



Tutoriel et exercices avec Concord de WordSmith Tools, niveau II

WordSmith Tools, conçu par Mike Scott, est un outil d'analyse de corpus qui intègre notamment Concord, un concordancier unilingue qui affiche des occurrences d'une chaîne spécifiée en contexte, et ce, en format KWIC (*key word in context*). (Il propose aussi une série d'autres outils d'analyse de textes qui pourront être explorés en faisant les exercices sur l'outil [WordList](#).)

Pour en savoir plus sur WordSmith, visitez le site www.lexically.net et lisez les fichiers d'aide portant sur Concord (stockés au Centre d'écriture dans le répertoire V:\WordSmith4 Help files et aussi disponibles en ligne). Vous pouvez aussi télécharger une version démo de WordSmith Tools (qui offre toutes les options de la version commerciale, mais qui limite le nombre d'occurrences affichées à 25) à partir du même site (www.lexically.net).

Concord de WordSmith Tools peut traiter des textes en format .html, .xml et .txt. Il offre de plus une variété de fonctions de recherche, de modes d'affichage des résultats et de calculs statistiques des cooccurrences.

Les exercices ci-dessous font suite au *Tutoriel et exercices avec Concord de WordSmith Tools, niveau I* sur le site de la CERTT (**ACCÈS PAR NOM D'OUTIL > WORDSMITH TOOLS > WORDSMITH TOOLS CONCORD, NIVEAU I**). Ils vous montreront comment :

- [choisir les textes à exploiter](#);
- [utiliser des fonctions de recherche avancées](#);
- [faire des recherches complexes](#); et
- [identifier des cooccurrents du mot recherché](#).

Préparation

1. Créez un sous-répertoire du lecteur U: (aussi appelé *Mes documents*). (Pour des instructions à ce sujet, consultez le document *Créer un sous-répertoire* sur le site de la CERTT (**ACCÈS PAR NOM D'OUTIL > WINDOWSXP > CRÉER UN SOUS-RÉPERTOIRE**)).
2. Téléchargez dans ce sous-répertoire le fichier contenant des décisions de la Cour suprême du Canada en français et en format txt à partir du site CERTT (**CORPUS ET AUTRES RESSOURCES > DÉCISIONS DE LA COUR SUPRÊME DU CANADA > SUPREME COURT DECISIONS (FRANÇAIS, TXT)**).
3. Extrayez les fichiers du dossier compressé dans le sous-répertoire que vous venez de créer. (Pour des instructions à ce sujet, consultez le fichier *Extraire des fichiers d'un dossier compressé* sur le site de la CERTT (**ACCÈS PAR NOM D'OUTIL > WINDOWSXP > EXTRAIRE DES FICHIERS D'UN DOSSIER COMPRESSE**)).

4. Ouvrez WordSmith Tools (**DÉMARRER > PROGRAMMES > TRADUCTION > WORDSMITH 4 > WORDSMITH.EXE**).
5. Lorsque WordSmith vous demande si vous voulez voir uniquement les fonctions de base du logiciel, cliquez sur **NO**. Vous aurez ainsi accès à l'ensemble des fonctions.
6. Le logiciel s'ouvre avec une fenêtre centrale à partir de laquelle les différentes options offertes sont accessibles (les principales à partir des trois boutons **CONCORD**, **KEYWORDS** et **WORDLIST**, les autres par le menu **UTILITIES**).

Choix des textes à exploiter

1. Sélectionnez un ou plusieurs fichiers à traiter :
 - a. Dans la fenêtre principale de WordSmith, ouvrez le menu **FILE** et choisissez l'option **CHOOSE TEXTS**.
 - b. Dans la fenêtre *Choose texts* qui apparaît, les fichiers à exploiter sont affichés dans la partie droite, et l'arborescence des répertoires, dans la partie gauche.
 - c. Un fichier démo fourni avec le logiciel (un chapitre tiré de *A Tale of Two Cities* de Dickens) apparaît dans la liste à droite. Enlevez-le en le sélectionnant et en appuyant sur la touche **SUPPR** pour ne pas avoir dans vos résultats des occurrences provenant de ce fichier.
 - d. Dans le menu déroulant situé dans le coin supérieur gauche de la fenêtre *Choose texts*, choisissez le lecteur où vous avez stocké les documents à analyser (le U: ou *Mes documents*).
 - e. Vous pouvez aussi choisir de voir uniquement des fichiers dans un certain format (par exemple, texte brut ou .txt), en sélectionnant l'option correspondante dans le menu déroulant qui se trouve immédiatement à droite de la liste des lecteurs.
 - f. Dans la partie gauche de la fenêtre, parcourez l'arborescence des répertoires dans le lecteur U: pour trouver les fichiers que vous voulez exploiter. (Double-cliquez sur les icônes des dossiers jaunes pour ouvrir les répertoires.)
 - g. Sélectionnez tous les fichiers. (Vous pouvez utiliser la touche **MAJ** pour sélectionner les fichiers en bloc.)
 - h. Cliquez ensuite sur le long bouton vertical avec deux flèches bleues pointant vers la droite. Les fichiers apparaissent dans la partie droite de la fenêtre.
2. Une fois la sélection des fichiers terminée, sortez de la fenêtre *Choose texts* en cliquant sur l'icône de la coche verte (✓).

Recherches en respectant la casse

1. Ouvrez **CONCORD** en cliquant sur le bouton qui porte son nom.
2. À partir du menu **FILE**, choisissez l'option **NEW**.
3. Dans la boîte de dialogue *Getting started* qui s'affiche, cliquez sur l'onglet **SEARCH WORD** pour vous assurer qu'il est actif.

4. Pour rechercher une chaîne spécifiquement avec ou sans majuscules, entrez la chaîne en question dans le champ **SEARCH WORD**, en le faisant précéder et suivre de deux signes d'égalité (==).
 - a. Essayez de chercher *cour* et *Cour*. **Qu'est-ce qui arrive?**
 - b. Cherchez *cour* et ==*cour*==. Évaluez les résultats. **Quelle est la différence? Qu'est-ce que cela peut vous indiquer à propos de l'usage de l'unité dans ces textes?**
 - c. Cherchez *cours* et ==*cours*==. Évaluez les résultats. **Quelle est la différence? Qu'est-ce que cela peut vous indiquer à propos de l'usage de l'unité dans ces textes?**
 - d. Les occurrences de *cours* sont-elles toutes pertinentes pour la recherche de formes du mot *cour*? **Pourquoi?**
 - e. **Pouvez-vous penser à d'autres difficultés que poserait la recherche de cette unité? Dans quelles conditions?**

Recherches avec troncature

1. Pour chercher une série de chaînes qui partagent seulement une partie de leurs caractères, entrez la partie commune de la chaîne dans le champ *Search word*, et, à la place du caractère qui diffère/des caractères qui diffèrent, entrez le symbole de troncature qui convient :
 - a. Pour remplacer un caractère, entrez un point d'interrogation (?).
 - i. Lancez une recherche pour trouver *cour* et *cours* avec une seule requête.
 - b. Pour remplacer plusieurs caractères, entrez un astérisque (*).
 - i. Lancez une recherche pour trouver des mots qui commencent avec *législ*.
 - ii. **Quels mots trouvez-vous?**
 - iii. Lancez une recherche pour trouver des mots qui contiennent la chaîne *appel*.
 - iv. **Quels mots trouvez-vous?**
 - v. **En quoi ce genre de fonction peut-il être utile pour interroger un corpus?**

Recherches complexes

1. Pour combiner des critères avec l'opérateur booléen OU (c'est-à-dire, pour trouver des contextes qui contiennent au moins une des chaînes entrées), entrez les deux chaînes dans le champ *Search word*, séparées par une barre oblique (/). Cliquez sur le bouton **OK** pour lancer la recherche.
 - a. Lancez une requête pour trouver à la fois les chaînes *appel* et *pourvoi*.
 - b. Triez les résultats pour séparer les occurrences des deux chaînes, en choisissant comme premier critère de tri **CENTRE**.

NOTE : Pour une description des fonctions de tri et de leur usage, consultez le *Tutoriel et exercice avec l'outil Concord de WordSmith Tools, niveau I* sur le site de la CERTT (**ACCÈS PAR NOM D'OUTIL > WORDSMITH TOOLS > WORDSMITH TOOLS CONCORD, NIVEAU I**).

- c. Retriez les résultats pour mettre en évidence les régularités dans les mots qui précèdent ou suivent ces deux chaînes.

- d. Les cooccurrents et les structures dans lesquelles ces unités se présentent sont-ils les mêmes pour les deux chaînes ? Est-ce que ce tri met en évidence de l'information utile pour le traducteur ou le terminologue ? Pourquoi ?
2. Pour combiner des critères avec un opérateur de proximité (c'est-à-dire, pour trouver des contextes qui contiennent deux chaînes situées à une certaine distance l'une de l'autre) :
 - a. Entrez la première chaîne dans le champ *Search word*.
 - b. Entrez la deuxième chaîne dans le champ *Context word* qui apparaît sur l'onglet **ADVANCED**.

NOTE : Si vous ne voyez pas l'onglet **ADVANCED**, il se peut que, lors de l'ouverture du logiciel, vous ayez sélectionné **YES** lorsque WordSmith vous a demandé si vous vouliez voir uniquement les fonctions de base dans le logiciel. Annulez en cliquant sur le bouton **CANCEL**. Du menu **SETTINGS**, choisissez l'option **SPECIFIC TO CONCORD** et cliquez ensuite sur l'onglet **GENERAL**. Décochez la case **KEEP THINGS SIMPLE** et cliquez sur le bouton **OK**. Relancez la recherche; l'onglet devrait apparaître.

- c. Choisissez la distance admissible entre les deux chaînes en spécifiant le nombre de mots qui peuvent apparaître entre les deux chaînes à gauche et à droite de la chaîne principale.
- d. Cliquez sur le bouton **OK** pour lancer la recherche.

NOTE : Cette fonction de recherche avec proximité prend la place de l'opérateur booléen **ET** disponible dans certains autres outils. Elle est cependant plus précise étant donné qu'elle spécifie non seulement la présence de deux chaînes dans un contexte, mais aussi leurs positions relatives.

- e. Essayez une recherche pour les occurrences de *cour* ou *cours* avec la chaîne *instance* à l'intérieur de 3 mots à sa droite.
 - f. Quels sont les résultats? Comment ce type de recherche pourrait-il être utile au traducteur ou au terminologue?
3. Pour combiner des critères avec l'opérateur booléen **SAUF** (c'est-à-dire, pour trouver des contextes qui contiennent une chaîne mais pas une autre) :
 - a. Entrez la chaîne qui doit être présente dans le champ *Search word*.
 - b. Entrez la chaîne qui ne doit pas être présente dans le champ **EXCLUDE IF CONTEXT CONTAINS** sur l'onglet *Advanced*.
 - c. Cliquez sur le bouton **OK** pour lancer la recherche.
 - d. Lancez une recherche pour *Cour* (en respectant la casse).
 - e. Répétez la recherche, mais cette fois en excluant les contextes où *appel* apparaît à l'intérieur d'un espace de deux mots à droite de *Cour*.
 - f. Quelle est la différence entre les résultats?
 - g. Triez les occurrences pour mettre en évidence les différents types de *Cours* dont il est question dans les textes.
 - h. Comment ce type de recherche peut-il être utile au traducteur ou au terminologue ?

Aide-mémoire

Le tableau ci-dessous résume les fonctions offertes dans WordSmith Tools.

Fonction	WordSmith Tools
Requête verbatim (exacte)	<i>chaîne1 chaîne2</i>
Respecter la casse	<i>==chaîne1==</i>
Troncature (un caractère)	<i>chaîne?</i>
Troncature (plusieurs caractères)	<i>chaîne*</i>
Opérateur booléen OU	<i>chaîne1/chaîne2</i>
Opérateur booléen SAUF	<ul style="list-style-type: none"> Entrez la chaîne principale à rechercher dans le champ <i>Search word</i>. Entrez la forme à exclure dans le champ <i>Exclude if context contains</i>.
Opérateur de proximité	<ul style="list-style-type: none"> Entrez <i>chaîne1</i> dans le champ <i>Search word</i>. Entrez <i>chaîne2</i> dans le champ <i>context word</i>. Spécifiez le nombre de mots à gauche et/ou à droite de la première chaîne.
Opérateur d'adjacence	<ul style="list-style-type: none"> Entrez <i>chaîne1</i> dans le champ <i>Search word</i>. Entrez <i>chaîne2</i> dans le champ <i>context word</i>. Spécifiez 1 mot à gauche et à droite de <i>chaîne1</i>.

Repérage de cooccurrents d'une chaîne recherchée

1. Dans la fenêtre *Concord*, répétez la recherche pour la chaîne *cour*.

NOTE : N'oubliez pas de supprimer les mots contextuels de l'onglet *Advanced* pour avoir toutes les occurrences de la chaîne.

2. Cliquez sur l'onglet **CLUSTERS** pour voir des suites de chaînes récurrentes observées parmi les résultats. Celles-ci apparaissent en ordre de fréquence décroissante.
 - a. Quelles sont quelques-unes des unités intéressantes repérées par l'outil?
 - b. Comment cette fonction peut-elle être utile au traducteur ou au terminologue?
3. Ajustez les paramètres pour le calcul des groupes (*clusters*) :
 - a. Du menu **COMPUTE**, choisissez l'option **CLUSTERS**.
 - b. Dans la boîte de dialogue qui apparaît, spécifiez la longueur des suites de mots à repérer. (Il est recommandé d'essayer des groupes de 2 à 4 mots, mais vous pouvez essayer plus ou moins de mots.)

- c. Spécifiez si désiré la fréquence minimale pour l'identification des groupes.
 - d. Choisissez les horizons pour le calcul des groupes (c'est-à-dire, à l'intérieur de quelle distance du mot recherché les groupes récurrents seront repérés).
 - e. Cliquez sur le bouton **OK** pour recalculer les groupes.
 - f. Regardez la colonne *Related* à droite dans la fenêtre des résultats. Elle affiche des groupes de mots qui partagent une certaine partie des mots avec le groupe indiqué dans les résultats.
4. Cliquez sur l'onglet *Patterns* pour faire afficher les listes de mots qui apparaissent à proximité du mot recherché.
 - a. Comment ces renseignements peuvent-ils être utiles au traducteur ou au terminologue?
 5. Cliquez sur l'onglet **COLLOCATES** pour identifier des chaînes fréquemment observées à proximité de la chaîne recherchée.
 - a. Les collocatifs de la chaîne sont indiqués en ordre décroissant de fréquence dans la colonne à gauche, et leur emplacement par rapport à cette chaîne (en mots à gauche et à droite) à droite de chaque collocatif. L'emplacement le plus fréquent est indiqué en rouge.
 - b. Quels sont quelques-uns des collocatifs intéressants pour cette chaîne?
 - c. Est-ce que ces résultats mettent en évidence des cooccurrents que vous n'aviez pas encore remarqués?

NOTE : Le nombre de mots à analyser de chaque côté de la chaîne recherchée peut être défini en passant par l'onglet *Concord* dans les paramètres, et les collocatifs recalculés à l'aide de l'option **COLLOCATES** du menu **COMPUTE**.

6. Faites une recherche pour la chaîne *legislation*. Consultez les collocatifs repérés.
 - a. Combien des collocatifs que vous avez repérés manuellement au cours des exercices dans le *Tutoriel et exercice avec l'outil Concord de WordSmith Tools : niveau I* sur le site de la CERTT (**ACCÈS PAR NOM D'OUTIL > WORDSMITH TOOLS > WORDSMITH TOOLS CONCORD, NIVEAU I**) sont repérés automatiquement par l'outil?

Dernières étapes

1. Pour faire une copie de sauvegarde de vos fichiers ou pour les transférer sur un autre ordinateur à l'extérieur du Centre d'écriture :
 - a. À partir du *Poste de travail* de Windows ou le menu *démarrer*, ouvrez le U : (*Mes documents*) et trouvez le sous-répertoire où vous avez stocké vos fichiers.
 - b. Créez un dossier compressé qui contient ce sous-répertoire. (Pour des instructions à ce sujet, consultez le fichier *Créer un dossier compressé* sur le site de la CERTT (**ACCÈS PAR NOM D'OUTIL > WINDOWSXP > CRÉER UN DOSSIER COMPRESSE**)).
 - a. Copiez le dossier compressé sur une clé USB, ou, s'il fait moins de 2 Mo, envoyez-le en pièce jointe à votre adresse de courriel.

Questions de réflexion

1. Comment ces fonctions de recherche et d'analyse avancées se comparent-elles aux recherches manuelles que vous pouvez faire à l'aide de *Concord*? Pouvez-

vous accomplir plus ou moins les mêmes tâches manuellement? Lesquelles et comment?

2. Quelle option trouvez-vous :
 - a. la plus rapide?
 - b. la plus simple?
 - c. la plus efficace?
3. Pourquoi?
4. Quelles sont les difficultés qui peuvent être résolues au moins en partie par les fonctions avancées décrites ci-dessus? Comment?
5. Y-a-t-il des difficultés qui ne peuvent être prises en charge par ces fonctions?