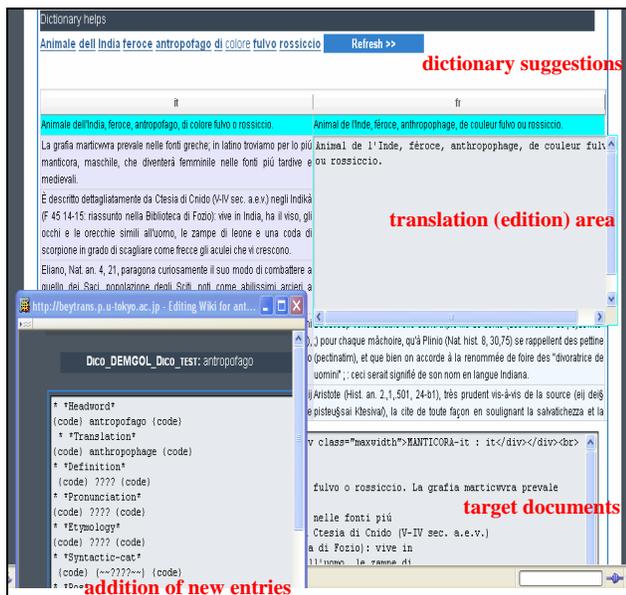


Figure 4. Main interface of BT-editor: translation of “MANTICORA” entry.

3.2.1 BT-editor Interface Design

Switching from the reading mode to the editing mode triggers the creation of a translation session. The TMX TC and the associated linguistic resources are prepared in advance (proactively).

Figure 3 presents a screenshot of the editor tuned to use “Systran Babel Fish” as MT server for the “pre-translation” from Italian into French [3]. Concerning the TM, the fuzzy matching threshold is fixed to 10. Translators may change this value to have more accuracy. The translation correspondence of Wiki space (or community) is selected as the default TM. The areas at the bottom of the Figure 3 contain TM and MT suggestions. Translators select the best suggestion and post-edit the content of the suggested segment. The figure shows complete TM matches of the current segments selected in Figure 4.

After segments have been translated and post-edited, they are gathered in a new TMX TC and sent to the XWiki repository. The new TMX TC is saved as a Wiki document that maintains old versions for historical flow. This is done by the XWiki API. Translators hence can follow translation at both document and segment level.

3.2.2 Multilingual TU Management

The interaction process between stored TUs and the editor requires conversions between XML formats. As explained before, the multilingual TUs are saved in specific TMX-based TC format. However, that format is not adequate to display what has to be displayed when translating documents.

One of the reasons is that translators are often interested to display limited number of languages, source and one or two targets, and not all possible target languages. That is why the editor calls a “search” function that makes requests for extracting the TUs from the TMX TC. Figure 4 shows a screenshot of the editor where only the Italian (source) and French (target) TUs are displayed.

The search function uses a parser that extracts the adequate TUs⁷ from the XML TC structure and transforms them to another XML format which is specific to the editor. This feature is shown in Figure 5 in step 1. Step 2 illustrates the opposite process, that transforms edited TUs from the editor format into the TC format and sends it to the database, where it is saved again as a Wiki document.

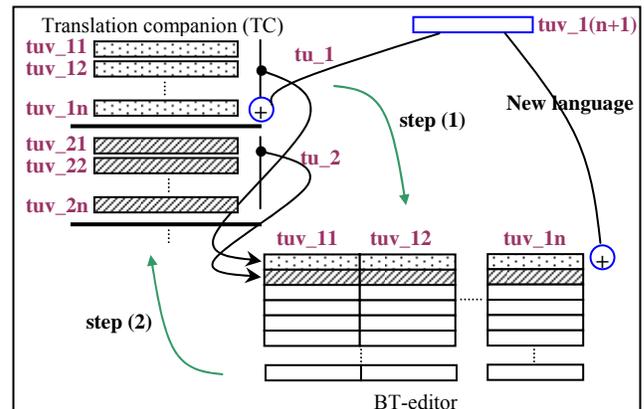


Figure 5. Conversions between the TC and BT-editor XML formats.

The TUs which correspond to the same “source segment” are stored between <tu> XML elements. In the editor, one line is the image of the content of one <tu> element. Each <tuv> contains the text of TU which is displayed by the editor.

Doing translation on the basis of segments allows to easily adding a new language without leaving the editor. This is done visually by adding a new column. In the background, BEYTrans creates a new <tuv> for each <tu> in the active TMX TC document.

⁷ The program is based on the XML DOM API (Document Object Modeling).

```
<?xml version="1.0" encoding="UTF-8"?>
<rows>
<row id="MANTICORA">
<cell>Animale dell'India, feroce, antropofago, di colore fulvo o
rossiccio.</cell>
<cell>Animal de l'Inde, féroce, anthropophage, de couleur fulvo
ou rossiccio.</cell>
</row>
<row>
<cell>...</cell><cell>...</cell>
</row>
<row>...</row>
</rows>
```

Figure 6. XML editor format of the DEMGOL “MANTICORA” entry.

4. EXPERIMENTATION

4.1 Adaptation for DEMGOL Project

4.1.1 Data Importation

The importation of the 1200 entries was done from a compiled database dump sent to us from Trieste (Italy) by the administrator of the DEMGOL server. The initial encoding was ISO-5988-1, which necessitates the installation of the Greek font for displaying specific Greek characters in Web navigators. Names in Greek like “MANTICORA” must be displayed as “Μαντιχώρας (Martichóras)”. For overcoming this problem, we have converted all entries into UTF-8. That also facilitates operations such as sorting and searching in the core of BEYTrans, because it internally processes all textual data in UTF-8.

The textual contents of entries were then extracted and segmented by Lingpipe [17]. For each entry, a TMX TC⁸ was created [6].

4.1.2 Creation of the DEMGOL Translation Lexicon

As explained before, we created an initially empty BEYTrans translation dictionary, specialized to terms related to the Greek mythology. To avoid confusion, we call it here the “translation lexicon”. It is used to automatically (proactively) offer lexical suggestions as shown in Figure 7. Simple words like (“o”, “di”) are excluded from the search.



Figure 7. Proactive dictionary suggestions for the segment “Animale dell’India, feroce, antropofago, di colore fulvo o rossiccio.”.

Lexicon entries having a translation are displayed differently from those missing a translation. In Figure 8, for example, lexicon entry “antropofago” misses a French translation. The translator can click on it (it is a Web link) for adding its translation to the lexicon.

Figure 8 also shows the structure of a lexicon entry. In addition to dictionary information (headword, translation, definition, etc.), there are formatting Wiki tags that serve for the presentation of the lexicon.

⁸ The TC concept was presented in our precedent work for managing TU of each document. It is also used by the online editor for managing segments.



Figure 8. Collaborative addition of a lexicon entry.

Proactive dictionary suggestions allow to avoid checking the (existing) paper dictionary, and to enrich the online lexicon when encountering a missing source headword or a headword missing its translation (in the current target language, here French).

4.1.3 Creation of the DEMGOL TM

The “TM_DEMGOL” translation memory was also created and left initially empty at the beginning of the experiment, because that is the most common case when one starts using a CAT tool. However, contrary to the lexicon, it could have been initialized with segment pairs produced by the alignment of already translated DEMGOL entries.

TM_DEMGOL is the set of all segments stored in all TCs of the same “space” (associated here to the DEMGOL community).

Suggestions are proposed systematically when the translator activates one segment. Figure 9 illustrates an activated segment that has to be post-edited to get a good translation in French.

it	fr
Animale dell'India, feroce, antropofago, di colore fulvo o rossiccio.	Animal de l'Inde, féroce, anthropophage, de couleur fulvo ou rossiccio.

Figure 9. Active segment in the BT-editor.

Suggestions	Rate
Animale dell'India, feroce, antropofago, di colore fulvo o rossiccio.	100.0%
Animal de l'Inde, féroce, anthropophage, à la couleur fauve ou rougeâtre.	
Eliano, Nat. an. 4, 21, paragona curiosamente il suo modo di combattere a quello dei Saci, popolazione degli Sciti, noti come abilissimi arcieri a cavallo.	12.0%
Eliano, Nat. an. 4, 21, compare curieusement sa méthode de combat à celle des Saces, une population des	

Figure 10. TM suggestions for an active segment.

Suggestions from the MT on this example are shown in Figure 10. The first suggestion found is a complete match (100% match rate). The MT suggestion method proposes matches that depend on how the threshold is tuned (Figure 3).

The relevant match is selected and automatically inserted in the correct editing position (cell).

Suggestions from MT servers are also prepared beforehand, by “asynchronous MT calls”. Hence, the translator’s work is not slowed down by waiting for MT calls to produce “pre-translations”.

4.2 Experimentation Protocol

The experimentation uses 2 sets of source Italian entries not yet translated into French. The first set (E₁) contains entries with

headwords starting with “L”. This set contains 60 entries (e.g. “LABDACOS”, “LACEDEMONE”, “LAKIOS” ...). The second set (E₂) contains 30 entries selected from the “M” list (e.g. “MACEDONE”, “MACHEREO”, “MACISTO” ...). E₁ will be translated without BT-editor (without any translation assistance) and E₂ under BT-editor using the available helps (segmentation, TM and lexicon look-up, calls to MT servers).

The experiment is done following several steps that have been defined in a specific protocol⁹:

Segmentation improvement: the segmentation process can produce incorrect TUs (Translation Units). For overcoming this weakness, we have integrated the possibility of merging and splitting erroneous detected segments. Note that the segmentation of French and Italian using the “Lingpipe” tool is sufficient and errors are rare. For example, for 90 Italian entries, the segmentation process produced only a single incorrect segmentation.

MT pre-translation: all DEMGOL entries of the E₂ set have been pre-translated by the online free MT server “Systran Babel Fish” [3]. This pre-translation is produced in systematic way. The translator has only to post-edit, but not to translate the whole entry from scratch.

Suggestions by TM: as explained above, TM DEMGOL was initially empty. It was gradually enriched by storing the results of translation (rather, post-edition) in new TUs. During post-edition, suggestions were proposed by fuzzy matching process.

Dictionary suggestions: as said above, the translation lexicon was also initially empty. The translator was asked to add or complete entries whenever BT-editor showed her that a translation was missing, and she did.

4.3 Results

Saving time when translating is important. In order to evaluate the efficiency of our environment, the main translator implied in the translation of entries was instructed to carefully report the translation times.

The duration of the whole translation of entries from the beginning to the end of the experiment was approximately 80 hours (2 weeks if working full time). Translation was actually not performed on an continuous basis, because the translator¹⁰ was not constantly available.

Her skills on the Greek mythology and languages fluency had of course a positive impact on the experiment¹¹.

⁹ The protocol is restricted for the set E2. Other set has to be translated with the classic method without any translation helps.

¹⁰ The experiment was supposed to be done by at least two volunteer translators. But in the scope of the project, especially for the Italian-French pair, only one Italian translator was available for the experiment. It proved impossible to find another translator with the same skills. Hence, we limited the experiment to translate different documents with the same translator, assuming that DEMGOL entries have the average size and complexity.

¹¹ A fluent speaker of French and a native speaker of Italian, she was the main volunteer translator involved in the BT-DEMGOL experiment, during which she translated 90 DEMGOL entries from Italian into French.

Before starting the experiment, we wrote a user guide and sent it to the translator.

During the experiment, we got from her many feedbacks for enhancing the editor and making it more user-friendly. For example, the first version of the editor did not allow seeing the whole document as an entity while translating “segment by segment”, and she felt it necessary. We hence added an area for displaying the target document generated from the translated segments.

The learning process has an important impact for enhancing the environment, because the translator has a point of view which is totally different from that of a computer scientist and developer. Feedbacks mainly concerned the usability and the simplification of access to functionalities like pre-visualization, suggestion from lexicon entries, automatic insertion of MT and TM at the right position (cell), how to save new translation segments in the TM, etc.

Translation times are presented in the **Table 1**.

As said above, all entries of the E₂ set have been translated using the free online MT server “Systran Babel Fish”¹² for Italian-French. **Figure 11** shows rough “pre-translation” produced for the entry “MANTICORA”.

The translator has to choose the best suggestion given by TM or MT. In both cases translator has to post-edit the relevant suggestion. In the case of 100% complete match of TM the suggestion has just to be inserted without post-edition.

Table 1. Translation duration with and without using BEYTrans

	Set	Nb	Time	Remarks
Without BEYTrans	E ₁	60	7,11h	Classic method (paper version dictionary and no automatic translation helps)
With BEYTrans	E ₂	30	5,13h	The duration of translation includes the time of adding new entries in the DEMGOL “translation lexicon” dictionary.

Animal de l'Inde, féroce, anthropophage, de couleur fulvo ou rossiccio. La grafia marticwvra prévaut dans les sources grecques ; dans latin nous trouvons pour le piú manticora, masculine, qui deviendra féminine dans les sources piú tardives et médiévaux. Il est décrit en détail de Ctesia de Cnido (V-IV sec. a.e.v.) dans l'Indikà (F 45 14-15 : repris dans la Bibliothèque de Fozio) : il vit en Inde, a le visage, les yeux et l'orecchie semblables all'uomo, les pattes de lion et une queue de scorpion en degré de lancer comme flèches les aiguillons qui vous croissent. Eliano (Nat. an. 4, 21) compare curieusement sa mode de combattre à celui des Saci, population des Sciti, remarques comme très habiles archers à cheval.

Figure 11. MT rough “pre-translation” of the “MANTICORA” entry.

As for the evaluation, these results presented in **Table 1** don't reflect the exact performance of the environment because the

¹² Babel Fish online MT is based on the Systran MT technology.

translation duration of the E_2 set does not include only the translation duration of entries, but it also includes the duration of dictionary inputs, which were not reported by the translator because it would have disturbed the rhythm of translation.

For estimating the exact time of dictionary inputs, the following steps have to be considered when adding new entry to the translation lexicon:

- Detection of the missing entry from the suggestion list.
- Click on the right link of the absent word.
- Showing the input window.
- Positioning on the source field.
- Inputting the source headword.
- Inputting the translation.
- Saving the new entry.

At the end of the translation of the E_2 set (letter M), the total number of additions in the translation lexicon was 174 entries. For having more clear idea about the difference in translation duration with and without using BEYTrans the duration is calculated on basis of 30 entries. The lexicon inputs duration have to be deducted from the total translation duration. We propose here an approximation of the time of one lexicon input. This situation is summarized in two cases:

- (1) **case 1:** if a dictionary addition lasts “1 minute”, then the total time of the translation, taking into account the 174 entries, is “3.68 h”.
- (2) **case 2:** if a dictionary addition lasts “0.5 mn” with the same number of entries added, the total time of the translation is “2.23 h”.

According to constraints of the experiment, we assumed that the *average size* and *complexity* of entries are almost the same. The duration of translation of 30 entries without using BEYTrans become “7.11h/2 = 3.5h”. Hence, the difference in translation duration for the *case 1* is 1.31h (3.5h – 2.23h; time of entry addition is 1mn). As for *case 2*, translation duration 0.14h (3.5h – 3.68h; the time of addition is 0,5mn).

Although the experiment started with an empty MT and an empty lexicon, it was ahead of the traditional method. These results, though modest, leave us optimistic, because the collaborative translation in the Wiki will most probably further reduce the translation time. If the total translation duration of 1200 entries (all document of DEMGOL) using BEYTrans by one translator were “89.2h” (case 1), that duration might be divided by 10 if 10 translators were involved in the translation.

Figure 12 shows the time of translation of E_1 . The translator had consulted her paper dictionary and online useful information on Greek historical figures.

Figure 13 shows the time of the post-edition of E_2 entries using BEYTrans. The translator added a new translation entry in the dictionary whenever a term translation was missing and at the same time she checked whether or not MT fuzzy matching suggestion was the best for current segment.

Regarding the TM, more than 500 TU (translation units) were added systematically after finishing the translation (post-edition) of entries. Some of them were actually used during the translation of new DEMGOL entries. We believe that, as in the case of Wiktionary and Wikipedia, the MT could enormously grow if the number of volunteer translators increases.

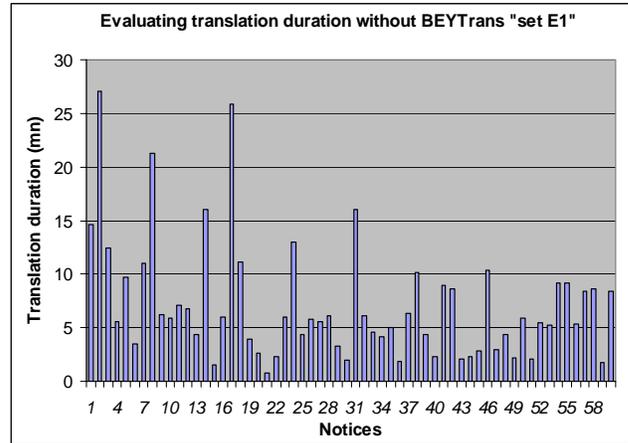


Figure 12. Translation duration of set E_1 .

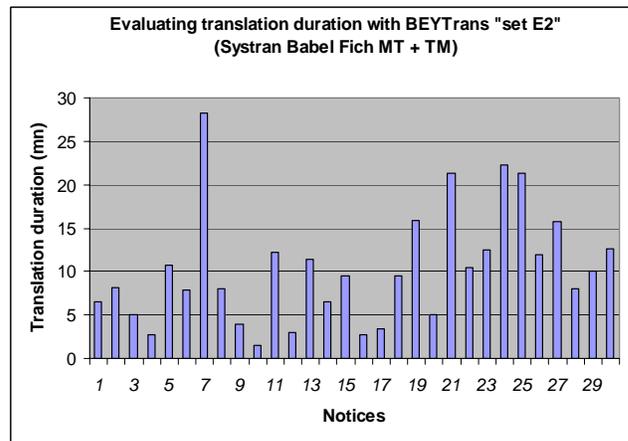


Figure 13. Translation duration of set E_2 .

In summary, this first experiment on BEYTrans was done specifically for testing the usefulness of the translation-specific functionalities and get feedback from the translator. These goals were achieved: we demonstrated a measurable reduction of translation time, even with initially empty lexicon and TM, and the feedback led to significant enhancements of all BEYTrans components, especially BT-editor. We have recently reinstalled BT-DEMGOL on a server accessible from anywhere (<http://javalig3.imag.fr/beytrans/>), so that it will remain open for other contributions on the Italian-English translation of new DEMGOL entries.

5. CONCLUSION

For increasing and facilitating the dissemination of free multilingual content on the Web, we developed the BEYTrans Wiki, using XWiki. It is dedicated to volunteer translators who translate various documents in different fields. For overcoming their needs, in particular, the lack of usable computer-assisted translation (CAT) tools, we developed several translation functionalities inspired by those available in CAT tools usable until now only by professional translators.

Our collaboration with the DEMGOL project was led to significant progress in the design and implementation of BEYTrans, and allowed us to make a first, positive evaluation of it in a real setting. The entries selected from DEMGOL allowed

evaluating BEYTrans by comparing the traditional and the new method. The result showed that our method (using BEYTrans) is ahead of the traditional method. However, only one translator participated, and we are looking forward to find a situation in which many volunteer translators would be involved, to get more significant results, and also to evaluate the speedup when the Wiki collaboration feature (support to a translators' community) is well exploited.

Our future work will be focused on 2 things: gathering more volunteers, and extending BEYTrans to support the localization of free software.

6. ACKNOWLEDGMENTS

We are very grateful to Dr Marzari Francesca for her voluntary translation and for her valuable constructive feedbacks. Special thanks go to the GRIMM group members who gave us the opportunity to evaluate BEYTrans on the DEMGOL translation subproject, and to the Graduate School of Education (University of Tokyo) for their help during the internship of the first author made possible by a scholarship from the "Collège Doctoral Franco-Japonais". Last but not least, this work was partly supported by grant-in-aid (A) 17200018 "construction of online multilingual reference tools for aiding translators" of JSPS (Japan Society for the Promotion of Sciences).

7. REFERENCES

- [1] Abekawa, T. and Kageura, K. 2007. Qredit: An Integrated Editor System to Support Online Volunteer Translators. In Proceeding of the Digital Humanities 2007 Conference, Annual Joint Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing (University of Illinois in Urbana-Champaign, US).
- [2] Arabeyes. 2008. DOI= <http://www.arabeyes.org>.
- [3] Babel-Fish. (2008). DOI= <http://world.altavista.com>.
- [4] Augar, N., Raitman, R., and Zhou, W. 2004. Teaching and Learning Online with Wikis. In Proceeding of the 21st Australasian Society of Computers in Learning in Tertiary Education Conference (Australia). 95-104.
- [5] Y. Bey, K. Kageura, C. Boitet, "Data Management in QRLex, an Online Aid System for Volunteer Translators", *Journal of Computational Linguistics and Chinese Language Processing*, 11, 4, 349-376, 2005.
- [6] Bey, Y., Boitet, C., and Kageura, K. 2006. The TRANSBey Prototype: An Online Collaborative Wiki-Based CAT Environment for Volunteer Translators. In Proceeding of the 3rd International Workshop on Language Resources for Translation Work, Research & Training (LR4Trans-III). LREC 2006 - Fifth International Conference on Language Resources and Evaluation, Paris: ELRA / ELDA (European Language Resources Association, European Language Resources Distribution Association), (Magazzini del Cotone Conference Centre, Genoa, Italy). 49-54.
- [7] Bryant, S., L. Forte, A., and Bruckman, A. 2005. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In Proceeding of the GROUP International Conference on Supporting Group Work, (Sanibel Island, Florida, US.). 1-10.
- [8] Buffa, M. 2006. Intranet Wikis. In Proceeding of the Intranet web workshop (WWW Conference). Edinburgh, Scotland, UK.
- [9] DEMGOL. 2008. DOI= <http://demgol.units.it>.
- [10] Désilets, A., Gonzalez, L., Paquet, S., and Stojanovic, M. 2006. Translation the Wiki Way. In Proceeding of the Wikiym2006 - The Conference Wiki of the 2006 International Symposium. NRC 48736, (Odense, Denmark.). 19-32.
- [11] DHTMLxGrid. 2007. DOI= <http://www.scbr.com>
- [12] Edward, A., Erich, F., Neuhold, J., Premssmit, P., and Wuwongse, V. 2005. Eyes of a Wiki: Automated Navigation Map. In Proceeding of the 8th International Conference on Asian Digital Libraries (Bangkok, Thailand). 186-193.
- [13] Jones, M., C., Rathi, D., and Twidale, M. B. 2006. Wikifying your Interface: Facilitating Community-Based Interface Translation. In Proceeding of the 6th ACM Conference on Designing Interactive Systems (University Park, PA, USA). 321-330.
- [14] Kageura, K. 2006. Improving the Usability of Language Reference Tools for Translators. In Proceeding of the 10th Annual Meeting of the Japan Association for Natural Language Processing (Japan). 707-710.
- [15] Leuf, B. and Cunningham, W. 2001. *The Wiki Way: Quick Collaboration on the Web*. Upper Saddle River (NJ, USA), Addison Wesley.
- [16] V.I. Levenshtein, "Binary Codes Capable of Correcting Deletion, Insertions and Reversals", *Soviet Physics Doklady*, 8, 10, 707-710, 1966
- [17] LingPipe, Linguistic Tool (Sentence-Boundary Detection, Named-Entity Extraction, Language Modeling, Multi-Class Classification, etc.). 2006. DOI= <http://www.alias-i.com/lingpipe/demo.html>.
- [18] Lisa, Localization Industry Standards Association. 2008. <http://www.lisa.com>.
- [19] Content and Metadata Management System. In Proceeding of the IADIS International Conference on WWW/Internet. 261-264.
- [20] Muljadi, H., Takeda, H., Kawamoto, S., Kobayashi, S., and Fujiyama, A. 2006. Towards a Semantic Wiki-Based Japanese Biodictionary. In Proceeding of the 1st Workshop on Semantic Wikis. 202-206.
- [21] Omega-T. 2008. <http://www.omegat.org/fr/omegat.html>.
- [22] Pach, J. 2004. Open Problems Wiki. In Proceeding of the 12th International Symposium of Graph Drawing (New York, NY, USA). 3383-2005.
- [23] Pootle. 2008. DOI= <http://translate.sourceforge.net/wiki/>.
- [24] Redox, XWiki engine 2005. DOI= <http://radeox.org/space/start/>.
- [25] Schwartz, L. 2004. Educational Wikis: Features and Selection Criteria. *Canada's Open* (ed.), *International Review of Research in Open and Distance Learning*, Athabasca

- University. DOI=
http://madlib.athabascau.ca/oldirrod/content/v5.1/technote_xxvii.html.
- [26] Similis, Second Generation of Online Computer Aided Translation Tools. 2005. DOI= <http://www.lingua-et-machina.com>.
- [27] Teanotwar, Human Rights Documents Translation, English to Japanese News Translation. 2005. DOI= <http://teanotwar.blogtribe.org>.
- [28] Translationwiki, Collaborative Wiki-Based Translation. 2006. DOI= <http://www.translationwiki.com>
- [29] R.A. Wagner, M.J. Fisher, "The String-to-String Correction Problem", *Journal of the ACM (JACM)*. 21, 1, 168-173, 1974.
- [30] Wales, J. 2005. Wikipedia: Social Innovation on the Web. In *Proceeding of the Web Design for Interactive Learning Conference* (Ithaca, New York, USA).
- [31] Wikipedia. 2007. DOI= <http://www.wikipedia.org>.
- [32] Wiktionary. 2007. DOI= <http://www.wiktionary.org>.
- [33] XWiki. 2006. Open Source Wiki. DOI= <http://www.xwiki.com>.
- [34] Yakushite.net. 2007. DOI= <http://www.yakushite.net>.