



Le Migou Online Concordancer Tutorial and Exercise: Level I

Introduction

Le Migou is an online monolingual concordancer that was developed at l'Observatoire de linguistique Sens-Texte ([OLST](http://www.olst.ling.umontreal.ca)) at l'Université de Montréal by Patrick Drouin (<http://www.mapageweb.umontreal.ca/drouinp/>). It allows a user to consult general and specialized language corpora which, for the most part, have been compiled by members of OLST. The corpora are mainly in French, but there is one in English (the SACOT corpus on terrorism), and one in Korean (on computer science).

A concordancer such as Le Migou can make it easier to search through texts by allowing a user to search for one or several character strings and display all occurrences of these character strings in context.

Le Migou is conveniently available online at all times. However, since the corpora are already built, the user has less flexibility in terms of content; you are limited in the domains or corpora you wish to consult.

NOTE : Le Migou interface is French-only. As previously mentioned, there is an English corpus, in which you can search in English, but the interface will always remain in French.

Doing these exercises will help you:

- [Connect to Le Migou](#);
- [Get to know the interface](#);
- [Complete basic and advanced searches](#);
- [Display and analyze search results](#).

NOTE : To use Le Migou, you need an account. (The account and use of the tool are free). The University of Ottawa has a group account. The login information for this account appears in the exercises below. If wish to request your own personal account, on the Le Migou access page (<http://olst.ling.umontreal.ca/ressources/corpus-du-migou/>) click the REGISTRATION button and provide the information necessary to create your account. It will be activated within the next few days and the account information will be e-mailed to you.

Getting ready

1. Open the web browser of your choice, for example, Internet Explorer (from the INTERNET EXPLORER shortcut on the Desktop), or Firefox (from the START menu).
2. Open Le Migou's home page by clicking on the following link:
<http://olst.ling.umontreal.ca/ressources/corpus-du-migou/>

3. Click the Access button and log in:
 - a. As your username, enter *UdO*.
 - b. As your password, enter *UdO* again.
 - c. Click the **OK** button to log in.

Getting to know the interface

1. The search interface has several fields:
 - a. In the first search field next to **PREMIER MOT**, you can enter a word or group of words (character strings) which to search for in a corpus;

NOTE: By default, Le Migou searches single words, that is, character strings which are graphically delimited by spaces, punctuation marks, symbols and carriage returns.

- b. In the drop-down list next to the search field, you can specify how you want to search for your character strings:
 - i. Exactly the way it appears in the search field (without preceding or following characters) (**SÉQUENCE EXACTE**);
 - ii. Potentially followed by more characters (**DÉBUTE PAR CETTE SÉQUENCE**);
 - iii. Potentially preceded by more characters (**SE TERMINE PAR CETTE SÉQUENCE**);
 - iv. Potentially preceded and followed by more characters (**SÉQUENCE IMBRIQUÉE**).
 - c. Below the first field, there is a drop-down list of available Boolean operators which help a user to combine search strings. The first in the list is the ET (AND) operator which searches for two words together. The second in the list is the OU (OR) operator which searches for one word or the other. If the ET operator is used, you can specify that you want the words searched in the order in which you entered them by checking the **RESPECTER L'ORDRE DES MOTS** checkbox next to the drop-down list.
 - d. The **DEUXIÈME MOT** search field allows you to enter a second search string of characters, if you wish. The drop-down list next to it allows you to specify how you want to search for your character strings (as in the **PREMIER MOT** drop-down list).
 - e. The drop-down list next to **CORPUS** allows you to specify which corpus you wish to search. Le Migou gives you access to a series of French corpora; most of these are specialized, but there are also two newspaper corpora (articles from *Le Monde* and the Canadian press). You can search through one corpus at a time by selecting the one you wish to search, or you can search all of them at once by selecting **TOUS LES CORPUS**.
 - f. The drop-down list next to **DISTANCE MAXIMALE ENTRE LES MOTS** allows you to specify how many words can appear between the specified search strings if you want to combine two strings using the **ET** operator.
 - g. The drop-down list next to **NOMBRE DE CONTEXTES PAR PAGE** allows you to specify how many contexts will be displayed per page in the search results.
 - h. Finally, the **LANCER LA RECHERCHE** button allows you to start your search.

Complete a search and analyze the results

1. Complete a search for the term *terror* :
 - a. Enter *terror* in the **PREMIER MOT** search field;
 - b. From the drop-down list next to it, select **SÉQUENCE EXACTE**;

- c. From the **CORPUS** drop-down list, select *Corpus SACOT (EN)*;
- d. Leave the **DEUXIÈME MOT** search field blank for now;
- e. Leave the rest of the default values;
- f. Click the **LANCER LA RECHERCHE** button.

A new page opens displaying the search results.

2. Analyze the search results, which are displayed in table form.
 - a. The *Identificateur* column identifies the sentence in which the search string occurs. You can click on the sentence number to display the sentence in a longer context.
 - b. The *Corpus* column identifies the corpus in which the search string occurs.
 - c. The *Document* column identifies the name of the file in which the search string occurs.
 - d. The *Contextes* column displays all occurrences of the search string in the context of about one sentence. The search string is indicated in red.

NOTE : To make searching easier, the format of the texts has been modified slightly, for example, to separate punctuation marks from the words they follow. Sometimes some effort will be required to interpret the contexts as they appear.

- e. Above the table, the number of contexts that are displayed on the page is indicated. (If this is the maximum number of results per page, there may be some results on subsequent pages.)
 - f. Below the table, you can click the **SUITE DES OCCURRENCES** button to see the following occurrences.
 - g. Finally, you can click the **NOUVELLE RECHERCHE** button to return to the search page.
3. Try the other search possibilities (by searching for words that begin or end with the string *terror* or words which contain this string). Look at the different strings that are displayed.
 - a. Of these possibilities, what is the most efficient search? Why?
 - b. What are the advantages of using this type of search? What are the disadvantages?
4. Try some searches that combine this string with another (for example, *war*, *cell* or *violence*) using ET or OU.
 - a. What searches did you launch? What functions did you use? Why?
 - b. What type of information can you gather from doing these types of searches? How is this information useful for translators or terminologists?
5. Complete a search for all forms of the verb *to wage*:
 - a. How can you search for all occurrences of the different forms of the verb?
 - b. What effects will the different ways of searching have on the results? What are the difficulties you might encounter when searching for various forms of verbs using this type of tool?
 - c. In what type of structures does this verb appear? With what types of words is this verb used (e.g., what are its typical subjects and objects, which in French are sometimes called its *combinatoire*)?
6. Complete a similar search in the French corpus Corpus medical OLST (FR), for the verb *traiter*, and observe its structures and *combinatoire*.

- a. Do you think all occurrences of this verb have the same meaning? Why?
- b. Do you think the search results represent all possible meanings of the verb? Complete some searches in different corpora (for example, the computer science corpora) to test your hypothesis.
- c. Compare the structures and common subjects and objects of the occurrences in the other corpora with those in the medical corpus. Do you notice any differences? Any similarities?
- d. Do you see any differences between these results and those from your analysis of the verb *to wage*? Might these be significant? Why?
- e. How is this information useful for translators or terminologists?

Questions for reflection

1. Are the search functions in Le Migou similar to those in other concordancers that you have worked with? What about the interface? Is it similar or different? Is the presentation of results in Le Migou similar to the presentation of results in other concordancers you have worked with?
2. What are some of the advantages and disadvantages of using an online concordancer compared to a concordancer that searches through corpora stored on your own computer?
3. What are the main difficulties that you face when searching through corpora using a concordancer such as Le Migou?