



Tutoriel et exercices avec le concordancier en ligne Le Migou : niveau I

Introduction

Le Migou est un concordancier unilingue en ligne conçu à l'Observatoire de linguistique Sens-Texte ([OLST](http://www.mapageweb.umontreal.ca/drouinp/)) de l'Université de Montréal par Patrick Drouin (<http://www.mapageweb.umontreal.ca/drouinp/>). Ce concordancier permet de consulter des corpus de langue spécialisée et générale qui ont, pour la plupart, été construits par des membres de l'OLST dans le cadre de travaux en terminologie. Les corpus sont principalement de langue française, mais il y en a aussi en anglais (le corpus SACOT, dans le domaine du terrorisme) et en coréen (informatique KR).

Un concordancier comme le Migou puise à même une banque de textes pour y chercher les occurrences d'une ou de plusieurs chaînes de caractères et les affiche ensuite dans leur contexte d'utilisation.

Un des avantages du Migou est sa disponibilité : offert gratuitement, vous pouvez le consulter en tout temps et en tous lieux, à condition d'avoir une connexion Internet. Cependant, puisque les corpus sont déjà construits, l'utilisateur a peu de souplesse sur le plan du contenu; il ne peut que choisir le domaine ou le corpus qu'il souhaite consulter.

Ces exercices vous aideront à :

- [vous familiariser avec l'interface du Migou](#);
- [faire des recherches simples et complexes et analyser les résultats](#).

Préparation

1. Ouvrez le navigateur de votre choix, par exemple Mozilla Firefox (disponible à partir du raccourci sur le Bureau de votre ordinateur).
2. Ouvrez la page d'accueil du Migou en tapant l'URL dans la barre d'adresse du navigateur ou en cliquant sur le lien suivant : http://olst.ling.umontreal.ca/?page_id=54.
3. Cliquez sur le bouton **ACCÉDER** pour ouvrir la page de recherche du Migou.

Vous familiariser avec l'interface du Migou

1. Explorez l'interface d'interrogation du Migou, qui comporte plusieurs champs :
 - a. Le champ de recherche à côté de *Premier mot* sert à entrer le mot ou le groupe de mots (chaîne de caractères) qui vous permettra de lancer la recherche dans un corpus;
 - b. Une liste déroulante à côté de ce champ permet de spécifier si vous voulez que la chaîne de caractères soit recherchée :
 - i. Sans caractères adjacents (**SEQUENCE EXACTE**);
 - ii. Potentiellement suivie par d'autres caractères (**DÉBUTE PAR CETTE SÉQUENCE**);
 - iii. Potentiellement précédée par d'autres caractères (**SE TERMINE PAR CETTE SÉQUENCE**);

iv. Potentiellement précédée et suivie par d'autres caractères (**SÉQUENCE IMBRIQUÉE**).

NOTE : Le Migou cherche par défaut des mots simples, c'est-à-dire des chaînes de caractères délimitées graphiquement par des espaces, des signes de ponctuation, des symboles et des retours de chariot.

- c. En dessous de ce premier champ se trouve une liste déroulante des opérateurs booléens disponibles pour combiner des mots-clés. Il s'agit des opérateurs ET pour chercher deux mots à la fois et OU pour chercher l'un ou l'autre de ces mots. Si le ET est utilisé, il est possible de spécifier l'ordre dans lequel les mots apparaissent en cochant la case **RESPECTER L'ORDRE DES MOTS** à côté de la liste déroulante.
- d. Le champ de recherche à côté de *Deuxième mot* permet de spécifier une deuxième chaîne à rechercher, si désiré, et le menu déroulant à côté de celui-ci permet de spécifier comment la chaîne sera recherchée (comme dans le cas de la première chaîne).
- e. La liste déroulante à côté de *Corpus* permet de spécifier le corpus à interroger. Le Migou donne accès à une série de corpus en français; la plupart sont spécialisés, mais il y a aussi deux corpus journalistiques (tirés du quotidien *Le Monde* et de la presse canadienne). Vous pouvez interroger un corpus à la fois en choisissant un corpus dans la liste, ou les interroger tous en même temps en choisissant l'option **TOUS LES CORPUS**.
- f. La liste déroulante à côté de *Distance maximale entre les mots* permet de préciser combien de mots peuvent apparaître entre les chaînes spécifiées si on souhaite combiner deux chaînes.
- g. La liste déroulante à côté de *Nombre de contextes par page* permet de spécifier combien de contextes seront affichés par page dans les résultats de recherche. (Notez que des pages avec moins de contextes d'afficheront plus rapidement que celles avec beaucoup de contextes.)
- h. Finalement, le bouton **LANCER LA RECHERCHE** permet de démarrer la recherche.

Faire des recherches simples et complexes et analyser les résultats

1. Faites une recherche à partir du terme *vasculaire* :
 - a. Entrez *vasculaire* dans le champ de recherche *Premier mot*;
 - b. À partir de la liste déroulante à côté de ce champ, choisissez l'option **SÉQUENCE EXACTE**;
 - c. À partir de la liste déroulante *Corpus*, choisissez l'option **CORPUS MÉDICAL OLST (FR)**;
 - d. Laissez le champ de recherche *Deuxième mot* vide pour l'instant;
 - e. Laissez les autres valeurs par défaut;
 - f. Cliquez sur le bouton **LANCER LA RECHERCHE**.

Une nouvelle page apparaît et présente les résultats de la recherche.

2. Explorez les résultats de recherche, qui sont présentés sous forme de tableau.
 - a. La colonne *Identificateur* indique le numéro de la phrase dont l'occurrence est tirée. Si désiré, vous pouvez cliquer sur ce numéro de phrase pour faire afficher un contexte plus long.
 - b. La colonne *Corpus* indique le corpus dont l'occurrence est tirée.
 - c. La colonne *Document* indique le nom du fichier dont l'occurrence est tirée.
 - d. La colonne *Contextes* présente les occurrences de la chaîne, dans un contexte d'à peu près une phrase. La chaîne recherchée est en rouge.

NOTE : Afin de faciliter l'interrogation, le format des textes a été modifié un peu, par exemple, pour séparer les signes de ponctuation des mots qu'ils suivent. Une certaine interprétation des contextes est parfois nécessaire.

- e. En haut du tableau, une ligne indique le nombre de contextes affichés sur la page. (Si c'est le nombre maximal de résultats par page, il peut y avoir d'autres résultats sur des pages subséquentes.)
 - f. En dessous du tableau, le bouton **SUITE DES OCCURRENCES** permet de passer aux occurrences des pages suivantes.
 - g. Finalement, le bouton **NOUVELLE RECHERCHE** permet de retourner à l'interface d'interrogation.
3. À partir du même corpus, essayez les autres possibilités de recherche (en cherchant des mots qui débutent ou se terminent par *vasculaire* ou qui contiennent cette chaîne). Observez les différentes possibilités d'occurrences de la chaîne.
 - a. Quelle est la recherche la plus polyvalente parmi ces possibilités? Pourquoi?
 - b. Dans quelles circonstances la polyvalence d'une recherche comme celle-ci serait-elle importante?
 - c. Quels sont les avantages d'utiliser ce type de recherche? Quels sont les inconvénients?
 4. Essayez des recherches qui combinent cette chaîne avec d'autres unités (par exemple, *accident* ou *artériel*) à l'aide des opérateurs ET et OU.
 - a. Quelles recherches avez-vous faites? Avec quelles fonctions? Pourquoi?
 - b. Quel genre d'information pouvez-vous repérer en faisant ces types de recherches? Comment cette information peut-elle être utile au traducteur ou au terminologue?
 5. À partir de la liste déroulante du champ *Corpus*, choisissez le corpus anglais **SACOT (EN)** et faites une recherche pour toutes les formes possibles du verbe *to wage*.
 - a. Comment peut-on rechercher les occurrences des différentes formes du verbe?
 - b. Quels effets les différentes façons de faire la recherche vont-elles avoir sur les résultats? Quelles sont les difficultés qui se présentent lorsqu'on veut rechercher des formes variables telles que des verbes avec ce genre d'outil?
 - c. Dans quelles sortes de structures ce verbe intervient-il? Avec quelles sortes d'unités se combine-t-il (en d'autres mots, quelle est sa *combinatoire*)?
 6. À partir de la liste déroulante du champ *Corpus*, choisissez de nouveau le **CORPUS MÉDICAL OLST (FR)** et faites une recherche des formes du verbe *traiter* :
 - a. Croyez-vous qu'il s'agit toujours du même sens pour ce verbe? Pourquoi?
 - b. Croyez-vous que ce sont tous les sens possibles de cette unité? Faites des recherches dans différents corpus (par exemple, dans le **CORPUS INFORMATIQUE FR**) pour tester vos hypothèses.
 - c. Comparez les structures et la combinatoire des occurrences dans d'autres corpus avec celles provenant du corpus médical. Remarquez-vous des différences? Des similarités?
 - d. Trouvez-vous des différences entre les résultats de cette analyse et de celle du verbe *to wage*? Ces différences sont-elles significatives? Pourquoi?
 - e. Comment cette information peut-elle être utile au traducteur ou au terminologue?

Questions de réflexion

1. Les fonctions de recherche du Migou sont-elles similaires à celles offertes par d'autres concordanciers que vous connaissez? Qu'en est-il de son interface? Est-elle similaire ou différente? La présentation des résultats de recherche dans Le Migou ressemble-t-elle à celle des autres concordanciers que vous connaissez?



2. Quels sont les principaux avantages ou inconvénients d'utiliser un concordancier en ligne par rapport à un concordancier qui exploite des corpus stockés sur votre poste de travail?
3. Comment un concordancier comme Le Migou peut-il être utile pour un traducteur ou terminologue? Pour faire quoi?
4. Quelles sont les principales difficultés auxquelles des concordanciers comme Le Migou peuvent être confrontés lors de recherches dans des corpus?