myCat for the Translation Centre

USER MANUAL

1. Context

myCat is a Computer-Assisted Translation (CAT) tool which has two essential uses:

- It allows searching for terminology in previous documents and shows how terms were translated in their complete context; and

- It detects whether some parts of a text to be translated were already translated in a previous document, and displays that reference document along with its translation.

Thus myCat is a tool which facilitates the re-use of previously-translated documents as a source of references for terminology and for whole phrases.

Please note that the version of myCat you are about to use is a beta-test version, *i.e.* it is entering its serious testing phase and thus still bears several defects and bugs. A bug report for myCat is available for beta-testers in order to report in a structured way the various issues you could meet when using this tool.

2. Accessing myCat

myCat is a web-based tool: it can be accessed like any web site by entering its URL (address) in the navigation bar of a browser.

Open any browser (myCat supports Internet Explorer 8 or later, Mozilla Firefox 3.6 or later, and Chrome 8 and later). In the navigation bar, enter this URL:

http://62.147.229.43:5555/TranslationText

Version	Date	Author(s)	Status
0.3	6 June 2011	Jacques Guyot Karim Benzineb	Provisional

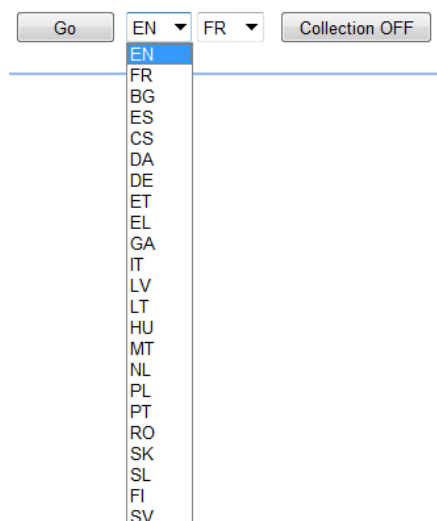You should see the following page:

You are now under the Text Aligner application, whose use is described in the next Section.

3. Using the Text Aligner

The Text Aligner is meant to retrieve terms or phrases from an existing corpus of previously-translated documents. It is extremely simple to use: The user defines the source and target languages, possibly a specific collection of documents (like Budget, etc.) and enters the term(s) to be looked for.

3.1. Defining the Language Pair

First define your source and target languages from the scroll-down menus at the top of the screen:

The 23 languages are indicated in the ISO-639 code :

EN: English	ET: Estonian	NL: Dutch
FR: French	EL: Greek	PL: Polish
BG: Bulgarian	IT: Italian	PT: Portuguese
ES: Spanish	LV: Latvian	RO: Romanian
CS: Czech	LT: Lithuanian	SK: Slovak
DA: Danish	HU: Hungarian	SL: Slovenian
DE: German	MT: Maltese	FI: Finnish
GA: Gaelic/Irish (*not available for the test*)	SV: Swedish	

Please note that in the beta version, only the language pairs EN-FR and EN-ES (and reverse direction) are aligned according to our new technology based on statistical machine translation. All other pairs are aligned according to a simple geometrical rule (*i.e.* their alignment should be of lesser quality).

3.2. Choosing Collections

The terms you are looking for may be very specific to a given type of documents, for example the Budget, or a technical report of a particular kind. You can limit the search to one or several collections by clicking in the corresponding tick-box in the Collection box:
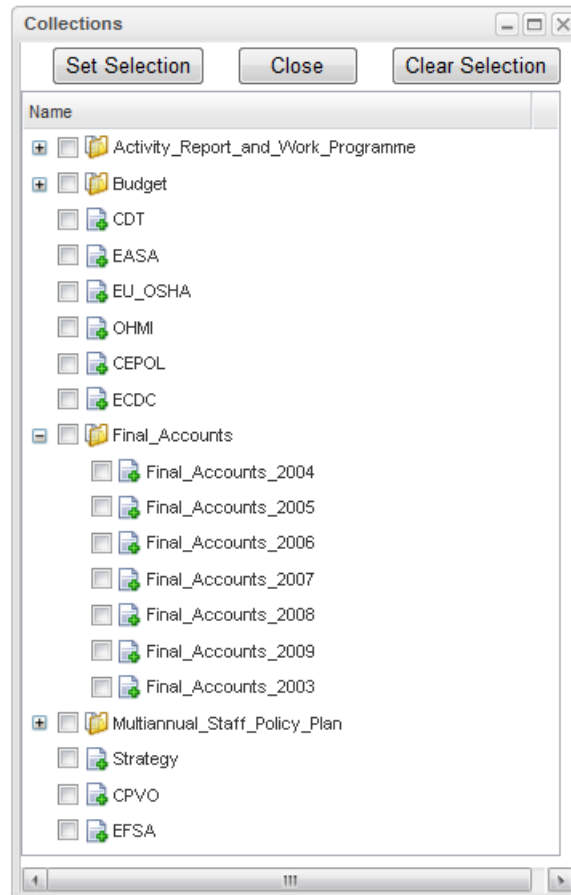

Some collections bear a **+** sign on their left side: it means that a number of sub-collections are available. You can expand the list of sub-collections by clicking on the **+** sign.

When all the relevant collections are ticked, click on the "Set Selection" button to validate your choice. If you click on "Close" the Collection box will be closed but your selection will not be saved. If you click on "Clear Selection" all the ticked boxes will be cleared so you can make other choices.

3.3. Searching for Terms

After you selected the relevant language pair and collections (which you normally do only once before translating), all you have to do is typing your term query in the text field and executing the search by clicking on the "GO" button (or hitting the ENTER key on your keyboard) :
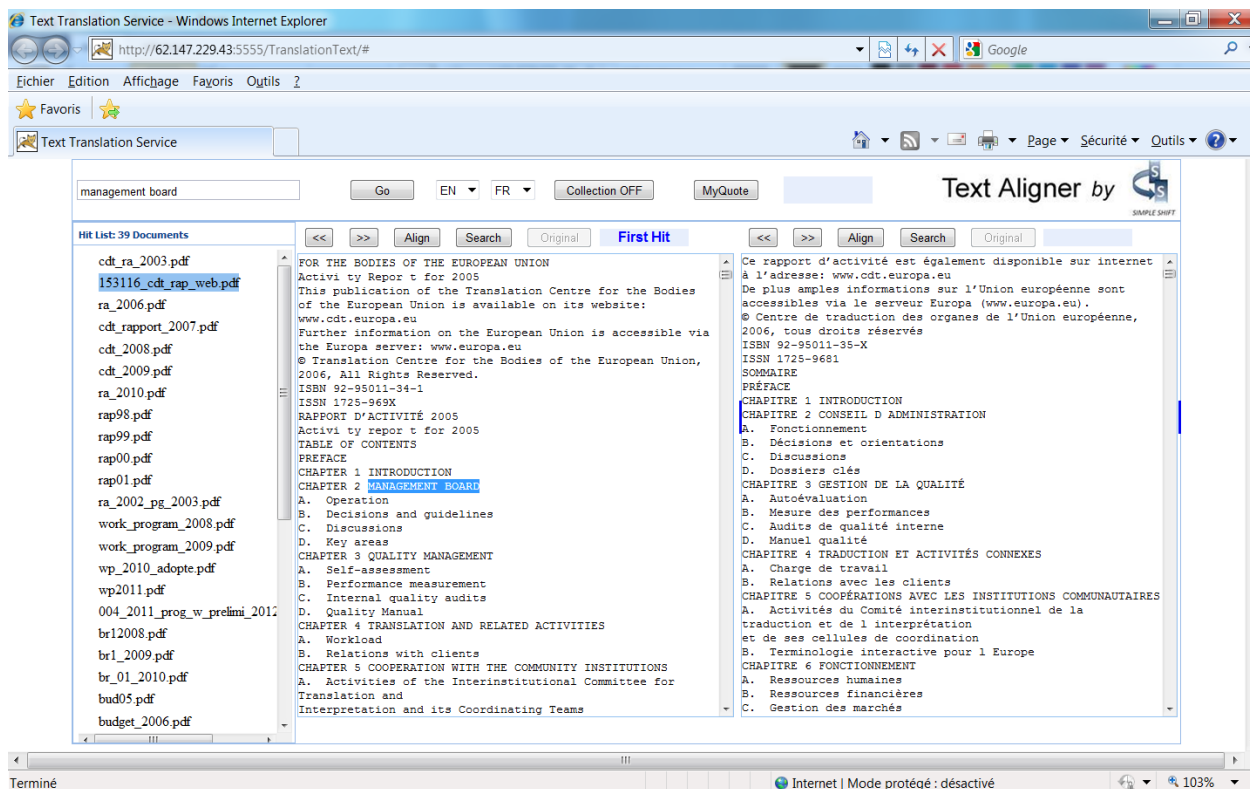
In the example above we searched for the expression "management board". The search produced 39 results, as indicated on the top of the Hit List frame.

The result list includes all the documents which contain the expression. If more than 200 results are found, the other results are discarded. (*Note: this maximum level may be changed by the application Administrator.*) You can navigate in the result list with the scroll bar located at the right of the frame.

The documents are listed by their file name. To display the entire path (*i.e.* the collection and sub-collections in which they are located), simply position the mouse on the file name; the complete path will be shown in a label as illustrated below. (*Note: this option may be de-activated by the application Administrator.*)
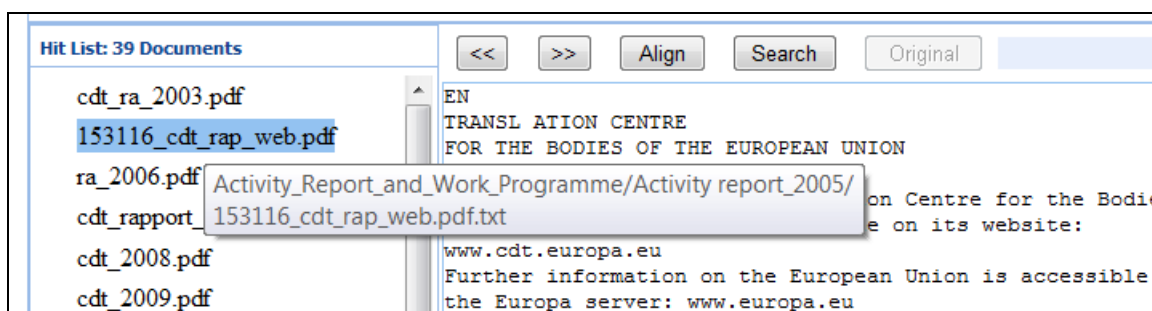
Clicking on any file name will display the source and target versions of the corresponding documents, and will align them on the first occurrence of the term in the text. That occurrence is highlighted in the source frame, and an indicative vertical blue bar shows where the corresponding term should be located in the target frame.

You can then copy the corresponding part of the target text and paste it into your current translation work. Copies can be performed through the usual CTRL+C keys, or through the browser's Edit/Copy menu.

3.4. Using the Navigation Buttons

The source and target frames have an identical button bar. You can navigate from one term occurrence to the next one by clicking on the **>>** button, and to the previous occurrence by clicking on the **<<** button.

When the highlighted term or expression is the first one in the reference document, the message "First Hit" appears in blue in the button bar of the source frame (as shown in the first screenshot in section 3.3 above).

If you use the navigation buttons, you will navigate from a hit to the next one until you reach the final hit in the reference document. In that case the message "Last Hit" will appear.
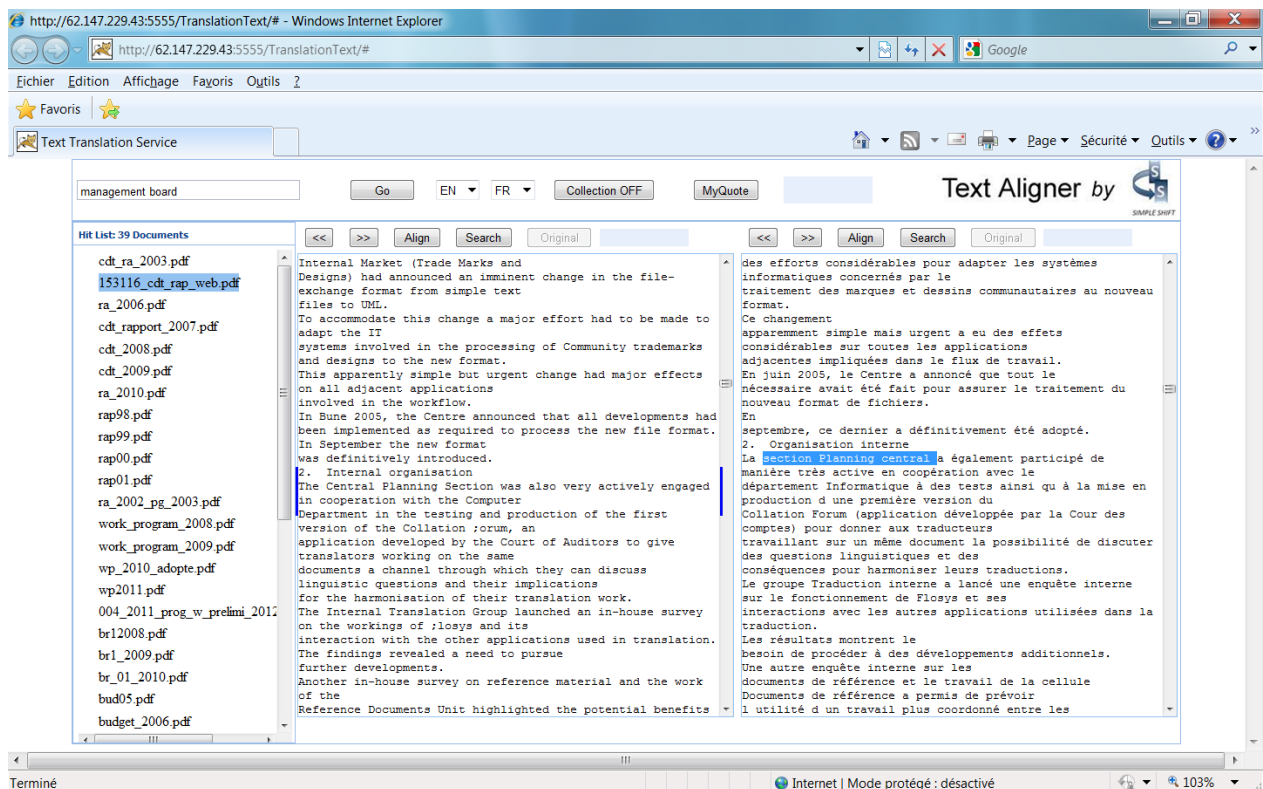
3.5. Using the « Align » Button

One of the core features of myCat is that the complete reference document is always provided both in the source and in the target languages. Thus you can always use the scroll bar on the right side of the frame to read more of any reference document. If you navigate far up or down in one version (source or target) of the document, you may want to display the corresponding part of the other version. In that case, simple double-click on the term you are interested in and then click on the "Align" button. The other version of the document will be automatically aligned.

Please remember to double-click a word in the text in which you are interested, otherwise the other version will not be aligned (because the system will not know on which part of the text it should focus).
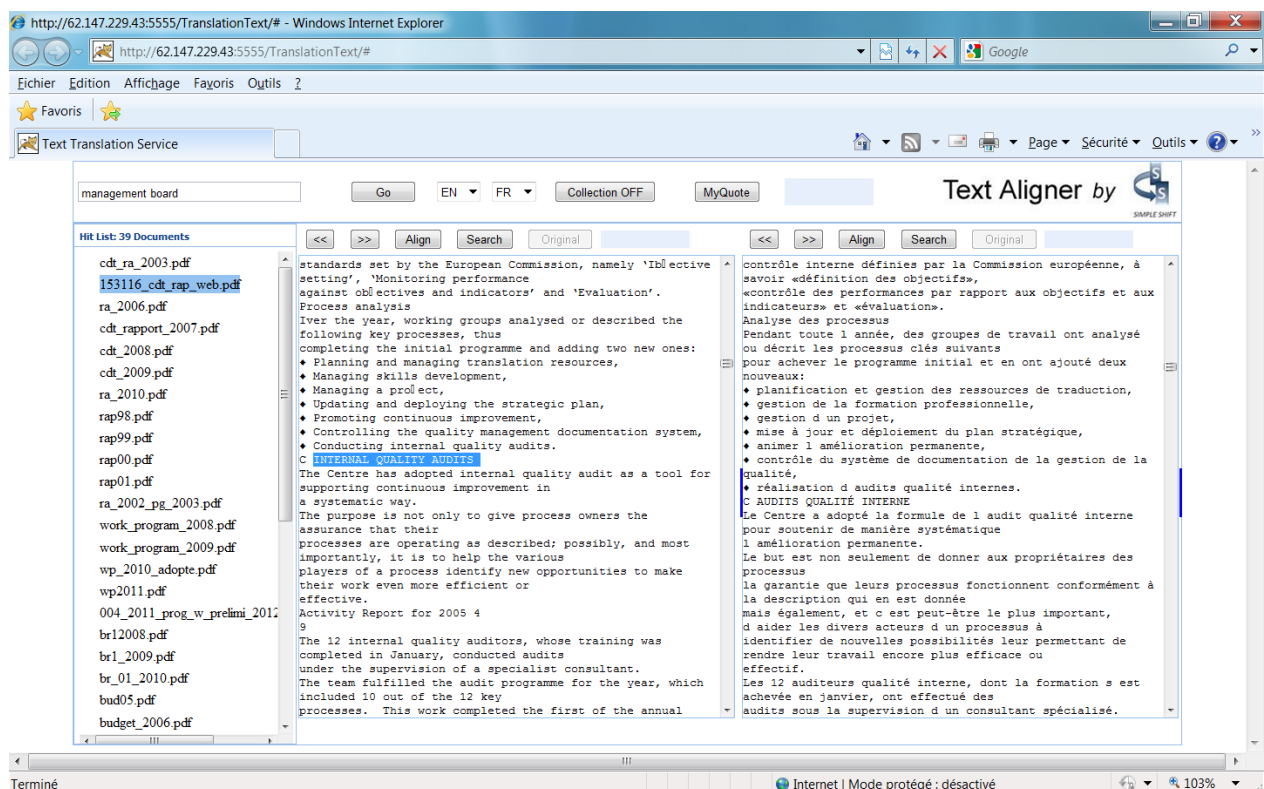
In the example above we highlighted the expression "Internal Quality Audits" and clicked on the "Align" button. The corresponding part of the target version was then displayed and correctly aligned.

Conversely you can highlight a term and click on the "Align" button in the target frame; then the source frame will be aligned, as illustrated below:

Here the highlighted target term was "section Planning central"; the "Align" button allowed displaying the corresponding part of the source document.

3.6. Using the « Search » Button

Some reference documents may include a lot of useful terminology for your current translation work. If such is the case, you may not want to perform again a complete search under the Text Aligner, but rather limit the search to the document at hand.

To do so, click on the "Search" button. A text field will appear, in which you can type your new query:
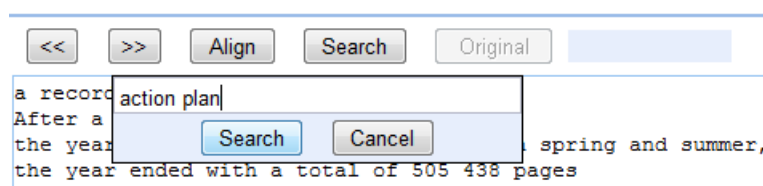
In this example we entered a new search on the expression "action plan". To execute that search, click on the "Search" button. (Clicking on the "Cancel" button will simple close the box without any further action.) The result is displayed below:
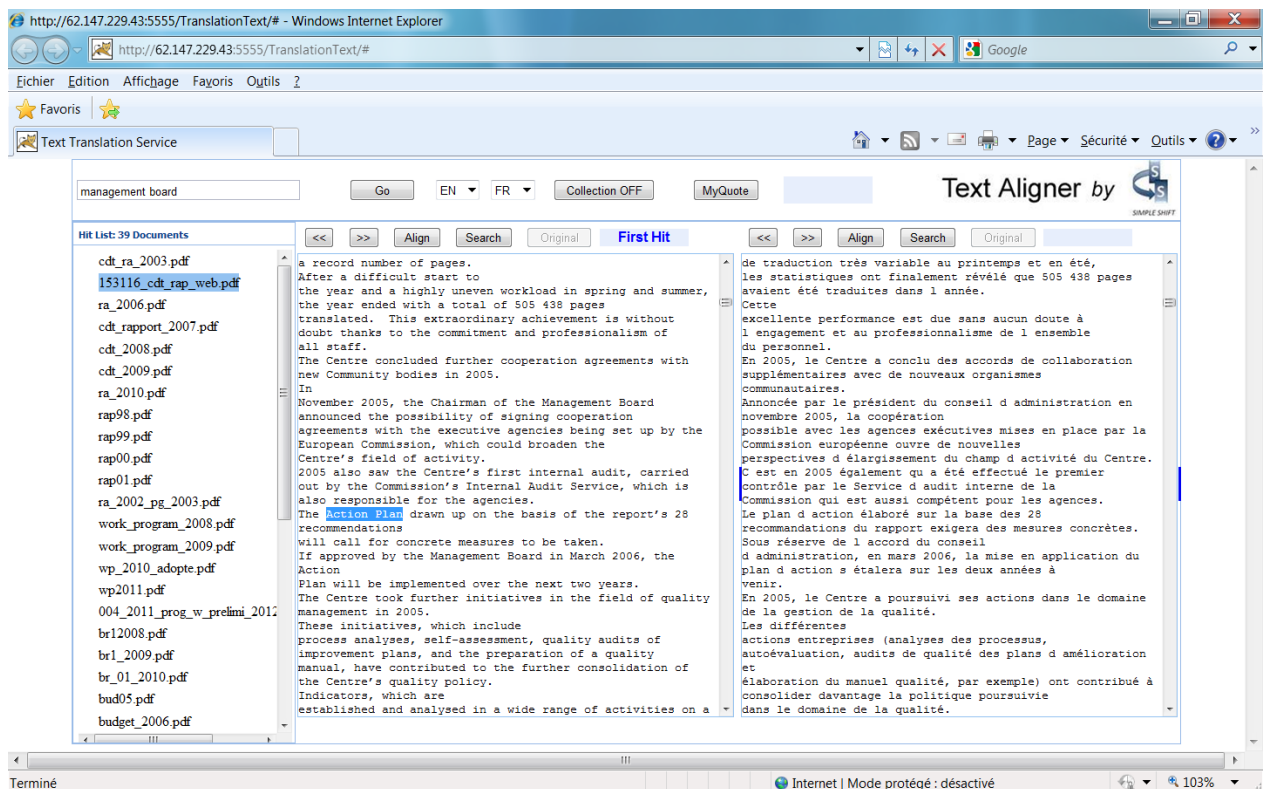
Note that we are still in the same document (153116_cdt_rap_web.pdf) as in the previous screenshots, but the highlighted term is now "Action Plan". You can use the navigation buttons to see the other occurrences of that new expression.

3.7. Using the « Original » Button

This button is not active in the beta version.

The "Original" button will allow displaying the document in its original format (generally MS-Word or PDF) so the user can see and possibly import the real page layout.

3.8. Boolean Operators and Other Search Tips

Below is a list of the Boolean operators (i.e. words which allow for various types of search operations) available under myCat.

Three types of requests are possible under myCat:

- term1 term2 …. termN: Search for a single term or a complete expression or phrase. No Boolean operator and no quotation marks are needed. All the terms are searched for and highlighted. Thus the AND Boolean operator is implicit.
- term1 OR term2: Only one of the terms is included in the hit document. This operator is useful to find a term which can be spelled out in various ways (such as "labour" and "labor", "organization" and "organisation", etc.)
- NEAR term1 term2 : Both terms are included in the hit document and they appear at a maximum distance of 15 words. Not-indexed words (see below) are excluded from that count. (*Note: This limit of 15 words may be changed by the application Administrator.*) Only the first term is highlighted.

Below are a few additional search tips:

- Search is case insensitive
  - Example: term = Term = TERM

- Stop Words (not-indexed words)
  - Articles, prepositions and other very short and common words in English and French are not indexed, in order to streamline the index size and improve the reply time. Thus words such as le, la, les, du, des, the, of, etc... will not be found.
  - (*Note: the content of the stop words file may be changed or entirely deleted by the application Administrator.*)

3.9. Using the « Quote Detector » (MyQuote) Button

If you want to leave the Text Aligner environment and access the Quote Detector, simply click on the "Quote Detector" (or "MyQuote") button. If you want to keep both environments open at the same time, you can right-click on the "Quote Detector" button and open it in another tab.

4. Using the Quote Detector

The Quote Detector was designed to analyze a new document before it is translated, so as to identify all the parts (from a few words to entire paragraphs or pages) which already appear in a previous document, and for which a translation already exists.

The Quote Detector environment looks like this:
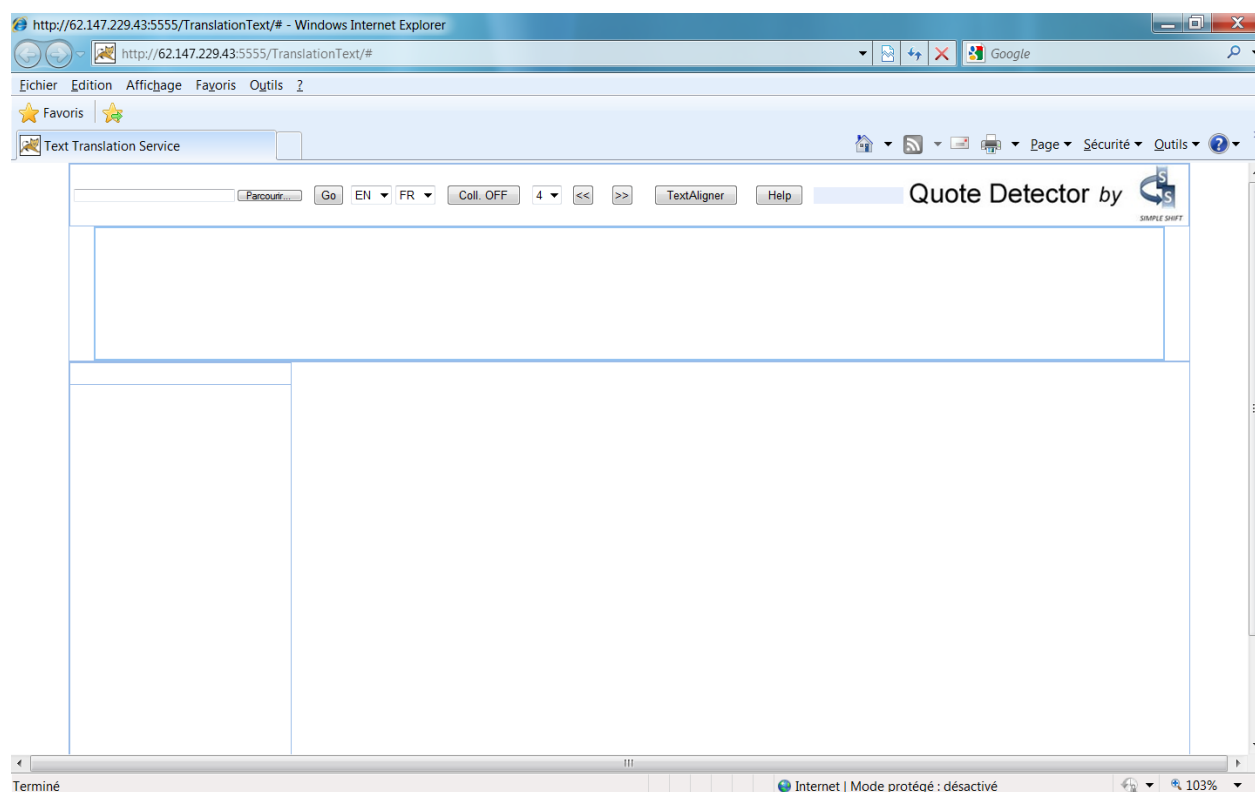
Note that the language scroll-down menus and the Collection button are still available; you can use them as described under the "Text Aligner" part of this manual.

## 4.1. Converting the document to be analyzed into Unicode UTF-8

Before submitting a document to be analyzed, that document <u>must be converted</u> into TXT and Unicode UTF-8 format. **(*Note: This constraint is only necessary for the beta-test; in the final version the conversion will be done automatically.*)**

The simplest way to convert your document into Unicode format is to open the document in your text processor (such as MS-Word or Open Office Writer) and to save it in the required format through the "Save As..." function.

In the example below, a document is open under MS-Word 2010 and the "Save As" function is activated from the "File" menu. The Type option allows choosing the "Plain Text (*.txt)" format, as shown above (this is the French version of Word so the option is called "Texte brut") :
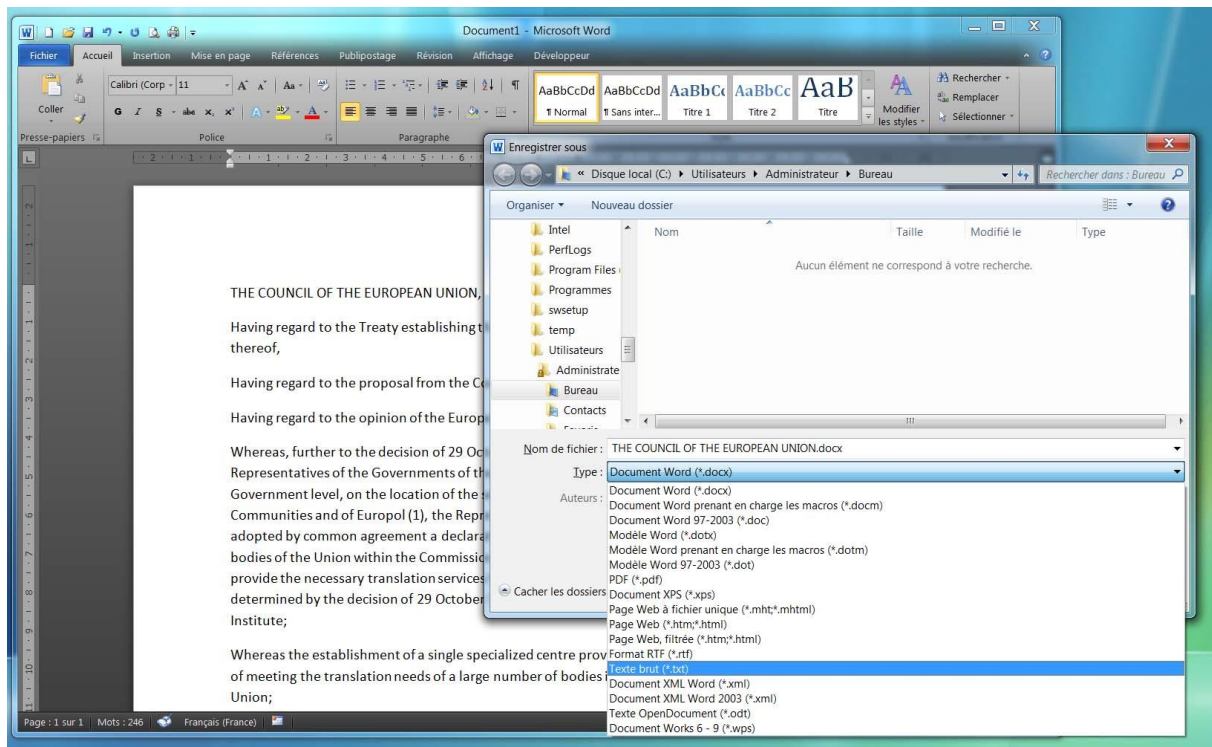
When selecting "OK" to save in TXT format, MS-Word proposes several encoding formats for your plain text file :
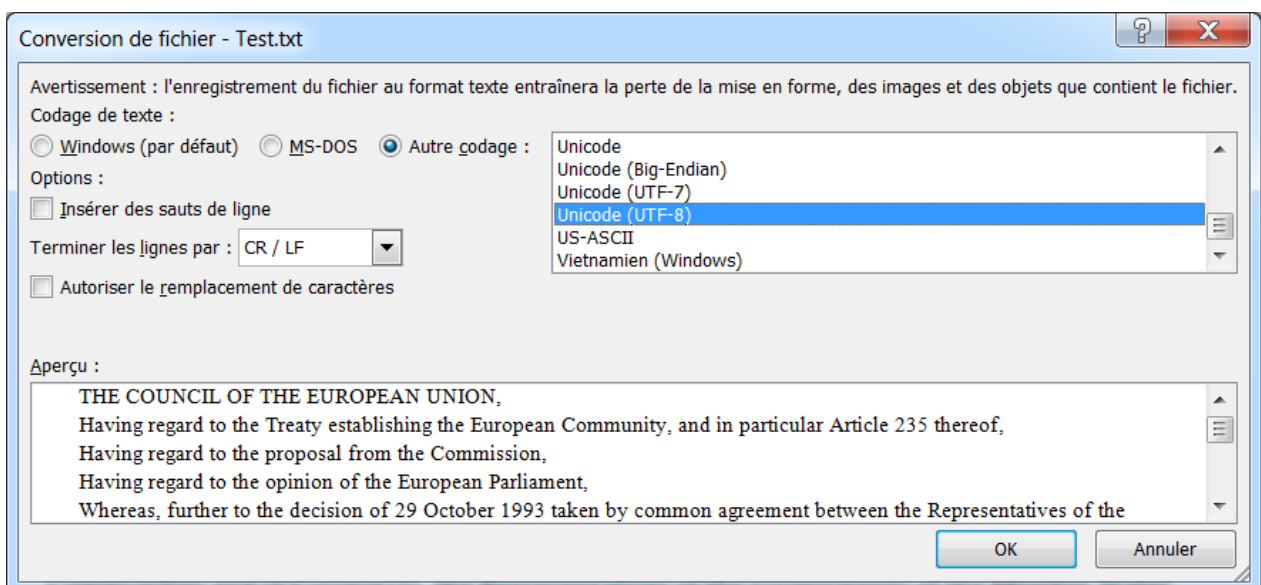

Choose the "Other encoding:" (in French: Autre codage) radio button and in the list, scroll down until you find "Unicode (UTF-8)". Then click the OK button : your file conversion is completed !

Please note that it is compulsory to choose the UFT-8 format: if you choose Unicode alone, or UTF-7, the Quote Detector will probably not be able to analyze the document.

4.2. Launching the Quote Detection

To launch the document analysis, click on the "Browse" button in the top left corner of the Quote Detector. A standard Browse box will open:
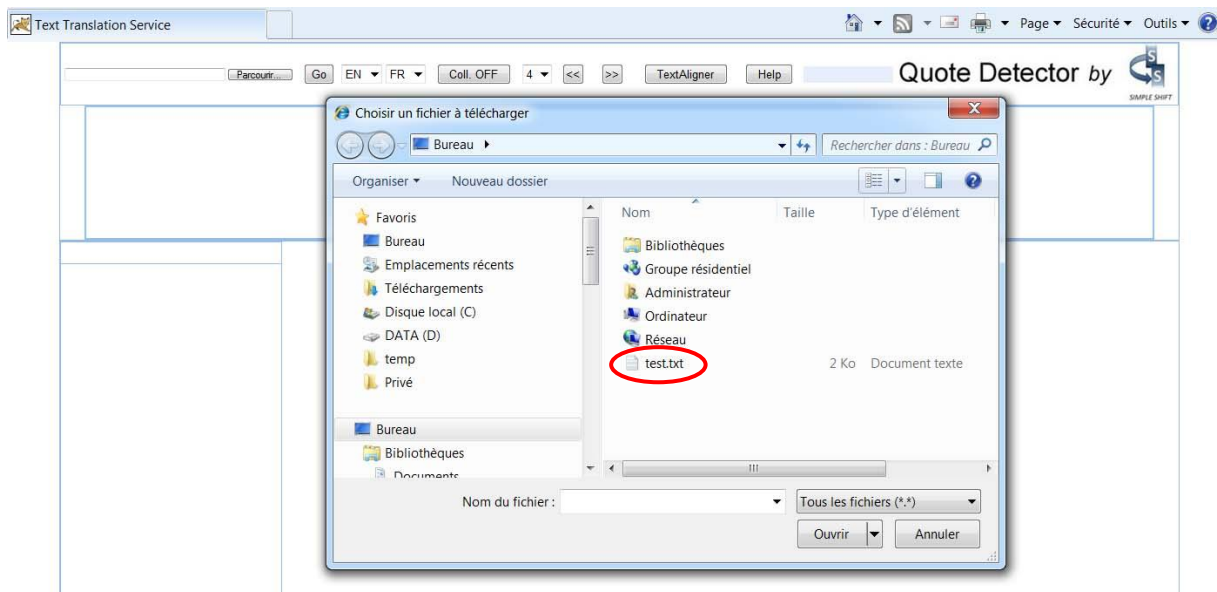

In that Browse box, go to the location of the document to be analyzed (in this example it is located on the Desktop and is called "test.txt") and click on the "Open" ("Ouvrir") button. This operation is completed when the "File uploaded" message appears after the "Help" button, as shown below:
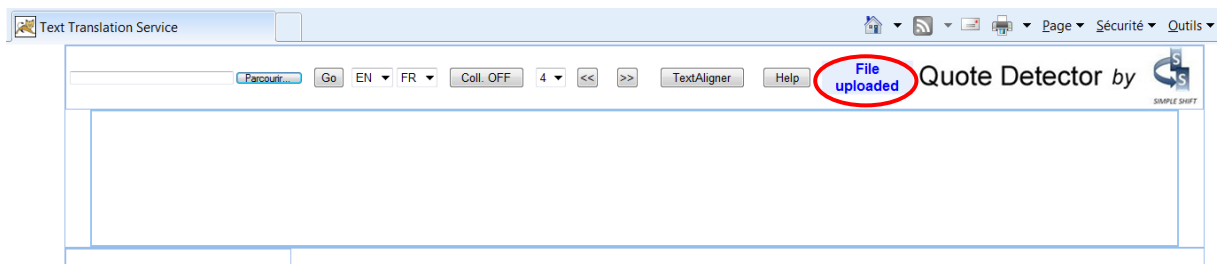

Before launching the analysis, you can select which source and target languages you are working in, and possibly select one or several collections.

You can also choose the minimum number of words you would like to see in the detected quotes. This function avoids detecting very small quotes, which could be very numerous but not very useful. The lowest number is 4 consecutive words; the highest is 9. Please note that this highest possible value only applies to the <u>minimum</u> number of words to appear in a quote; there is no maximum length for the quote. In fact the system always looks for the longest possible quote; thus if the whole document was already translated it should appear entirely in blue (*i.e.* in hyperlink).
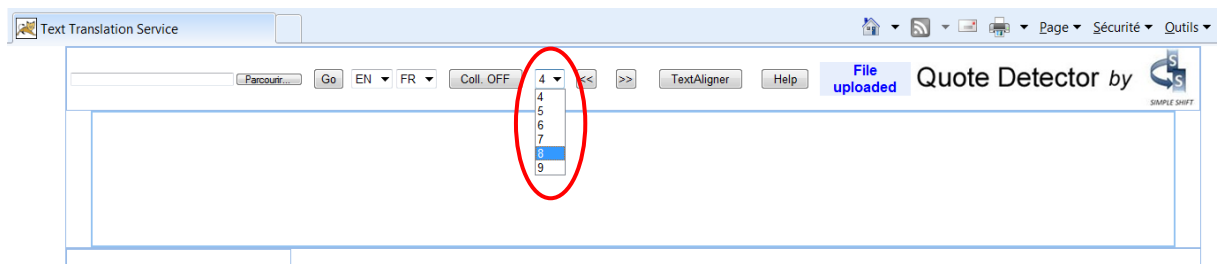
A possible strategy is to start out with the highest minimum number of words in the quotes, and to reduce that number if the system doesn't find any (or many) quotes.

When all the parameters all set up you can click on the "GO" button to launch the analysis. The task is completed when the "Content loaded" message appears next to the "Help" button and the document appears in the top text frame:
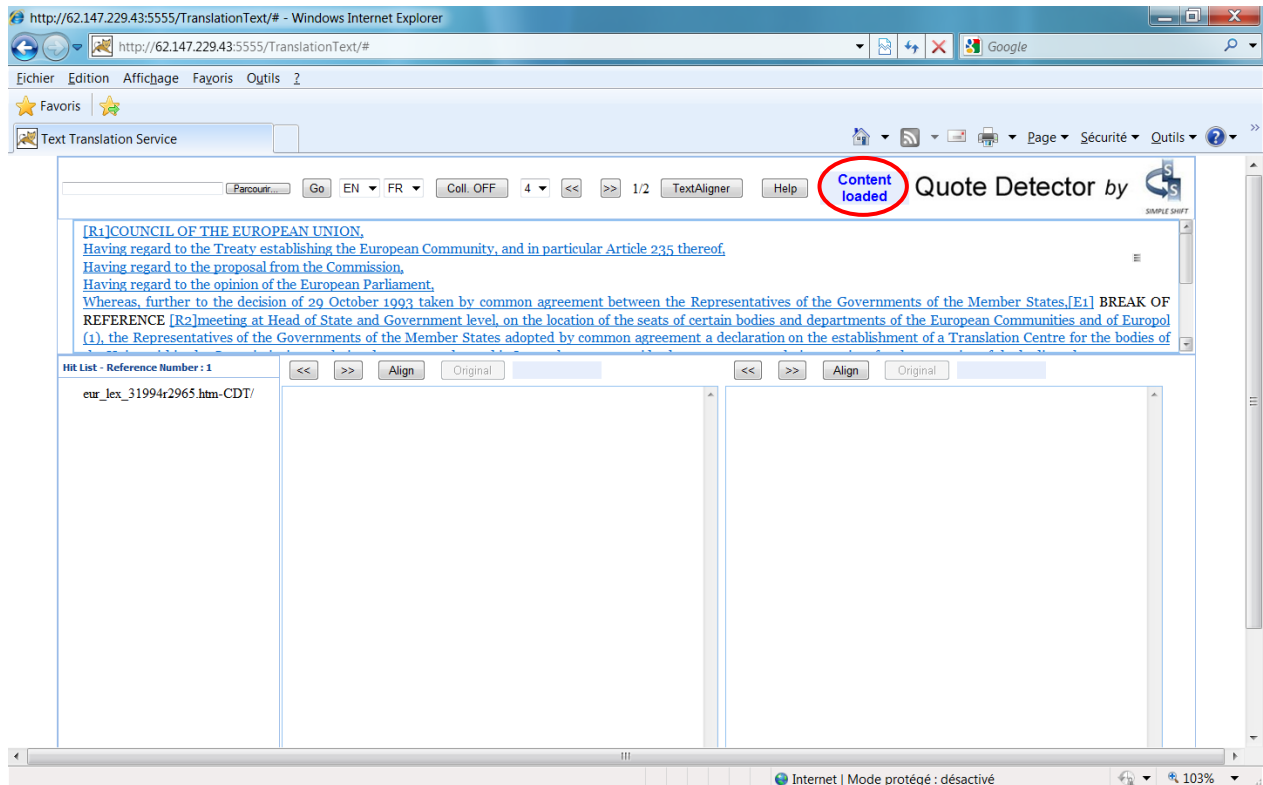
In the example above you can see that the parts of text for which a reference was found appear as hyperlinks (in blue and underlined), while the unknown parts are displayed in black (the expression "BREAK OF REFERENCE" which was introduced on purpose to test the application).

The quotes are numbered as [R1], [R2], etc. You can click on any of them to display the list of documents which contain them. (Note: the list of documents containing the first quote is automatically displayed after the document analysis is completed.)

As with the Text Aligner, clicking on a reference file name in the Hit List displays the content of that document in the source and target versions, and aligns both documents on the quoted part:
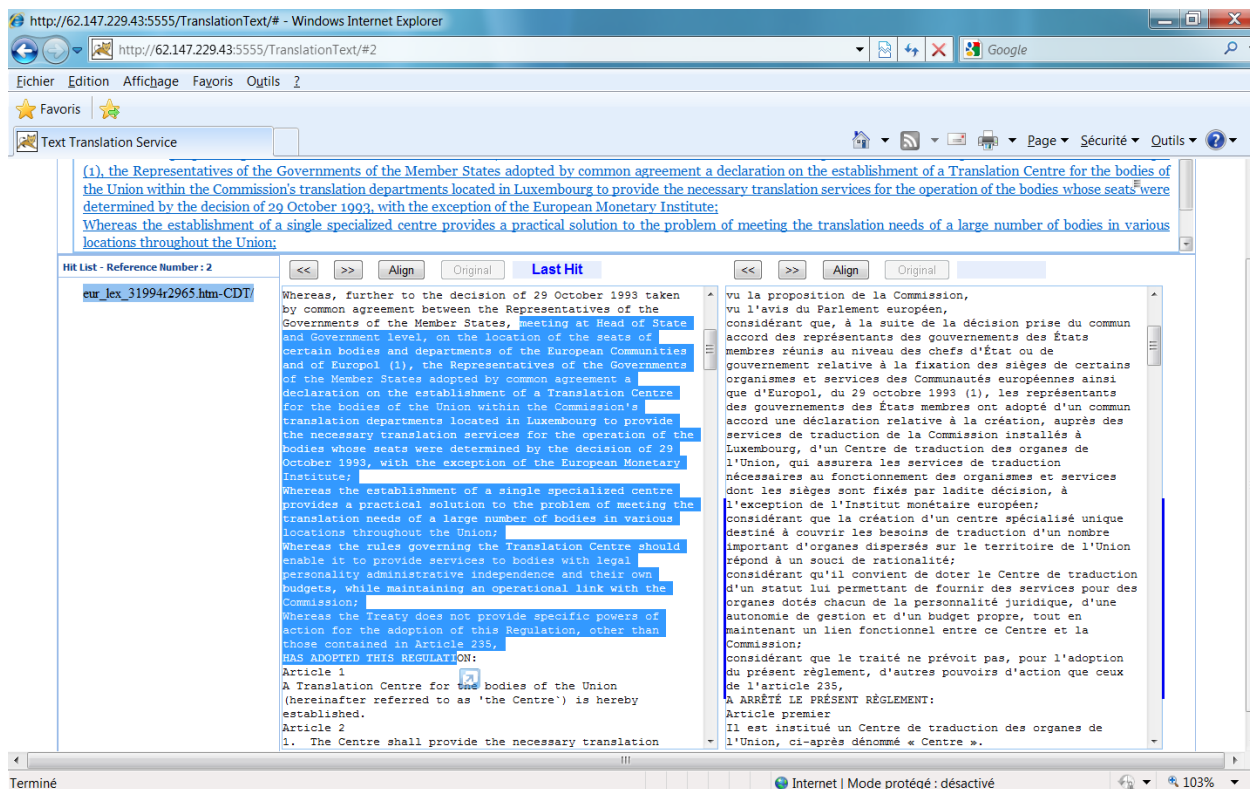
You can then copy the corresponding part of the target text and paste it into your current translation work.

The navigation buttons **>>** and **<<** allow an easy transition between the various quotes found for the analyzed document. The number of the quote currently displayed and the total number of quotes found for the document are also displayed next to the navigation buttons:

The button bar also includes a "Text Aligner" button which allows going back to the Text Aligner program (but again, you can open each program in a separate tab so that both of them remain available at all time during your work session).

Finally, a "Help" button opens an electronic version of this User Manual, so you don't have to keep a paper version on your desk.