



Tutoriel et exercices avec Concord de WordSmith Tools, niveau I

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 qui suivent vous montreront comment :

- [choisir les textes à exploiter](#);
- [faire des recherches simples](#);
- [manipuler les résultats](#);
- [consulter les contextes dont les résultats sont tirés](#).

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.

NOTE : Par défaut, l'option *.* est sélectionnée dans ce menu; cela signifie que tous les fichiers du répertoire sont affichés, peu importe leur nom ou l'extension qu'ils portent. Si vous sélectionnez *.txt, par exemple, seuls les fichiers qui portent l'extension .txt seront affichés; l'astérisque précédant le point indique cependant que tout nom de fichier sera accepté.

- 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.

NOTE : Si vous pensez devoir consulter cet ensemble de fichiers régulièrement, vous pouvez sauvegarder la liste en cliquant sur le bouton *Save favourites* (🔖). Vous pouvez même ajouter des commentaires sur ces fichiers dans la partie inférieure de la fenêtre. (Assurez-vous de stocker la liste de fichiers ainsi produite dans le sous-répertoire de votre U:, où elle sera disponible la prochaine fois.)


Les fichiers pourront alors être ajoutés en bloc la prochaine fois que vous utiliserez **Concord**, en ouvrant la liste à partir de l'icône *Get favourites* (📁). Le chargement peut prendre un petit moment, alors il faut être patient!

2. Une fois la sélection des fichiers terminée, sortez de la fenêtre *Choose texts* en cliquant sur l'icône de la coche verte (✅).

Production d'une concordance



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. Sur cet onglet, vous verrez un champ de recherche en haut, et, dans la partie inférieure de l'onglet, un rappel de quelques-unes des options de recherche offertes.

NOTE : WordSmith offre aussi la possibilité de chercher une série d'unités stockées dans un fichier, mais cette possibilité ne sera pas exploitée dans le cadre de ces exercices. Pour en savoir plus à ce sujet, consultez les fichiers d'aide.

4. Tapez *législation* dans le champ de recherche, puis cliquez sur **OK** () pour lancer l'analyse.
5. Les résultats apparaissent dans la fenêtre *Concord*. Chaque occurrence de *législation* apparaît sur une ligne distincte, avec cette chaîne alignée au centre de la ligne et indiquée en couleur. La source de chaque occurrence est affichée à droite dans la colonne *File*.
6. Observez les autres statistiques fournies. WordSmith calcule la position de l'occurrence affichée dans la phrase, dans le paragraphe, dans la section et dans le texte, et les exprime en valeurs brutes et en pourcentages. (Par exemple, une occurrence qui apparaît à la toute fin d'un texte aurait une position de 100 % dans la phrase, dans le paragraphe, dans la section et dans le texte. Une occurrence au tout début aurait une position de 0 %. Mais les valeurs brutes varient bien sûr selon le nombre de mots dans la phrase, le paragraphe et le texte.)





a. Pouvez-vous penser à des informations utiles que ces statistiques pourraient fournir?

Affichage des résultats

1. Pour ajuster la largeur des contextes, placez le curseur sur la division entre les colonnes *Concordance* et *Set* (il prend alors la forme d'une ligne verticale avec une flèche de chaque côté) et tirez vers la droite.
2. Pour ajuster la hauteur d'une ligne, placez le curseur sur la division entre cette ligne et celle d'en-dessous et tirez vers le bas.
3. Pour ajuster la hauteur de toutes les lignes de la concordance, ouvrez le menu **VIEW** et choisissez les options **GROW** (avec l'icône ) ou **SHRINK** (avec l'icône ). (Alternativement, appuyez sur la touche **F8** pour augmenter la hauteur de la ligne et **CTRL + F8** pour la diminuer.)
4. Si vous voulez voir uniquement la phrase dans laquelle la chaîne recherchée apparaît, choisissez l'option **SENTENCE ONLY** du menu **VIEW**.
5. Pour voir le texte intégral dont l'occurrence est tirée, double-cliquez sur la ligne en question pour la sélectionner. Le texte est affiché sur l'onglet **SOURCE TEXT** avec l'occurrence du terme surligné. Une fois la consultation terminée, cliquez de nouveau sur l'onglet **CONCORDANCE** pour retourner aux résultats.

6. Vous pouvez aussi classer certaines occurrences selon des classes que vous aurez vous-mêmes définies. Par exemple, vous pouvez marquer dans les occurrences de *législation* la présence de cooccurrents potentiellement intéressants de cette unité ou le type de cooccurrents présents.

NOTE : Ceci peut être très intéressant, par exemple, pour indiquer les parties du discours pour une chaîne qui présente une ambiguïté catégorielle ou pour différencier des cas intéressants ou problématiques de ceux qui ne le sont pas.

- a. Dans la fenêtre *Concord*, placez le curseur sur une ligne dans la colonne *Set*.
 - b. Tapez une lettre qui représentera la classe du cooccurrent que vous trouvez intéressant dans chaque contexte (s'il y a lieu). (Par exemple, vous pouvez utiliser *n* pour les cooccurrents nominaux, *v* pour les verbaux et ainsi de suite). Lorsque vous classez un contexte, le curseur avance à la ligne suivante.
 - c. Continuez à classer les occurrences en attribuant une lettre à chaque ligne. (Au n° 7, vous verrez comment trier les occurrences en fonction de ces classes, si désiré.)
7. Pour trier les résultats, à partir du menu **EDIT**, choisissez l'option **RESORT** (avec l'icône ) . (Alternativement, appuyez sur la touche **F6**.) Les occurrences sont automatiquement triées par ordre alphabétique selon le *Set*. (Celles sans *set* apparaissent en premier, suivi de ceux annotés *n*, suivi de ceux annotés *v*.)
 - a. Est-ce que ce genre de tri peut être utile pour étudier l'usage de l'unité cherchée? Comment?
 8. Pour sauvegarder les résultats d'une recherche sous forme d'une concordance ouvrable dans WordSmith Tools :
 - a. À partir du menu **FILE**, choisissez l'option **SAVE** (avec l'icône ) .
 - b. Dans la boîte de dialogue qui apparaît, choisissez le lecteur et naviguez à l'emplacement désiré dans l'arborescence de fichiers.
 - c. Tapez un nom de fichier dans le champ.
 - d. Cliquez sur le bouton **ENREGISTRER** pour sauvegarder le fichier.
 9. Pour sauvegarder les résultats d'une recherche sous forme d'une concordance en format Texte brut (.txt) qui pourra être visualisée et éditée dans un traitement de texte ou importée dans une base de données :
 - a. À partir du menu **FILE**, choisissez l'option **SAVE AS... > PLAIN TEXT** (avec l'icône ) .
 - b. Cliquez sur le bouton du dossier jaune () pour choisir le lecteur désiré et naviguez dans le sous-répertoire que vous avez créé pour faire ces exercices.
 - c. Tapez un nom de fichier et cliquez sur le bouton **OUVRIR**.
 - d. De retour dans la boîte de dialogue *Save as Plain text*, cliquez sur **OK** pour sauvegarder le fichier.

NOTE : Si vous enregistrez le fichier en format .txt, vous pouvez délimiter la longueur du contexte à extraire à partir de la fenêtre *Concord* (menu **SETTINGS > SPECIFIC TO CONCORD > CHARACTERS TO SAVE (PER ENTRY) > OK**).

10. Triez les occurrences selon un autre critère.

NOTE : Comme vous avez vu ci-dessus, lorsque des *Set* ont été attribués, WordSmith les utilisera comme critère de tri par défaut lorsque vous demandez un tri. Pour avoir accès aux

autres critères de tri disponibles, commencez en éliminant les *Set*. (Notez que vous avez sauvegardé une copie de la concordance annotée, donc vous pourrez toujours y référer si vous voulez.)

- a. À partir du menu **EDIT**, choisissez l'option **CLEAR SET COLUMN**. Les *Set* ajoutés disparaissent.
- b. À partir du menu **EDIT**, choisir l'option **RESORT**. Dans la boîte de dialogue *Concordance Sort* qui apparaît, vous pouvez spécifier jusqu'à trois critères de tri en cliquant sur les trois onglets un à un :
 - i. **FILE** trie les occurrences selon le fichier source des occurrences;
 - ii. **R1** trie les occurrences par ordre alphabétique selon le premier mot à droite de la chaîne recherchée, **R2** par le deuxième mot à droite, et ainsi de suite;
 - iii. **L1** trie les occurrences par ordre alphabétique selon le premier mot à gauche de la chaîne recherchée, **L2** par le deuxième mot, et ainsi de suite;
- c. La case **CASE SENSITIVE** permet de prendre en compte la casse lors du tri, si désiré.

11. Triez les résultats pour mettre en évidence différents cooccurents de *législation* :


NOTE : N'oubliez pas de cocher la case **ACTIVATE** pour chaque critère que vous voulez utiliser pour trier la concordance.

- a. Triez les occurrences par le premier mot à droite de *législation* en choisissant l'option **R1** et en cliquant sur le bouton **OK**.
 - i. Quels cooccurents repérez-vous dans les contextes?
 - ii. À quelle(s) catégorie(s) grammaticale(s) appartiennent-ils?
 - iii. Est-ce que cela correspond à l'étiquetage que vous avez fait des cooccurents dans la colonne *Set*?
 - iv. En quoi ces cooccurents peuvent-ils être utiles au traducteur ou au terminologue?
- b. Triez les occurrences par le troisième mot à gauche de *législation*.
 - i. Quelle option faut-il choisir pour faire ce tri?
 - ii. Quels cooccurents repérez-vous dans les contextes?
 - iii. À quelle(s) catégorie(s) grammaticale(s) appartiennent-ils?
 - iv. Est-ce que cela correspond à l'étiquetage que vous avez fait des cooccurents dans la colonne *Set*?
 - v. En quoi ces cooccurents peuvent-ils être utiles au traducteur ou au terminologue?
- c. Avez-vous relevé d'autres cooccurents pertinents de *législation*? Pouvez-vous trouver une façon de trier les occurrences pour les mettre en évidence?

NOTE : Vous auriez intérêt à noter les cooccurents que vous avez relevés pour cette unité. Vous verrez en effet dans le *Tutoriel et exercices avec Concord de WordSmith Tools : Niveau II* sur le site de la CERTT (ACCÈS PAR NOM D'OUTIL > WORDSMITH TOOLS > WORDSMITH TOOLS CONCORD, NIVEAU II) une autre façon d'identifier des cooccurents, cette fois automatiquement.

12. Pour éliminer des résultats non pertinents de la liste :

- a. Cliquez sur la ligne en question pour la sélectionner.

- b. Appuyez sur la touche **DEL** pour « désactiver » la ligne. (Si vous changez d'avis par la suite, sélectionnez la ligne et appuyez sur la touche *Ins* pour la « réactiver »).
- c. Répétez si désiré pour d'autres lignes à supprimer.
- d. Ouvrez le menu **EDIT** et choisissez l'option **ZAP** (avec l'icône ) pour supprimer les lignes désactivées des résultats.

NOTE : Vous pouvez aussi accéder à la fonction **ZAP** par le raccourci clavier **CTRL+Z**. Vous allez sans doute reconnaître ce raccourci utilisé par défaut par les applications Microsoft pour annuler une action. Il est important de ne pas l'utiliser dans WordSmith, à moins de vouloir « zapper » les occurrences supprimées.

13. Pour fermer la concordance sans la sauvegarder, choisissez l'option **CLOSE** du menu **FILE** et cliquez sur le bouton **NO** lorsque l'outil vous demande si vous voulez enregistrer le fichier.

NOTE : À cette étape, si vous désirez continuer tout de suite avec WordSmith Tools, vous pouvez commencer le *Tutoriel et exercices avec Concord de WordSmith Tools : Niveau II* sur le site de la CERTT (**ACCÈS PAR NOM D'OUTIL > WORDSMITH TOOLS > WORDSMITH TOOLS CONCORD, NIVEAU II**), en sautant les premières sections sur le téléchargement de fichiers et la construction de corpus et en débutant avec les premières recherches suggérées.

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 du menu *démarrer*, ouvrez le lecteur U: (*Mes documents*) et trouvez le sous-répertoire dans lequel 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 > EXTRAIRE DES FICHIERS D'UN DOSSIER COMPRESSE**).)
 - a. Copiez le dossier compressé sur une clé USB ou sur une disquette; s'il fait moins de 2 Mo, vous pouvez l'envoyer en pièce jointe à votre courriel.

Questions de réflexion

1. Au cours de ces exercices, qu'avez-vous remarqué quant au fonctionnement de Concord? Que pensez-vous de son interface? Des options de recherche qu'il vous offre? Des options d'affichage et de traitement des résultats?
2. Qu'est-ce que Concord pourrait vous aider à faire? Dans quelles conditions?
3. Quels sont les principaux avantages ou inconvénients liés à l'utilisation de Concord pour interroger des banques de textes? Par rapport à une approche manuelle? Par rapport à l'utilisation d'un autre outil?
4. Quelles sont les principales difficultés qui se posent lorsqu'on utilise un concordancier? Ces difficultés ont-elles une incidence sur la façon de faire les recherches et d'exploiter les résultats?
5. Quels critères pourriez-vous utiliser pour évaluer des concordanciers unilingues?



6. Pouvez-vous comparer Concord à d'autres concordanciers unilingues que vous connaissez?