



WeBiText Tutorial and Exercise: Level I

Introduction

WeBiText is a prototype online bilingual concordancer developed by Alain Désilets, Benoit Farley and Marta Stojanovic of the National Research Council Institute for Information Technology. It can help a user to search a million Web pages taken from the Canadian federal government site, www.gc.ca (Désilets 2007: 11), as well as pages from the European union site, www.europa.eu, and to display occurrences of English search words or other character strings and their translations in French or in other languages in context. Because WeBiText is an online tool, you can use it anywhere you have access to an Internet connection and a Web browser.

The results are presented in bitext format: the sentence in which the search word is found is aligned with and displayed beside its translation. If you wish, you can also see the entire aligned text and translation.

This version of WeBiText is a prototype, so its functions are still evolving. It is freely available temporarily, while it is in the development phase.

NOTE: The tool's developers would like to hear your thoughts about WeBiText. If you would like to send them some feedback about the tool or its use, you can do so by clicking the **FEEDBACK** link on the WeBiText search page, or by sending an e-mail to Alain.Desilets@nrc-cnrc.gc.ca and putting "Feedback from the WeBiText site" in the subject line.

You can learn more about WeBiText on the project home page, www.webitext.com, on one of the developers' home page, http://iit-iti.nrc-cnrc.gc.ca/personnel/desilets_alain_e.html, and read more discussion about it in the article *Translation wikified: How will massive online collaboration impact the world of translation?*, available [here](#).

Doing these exercises will help you learn to:

- [Search in Canadian government documents using WeBiText](#),
- [Navigate within the WeBiText results](#),
- [Compare search results from different sources](#), and
- [Answer some translation questions using WeBiText](#).

Getting ready

1. Open the Web browser of your choice (e.g., Mozilla Firefox or Internet Explorer) by double-clicking the icon that appears on the desktop (or from the Windows **START** menu).
2. Go to the WeBiText home page (www.webitext.com) by typing the address into the browser's address bar, or by clicking the link.

3. Read through the introductory text, and then click the WeBiText link that appears in the *Description* section.

NOTE: The WeBiText home page also hosts the tool WeBINuk, a bilingual English-Inuktitut concordancer that works using similar principles. You can access WeBINuk by clicking the corresponding link in the *Description* section of the home page.

Searching Canadian government documents online using WeBiText

1. The WeBiText search page appears. Read the introductory text on the page, which includes your options for searching.
2. Click the arrow that appears beside the drop-down list of available sites (called corpora). A number of sites are offered, and each is labeled with an indication of how quickly searches can be carried out for that site. (When working in larger groups, such as in a course or a lab, it is important to use the faster sites such as Health Canada.)
 - a. The *Government of Canada* option allows you to search within the whole Canadian federal government site, www.gc.ca. Because of its size and other factors, searches can be quite slow.
 - b. The *Health Canada* option allows you to search within the Health Canada site, www.hc-sc.gc.ca. This site has been prepared for faster searching.
 - c. The *Genomics* option gives you access to a number of different sites on the subject.
 - d. The *Government of Ontario* option allows you to search the Ontario provincial government site, www.gov.on.ca.
 - e. The *Gateway to the European Union* option allows you to search the European Union main site, www.europa.eu.
 - f. The *Federal Authorities of the Swiss Confederation* option allows you to search the main site of the Swiss Confederation, www.admin.ch.
3. Click the arrows that appear beside the drop-down lists to see the lists of source and target languages available. Note that not all language combinations are available for all sites; it depends on the official languages of each organization and the pages available in each language. For example, for the Government of Canada site, only English-French and French-English bitexts are available.
4. Do a search for the string *vendor* and its French equivalent(s) in the Government of Canada corpus:
 - a. Type *vendor* into the search field,
 - b. Ensure that **ENGLISH** is selected in the first dropdown list and **FRANÇAIS** in the second,
 - c. Ensure that **GC.CA GOVERNMENT OF CANADA (SLOW)** is selected in the second dropdown list, and
 - d. Click the **SEARCH** button that appears beside the search field.

5. Wait for a moment while WeBiText searches the gc.ca site for occurrences of this term. Note that this may take some time, as the tool must go through all of the pages searching for occurrences.

NOTE: If many users are working with WeBiText at once, the searches may be slower, so you may have to be patient!

Navigating within WeBiText results

1. After a moment (please be patient!) the results of the search will appear on screen.
 - a. Under the search field the total number of occurrences found is displayed. WeBiText displays a maximum of 200 occurrences, but also tells you how many were found in total.

NOTE: These are the results WeBiText could locate and prepare for display within the time allotted for the search (so as not to keep you waiting too long for your results). You should note that these may not be all of the results available on the site you are searching.

- b. Under the total number of occurrences, information is displayed in two columns. On the left-hand side, sentences from English texts containing the search string are displayed. On the right-hand side, sentences from French texts that have been identified as corresponding to the English results are displayed.
 - c. Within the English sentence, the search string is indicated with brown highlighting.
 - d. Read through some of the English sentences and their French equivalents. *Can you identify the equivalent(s) used for the string *vendor*?*
 - e. *Can you think of any reasons the equivalent for the string might vary?*
 - f. *Why do you think the equivalent of the string is not highlighted for you as well?*
 - g. *Does WeBiText find only the exact form that you enter in the search field, or does it allow for variations? What can this tell you about the searching approach WeBiText uses?*
 - h. *How could you test out the search function in WeBiText to learn more about how it works? Try a few searches to see what WeBiText will find.*
 - i. *What does this mean you should keep in mind when you do searches using WeBiText?*
2. Below each sentence, a hyperlink indicates the source of the sentence. You can click this hyperlink to see the original document online.
 - a. Try clicking the links for the first results in your search. New windows open to display the texts.
 - b. When you have looked at the texts, close the windows/tabs by clicking on the Close icons (ⓧ) at the top right-hand corners.
3. Between each pair of sentences, an icon of a page (📄) allows you to view the whole aligned documents (the *bitext* created by WeBiText). Click on this icon (📄) to view the bitext from which the first occurrence was taken.
 - a. *Do you recognize where the first aligned segments come from?*
4. Close this new window/tab and return to the list of results.
5. You can also view the documents side-by-side in their original format by clicking the Web page icon (🌐) below the page icon. Click on this icon to see the



difference between the two display modes. **When do you think each one might be useful?**

NOTE: This button is also available in the heading of the first bitext view.

6. Close the window/tab to return to the list of results, then choose another occurrence to view in bitext form.
7. Look at the alignment for the first few dozen segments.
 - a. Are they all perfectly aligned? If not, can you see why not?
 - b. Are you always able to find the correct French sentence, near to if not exactly beside the English sentence?
8. Look through the bitext for occurrences of the search string.
 - a. At the top of the document, a **FIRST** link appears. Click on it to move to the first time *vendor* appears in the document.
 - b. Note the brown highlighting of the sentence, and the green highlighting of the search string in the segment. Read the sentence and its equivalent, and locate the equivalent used to translate *vendor* (if any).
 - c. Click the **NEXT** link that appears in square brackets at the end of the English sentence to move on to the next segment that contains the string. (Once you have moved on, you can also use the **PREVIOUS** link to move to a previous occurrence.)
 - d. Is the equivalent used for *vendor* consistent within this single document? How does this compare with the overall results you looked at earlier?
9. When you have looked at a few occurrences of the search string, close the window/tab to go back to the list of search results.

Comparing WeBiText search results from Canadian government documents with results from European Union documents

1. Do two searches for the string *weekend*, one in the Canadian government sites, the other on the European Union sites.
 - a. Enter *weekend* in the search field, ensure that **ENGLISH** is selected in the first dropdown list, **FRENCH** in the second and **GC.CA GOVERNMENT OF CANADA (SLOW)** in the third, and then click the **SEARCH** button.
 - b. What equivalent(s) do you find for this term in the Canadian government pages?
 - c. Enter *weekend* in the search field, ensure that **ENGLISH** is selected in the first dropdown list and **FRENCH** in the second, but this time select **EUROPA.EU** in the second dropdown list to search European Union documents.
 - d. Click the **SEARCH** button.
 - e. What equivalent(s) do you find for this term in the European Union pages? Are the results different from what you found in the Canadian documents?
2. Do two searches for the string *design*, one in the Canadian government sites and the other on the European Union sites.
 - a. What equivalent(s) do you find for this term in the European Union pages? Are the results different from what you found in the Canadian documents?

3. Can you think of other examples of items for which results may be different in the two corpora? Try some searches for these to see if your intuitions are correct.
4. Judging from these results, how do you think WeBiText's option to search in two corpora may be useful?
5. Can you think of any reasons why you should be careful when using results found using this kind of tool?

Answering some translation questions using WeBiText

1. Search for occurrences of the string *eventually* and its equivalent(s).
 - a. Enter *eventually* in the search field, ensure that **ENGLISH** is selected in the first dropdown list, **FRENCH** in the second and **GC.CA GOVERNMENT OF CANADA (SLOW)** in the third, and click the **SEARCH** button.
 - b. Examine the results. What equivalent(s) do you find for this expression? Is it always easy to identify the equivalent? Why or why not?
2. Search for occurrences of the complex expression *privy council*, and its equivalent(s).
 - a. Enter *privy council* in the search field, ensure that **ENGLISH** is selected in the first dropdown list, **FRENCH** in the second and **GC.CA GOVERNMENT OF CANADA (SLOW)** in the third, and click the **SEARCH** button.
 - b. Examine the results. What form(s) do you find for this expression?
 - c. Judging from the results of this search, do you think WeBiText uses case-sensitive searching? Why or why not?
 - d. Does this expression form part of any more complex expressions in English? If so, what are they?
 - e. What equivalent(s) do you find for this expression and those that contain it? Is the equivalent of *privy council* consistent in all of the expressions?
3. Search for complex expressions that include the string *counsel*, and their equivalents.
 - a. Enter *counsel* in the search field, ensure that **ENGLISH** is selected in the first dropdown list, **FRENCH** in the second and **GC.CA GOVERNMENT OF CANADA (SLOW)** in the third, and click the **SEARCH** button.
 - b. Examine the results. Do you have any difficulties identifying the limits or forms of the expressions? If so, why?
 - c. Do you have any difficulties identifying the equivalent(s) of each of the expressions? Why?
 - d. Do you notice anything in particular about the form(s) in which this string appears in the results?
 - e. How do you think this kind of search compares to searches you can do in dictionaries or term banks?
4. Identify some collocates of the string *file* and its equivalent(s).
 - a. Type the string *file* into the search field, ensure that **ENGLISH** is selected in the first dropdown list, **FRENCH** in the second and **GC.CA GOVERNMENT OF CANADA (SLOW)** in the third, and click the **SEARCH** button.

- b. Examine the results. What kinds of occurrences do you find?
- c. Do they all correspond to the same lexical unit (i.e., word)? Do they all have the same meaning?
- d. What are some collocates of (i.e., words that appear frequently in connection with) each of the units and/or senses you identified? (You may want to think about adjectives, verbs and/or nouns that often appear in connection with this string.) Are they the same for all units and senses you identified?
- e. What are the equivalents used for the string *file*? What are the equivalents of the collocates you identified?
- f. What can analyzing the collocates of an item help you to do?

5. Analyze the occurrences of the string *grow*.

- a. Type the string *grow* into the search field, ensure that **ENGLISH** is selected in the first dropdown list, **FRENCH** in the second and **GC.CA GOVERNMENT OF CANADA (SLOW)** in the third, and click the **SEARCH** button.
- b. Examine the results. What kinds of occurrences and structures do you find? Do you think all of these would be found easily in a dictionary entry? Why or why not?

NOTE: If you like, you can consult a monolingual English dictionary such as the *Oxford English Dictionary* (www.oed.com) or the *LDOCE Online* (www.ldoceonline.com). You can learn more about using these dictionaries in the *Monolingual Electronic Dictionaries Exercise, Level I* on the CERTT site (ACCESS BY TOOL NAME > ELECTRONIC DICTIONARIES).

- c. How are these occurrences translated in French?

6. Analyze the occurrences of the string *prevent*.

- a. Type the string *prevent* into the search field, ensure that **ENGLISH** is selected in the first dropdown list, **FRENCH** in the second and **GC.CA GOVERNMENT OF CANADA (SLOW)** in the third, and click the **SEARCH** button.
- b. Examine the results. What forms and structures do you find?
- c. How are these forms and structures translated in French? Are they always parallel to the English structures?

7. Evaluate occurrences and equivalents for the string *incidental*.

- a. Type the string *incidental* into the search field, ensure that **ENGLISH** is selected in the first dropdown list, **FRENCH** in the second and **GC.CA GOVERNMENT OF CANADA (SLOW)** in the third, and click the **SEARCH** button.
- b. Examine the results. What kinds of occurrences do you find?
- c. Do they all correspond to the same lexical unit (i.e., word)? Do they all have the same meaning? What kinds of sentence structures does the string appear in?
- d. Does this string appear in any complex expressions? If so, what are they?
- e. What about the equivalent(s) for this string and/or expressions that contain it? How are they linked to the various units and/or senses you have observed?
- f. Compare the results of your research with the entries for *incidental* in an English monolingual and a French-English bilingual dictionary.

NOTE: You can compare the occurrences you find to the entry in a monolingual English dictionary such as the *Oxford English Dictionary* (www.oed.com) or the *LDOCE Online* (www.ldoceonline.com). You can also compare equivalents with those found in a bilingual

dictionary such as the *Oxford-Hachette*, installed in the Writing Centre (START > PROGRAMS > TRADUCTION > HACHETTE OXFORD > DICTIONNAIRE HACHETTE-OXFORD).

- g. How do the dictionary entries compare with your synthesis? Did you find anything that is not described in the dictionaries? Do you think any additional information found in the results of WeBiText can be useful to you as a translator? How?

Questions for reflection

1. As you did these exercises, what did you notice about how WeBiText works?
2. What kinds of tasks do you think WeBiText performs, to be able to produce the results you have seen? Why?
3. What could WeBiText help you to do? In what kind of situation?
4. Given that WeBiText searches the Canadian federal government and European Union sites, what subject fields do you think it might be particularly useful for working in? What text types do you think it is likely to help you with (e.g., technical documents, scientific articles, popularized articles, training documents, administrative correspondence, advertising, etc.)? Are there any subject fields or text types you don't think it would help with? Why or why not?
5. What criteria can be used to evaluate bilingual concordancers such as WeBiText?
6. How does WeBiText compare to others in the same class, such as TransSearch, ParaConc, BeeText Find and/or the LogiTerm bilingual concordancer?
7. What are some of the advantages and disadvantages of using WeBiText or other concordancers to search for occurrences of search strings and their equivalents, compared to searching in monolingual and bilingual dictionaries?
8. What are some of the advantages and disadvantages of using WeBiText to search for occurrences of search strings and their equivalents, compared to a manual approach? Compared to using another tool?
9. What did you like most about WeBiText?
10. Was there anything you didn't like about it?
11. How do you think WeBiText could be modified, expanded, or improved to be even more useful for you?

NOTE: If you wish, you may send WeBiText's developers your answers to questions 9, 10 and 11 by clicking the **FEEDBACK** link on the WeBiText search page, or by sending an e-mail to Alain.Desilets@nrc-cnrc.gc.ca with *Feedback from the WeBiText site* in the subject line. (If you wish to send your comments anonymously, feel free to use the CERTT anonymous g-mail account. To do this, just log on at www.gmail.com using the username *uo.certt.anon* and the password *uo.certt*.) Your feedback will be much appreciated and will help future versions of WeBiText meet users' needs even better.