BEYTRANS Ver 2.0 – Users guide

Introduction

BEYTrans is a collaborative translation environment dedicated to volunteer translators. It was started in the frame of research work (PhD) with the collaboration of University of Tokyo and Joseph Fourier University. It’s one of the systems that exploited early the Wiki technology for translation (wikitranslation, traduwiki).

For promoting open translation, the system is open to any translation project (medical, technical, localization, etc). Translators can be organized in community groups (called Space or ‘translation project’).

Each project consists of documents, linguistic resources, and a community of translators. Linguistic resources and functionalities are available during translation. Suggestions from various multilingual dictionaries and machine translation are automatically proposed when translating. Mainly, linguistic resources consist of translation memories, dictionaries, terminologies, glossaries, etc.

Any data that can be introduced in the environment can be modified by translators. They can also look up, search and check the quality of these (linguistic resources); it is possible thus to enhance translation units, dictionaries, and translation suggestions.

Mainly, the BEYTrans is open “as Wikipedia” online, any user or volunteer who is interested in translation may participate and enhance content. The only thing needed is registration.

In BEYTrans, linguistic resources are managed as follows:

a. Dictionaries and terminologies

- Import of existing linguistic resources (external -> internal) – internal refers to the environment;
- Construction of new resources during translation (internal -> internal);
- Exploitation of online dictionary (external -> external; ex. “Wordreference” dictionary).
b. Translation memories

All translation are saved, and suggested when a match is found. BEYTrans has a memory base where source and target sentences are saved together after any translation. For any comparison a score is computed. The suggestion threshold is tunable (ex. 50%, 80%, 90%, 100%).

c. External helps: machine translation

A machine translation (MT) systems translates your text automatically from source into target language. It proposes a first translation draft that can be enhanced by human translator. This process is called “post-edition”.

The MT systems used are not directly supported by the BEYTrans server. Online machine translation servers (ex. Google Translate) are called for getting suggestions. BEYTrans calls machine translation systems, and store their results.

Functionalities

Registration, users and communities

It is possible to check and read documents in many languages without any registration, but if a translator wants to practice translation, she/he needs to log as a registered user. Registration asks only minimal information: first name, last name, login and password. After the registration is done, translation functionalities become accessible. Registered users use many functions: importing documents, viewing, checking, post-editing, enhancing and checking dictionaries, and of course using available helps during translation. In the main page of the environment, in the top-right area, the “Register” link allows registering new users.

---

**Figure 1: Registration access**

The “Register” link lead user to the following page:

**Figure 2: Login creation**

---
Once the “login” information is input, save it by clicking on the “Register” button. Now it is possible to login into the environment BEYTrans.

By clicking the “Login” link (top-right), the translator gets the following interface:

![Figure 3: Login interface (user must be registered before logging)](image)

Let’s assume the username is “translatorguest”.

![Figure 4: Login creation](image)

To get the page introducing of BEYTrans, just click on the BEYTrans Logo! (Top-left side)

![Figure 5: Account creation](image)
Translation projects

- **System functionalities**: accessible from the main menu

  - « edit » or « translate » are for accessing system pages and translation text during viewing.
  - « Inline pages » allows importing a document in an existing project or creating a new project if it doesn’t exist!
  - Linguistic resource manipulation

  ![Figure 6: Pages (documents) functions](http://87.98.171.238:8080/beytrans/)

- **Importing documents in existing project**

  - Existing project name (1)
  - New project name if it does not exist. (2)
  - Document name: will be displayed in the project
  - Source language of the document
  - Source of the publication
  - Original author
  - Existing translation project name
  - Translation copyright
  - Text importing from the original web location – this task is done manually

![Figure 7: Preparing the communities and importing pages](http://87.98.171.238:8080/beytrans/)
Introduction project page

Each translation project has its own “home page” called “Web Home” which introduces the project itself: goal of translation, text source, source and target languages (translation), author and copyright... (See Figure 8). When a new project is created, its “Web Home” is created systematically. Translators can modify it by choosing the “Translate/Edit” submenu item (Figure 6).

Importing documents in new project

Any translation project must contain at least one document. To create a new project, a new project name has to be introduced rather than choosing an existing project. After import, the project, documents and the introduction pages will be displayed in a specific space.
Search your document by project

A project -> search displays the search form. The community list is detected and shown automatically. One can access the list by introducing the name of community (project translation) and the name of a document (if not the first char of the name). Note that “%” allows displaying all documents of the selected community.

Document management

Creation of communities and import of documents

Figure 7 shows the main interface for creating and importing pages. If it is subject of new translation project, the name and imported document are inserted in the right panel. The BEYTrans creates the new project systematically and attach to it different functionalities (translation, display in different formats, etc.).

The creation of a new project is illustrated in Figure 10:
The validation of the information introduced in interface of Figure 10 creates the project Paxhumana illustrated in Figure 11. The document and project description page are associated to the new project “Paxhumana” (check the web site of BEYTRans).

The Paxhumana project is created with two main documents: (i) imported document “whiteribboncampaign” and (ii) “Web Home”. This last document might be modified to introduce more information about the project (conditions, text type, community, original source, advices and helps,...)

*Viewing document in several formats*

When you select “Paxhumana” it opens a Tab which contains all documents of the translation project. If you want to translate any document, just click on it. You will get the following interface:
**Bilingual parallel documents: segment/segment**

- Source language
- Target language
- Source segments
- Target segments – no translation is done yet!
- Number of translated segments in target language
- Source segment amount
- Display formats.
- Source language of document
- Target language
- Display formats.
- Source language
- Target language
- Source segment amount
- Number of translated segments in target language

Figure 12: Tuning displaying format (before translation)

Figure 13: Segment/segment format
* Bilingual parallel documents: document/document

The interface presented in Figure 14 allows displaying text in: (1) document/document parallel format, (2) segment/segment parallel format.

It is also possible to tune this interface for readers. People who want to read text in any existing language, they may just choose source language and keep target language empty.

**Translation editor**

*How to translate?*

![Figure 14: document/document format](image)

![Figure 15: document/document format](image)
Choosing source and target languages: case (a)

The interface above (Figure 16) is often empty when accessing for the first time. After choosing the source and target languages, the result is displayed in Figure 17.

The interface above (Figure 16) is often empty when accessing for the first time. After choosing the source and target languages, the result is displayed in Figure 17.
Choosing source and target languages: case (b)

![Figure 17: Translation editor](image)

**How to save segments and their translations**

Segments are saved after any modification!

**MT suggestion, internal dictionary suggestions, online dictionary suggestions, fuzzy match suggestions**

To get suggestions from linguistic resources select (mouse right click) and ENTER. By selecting MT suggestion + ENTER, the Editor copies the suggestion in the current target cell.
Once the post-edition is finished, translators can click on any other cell to save the new translation. The bold characters are used to show that a segment (a row of the table) has not been post-edited. To directly translate it or to post-edit an MT result, double click on the target cell. The editor opens an editable area. Once the post-edition is finished, translators can click on any other cell to save the new translation.

** Tokens are automatically detected in the source segment. If a translation exists in the community’s dictionary then it is displayed, else the token is input as new headword in dictionary. Both existing and new headwords are subject of enhancement.
The “exit” button redirects the user to the viewing interface (see Figure 19).

** This is caused by import process. There is match between 2 sources sentences but the target sentence is empty, this situation can be useful for checking existing duplication in resource, and can also hidden when target sentence is empty.
* Using existing translation

Each community has its own dictionary and linguistic resources. By choosing source and target language, community and search expression, a list of dictionary entries is displayed. For enhancing the content and introducing new translations double click on the target cell. An edition area will be displayed. For saving it just click on other cell. The new translation will be suggested to other translators when translating segments containing the same word (Figure 21).

*Enhancing content when translating

Translation of words can also be introduced when translating. Information like domains is proposed by the editor (Figure 22).
Figure 22: Enhancing community dictionaries during translation

Main features

- 100% multilingual
- 177 languages with more than 200 fields domain for dictionary fields
- Text, HTML, PHP and PO file format management
- Automatic segmentation of the raw text in Indo-European languages (Processing Asian languages is under development)
- Transitive translation (English -> French -> Japanese): target language can be managed as source for other languages (EN (source) -> JP(target/source) -> FR(target))
- Machine translation capabilities + post-edition
- Translation memory suggestions + post-edition: the initial TM contains 15,378 thousand translation units translated completely or partially in more than 15 languages!
- Semi-proactive translation (background MT translation)
- Multilingual dictionaries management: construction and enhancement of dictionaries whole translating is possible
- Wiki-based content management: the current version is implemented over the java-based XWiki wiki, future versions will be developed under Tikiwiki (PHP) (under construction)
- Collaboration at all level of content
- Translation projects management
- Translators management
- Displaying documents in different formats: parallel segment/segment, source or target document, parallel document/document
- Exploitation of online dictionaries
- Automatic suggestions during translation
- No worry about losing data, the system works as a wiki. All or part of the data may be restored easily.
- Possibility of switching between languages without changing the interface

Recently added (before release 2.1)

(1) Managing date/time of translation and the source of translation (MT or Human)

![Figure 23](image)

Users’ login names are saved for any segment translation according to selected target language. Date/ time of translation are also kept with each segment. This information will be used in the future to compute statistics. Note that ‘MTH’ contains information about translation; if a translation is a MT suggestion; it takes the value ‘M’ or ‘H’ if it is the result of human post-edition or direct translation.
(2) Functionalities menu

MT translation as background tasks

Examples above show how to perform translation and post-edition one segment at a time. For enhancing efficiency and saving more translation time, it is also possible to run MT as a background task performed asynchronously. In parallel, the human translator may translate while he/she is getting one after one MT suggestion.

Example

Step (1) - Call MT as background task by using **MT -> background**

Figure 24: MT translation of remaining segments
Step (2) - display progressively MT translations by using the **Refresh** function

![Figure 25: Enhancing community dictionaries during translation](image-url)
Step (3) post-editing, enhancing MT translation

In some cases, translators need to copy the source content to the corresponding target cell. This function allows reducing time for editing and translation as shown in the following example.

**Step (1) Initial status**

```
44 Block description: -
```

**Step (2) after copying**

```
44 Block description: Block description:
```
Step (3) modification of source (same method as for target)

<table>
<thead>
<tr>
<th>Block description:</th>
<th>Block description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admin content</td>
<td>Bloc de description:</td>
</tr>
<tr>
<td>Show comments</td>
<td></td>
</tr>
<tr>
<td>Hide comments</td>
<td></td>
</tr>
<tr>
<td>Create or edit a file gallery using this form</td>
<td></td>
</tr>
</tbody>
</table>

Figure 27: post-edition of MT suggestion

**Switching languages**

The editor allows translators translating and modifying target segments, but not source segments. For modifying the source, it is possible now to switch languages. The source becomes target and vice versa. This function allows changing source content if language related errors are detected in the source (Figure 28 shows how to switch source and target segments for modifying the word “display” into “display”).

Figure 28: switching language for manipulating source segments
Managing translation time

**Start clock**: In order to activate translation timing, this function displays (in seconds) the progression in translation according to time duration. Timing activation depends on each user session and target language. Duration is computed according to each source language.

**Stop & Save clock**: If translator wants to take a pause, or if translation is finished, then translators might be able to save time duration.

**00:00:04 | Total: 00:00:26**: This formatted text displays the time duration. In red is the progression in seconds. The blue text time indicates the whole time spent of translation for the target languages!

**Chat**: The chat allows translators to have direct discussions. This function has been tested efficiently when the environment has recently been experimented.

![Chatting system](image)

**Figure 29: Chatting system**

Under development

- Processing Asian languages
- Ergonomic enhancement of interfaces
- Importing more free dictionaries and enhancing matching of sub-segment content
- Automatic dictionaries construction from MT
- Import of parallel translation resources
History of the project

- The BEYTrans system started on 2006. The first experimentation of the prototype was done in November 2007 with the collaboration of Francesca Marzari, DEMGOL project – Trieste University
- Enhancement during one year (2008) – technical aspects
- Bugs detection and ergonomic aspect development – several months on 2009 – after PhD defense.
- 2009 June first official production version. Switching from a prototype to a operational system (1 years of development).

Contact

Official site admin: Youcef BEY - M@il: youcef.bey (at) lts-tech.com

First interested volunteers

- Queralt Valls: Catalane, English, Dutch, Spanish (queraltvalls (at) hotmail.com) – Spain
- Sarah Edwards: Japanese, English (tappingcat (at) gmail.com) – Japan
- Alia Mourad: Arabic, English, French (mouradal (at) simula.no) – Norway
- Youcef BEY: Arabic, English, French (youcef.bey (at) lts-tech.com) – France
- Amer Radwan : Arabic, English – United Arab Emirates (UAE)
- Zilola Yuldashiva: Russian, Uzbek, Tajik, English, French – OHIO, USA

Experimental contribution

- Mutsuko Tomokyo: French, Japanese, English, LIG-GETALP, France
- Jean-Claude Durant: French, English, LIG-GETALP, France

Author of this tutorial

Yousef BEY - http://panflute.p.u-tokyo.ac.jp/~bey/