



Exercice avec le concordancier de TextSTAT : niveau I

TextSTAT est un outil d'analyse de corpus qui peut être téléchargé gratuitement à partir du site de son concepteur, Matthias Hüning, de la Free University of Berlin (<http://www.niederlandistik.fu-berlin.de/textstat/software-en.html>). TextSTAT permet de constituer des corpus à partir de fichiers stockés sur votre poste de travail ou directement à partir d'Internet, et offre un dépouilleur lexical ainsi qu'un concordancier KWIC.

Ces exercices vous aideront à :

- [installer TextSTAT](#);
- [construire un corpus à partir de textes enregistrés sur votre poste de travail](#);
- [faire des recherches simples avec TextStat](#);
- [ajuster la présentation des résultats](#);
- [faire des recherches multicritères avec opérateurs de proximité](#); et
- [exporter les résultats de vos recherches en le exportant vers un document Word](#).

Préparation

1. Enregistrez les fichiers dont vous aurez besoin pour ce tutoriel :
 - a. Créez dans le lecteur U : (aussi appelé *Mes documents*) un sous-répertoire pour stocker les fichiers utilisés dans le cadre de ce tutoriel. (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).)
 - b. Téléchargez à partir du site de la CERTT le dossier compressé qui contient des Décisions de la Cour suprême du Canada en français et en format .txt (CORPUS ET AUTRES RESSOURCES > DÉCISIONS DE LA COUR SUPRÊME DU CANADA > SUPREME COURT DECISIONS (FRENCH, TXT)).


NOTE : TextSTAT peut traiter des fichiers en format Word (.doc), HTML (.htm/.html) ou Texte seulement (.txt).

- c. Extrayez les fichiers dans le sous-répertoire que vous venez de créer. (Pour des instructions à ce sujet, consultez le document *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).)

Installation de TextSTAT

1. Consultez la page Web qui décrit TextSTAT (<http://neon.niederlandistik.fu-berlin.de/en/textstat/>) et lisez son texte de présentation.
2. Téléchargez dans votre lecteur U : (*Mes documents*) le dossier compressé qui contient la version 2.8g de TextSTAT pour Windows.
3. Ouvrez le répertoire *Mes documents* (le U :) à partir du raccourci qui apparaît sur le Bureau de votre ordinateur et double-cliquez sur le dossier compressé que vous avez téléchargé pour extraire automatiquement tous les fichiers qu'il contient.

Gardez l'emplacement suggéré par défaut pour les fichiers extraits en cliquant sur le bouton **OK** dans la boîte de dialogue qui apparaît. Ceci créera automatiquement un sous-répertoire *TextSTAT – 2* où seront stockés les fichiers extraits.

4. Ouvrez maintenant le logiciel : dans la liste de fichiers extraits, double-cliquez sur le fichier principal (celui appelé *TextSTAT.exe* qui est représenté par l'icône ). Le logiciel s'ouvre et est prêt à l'exploitation.

NOTE : Pour utiliser TextSTAT plus tard, il suffira de retrouver ce fichier et de double-cliquer dessus pour ouvrir le logiciel.



Si vous installez l'outil chez vous, vous pouvez aussi créer un raccourci sur votre Bureau pour en faciliter l'accès :

- a. Avec le bouton droit de la souris, cliquez sur le fichier principal.
- b. À partir du menu contextuel qui apparaît, choisissez l'option **ENVOYER VERS...** et puis **BUREAU (CRÉER UN RACCOURCI)**.


Construire un corpus à partir de textes enregistrés sur votre poste de travail

1. Créez un nouveau corpus (c'est-à-dire, une liste de fichiers) :

NOTE : Vous devez commencer par créer une liste vide, à laquelle vous allez par la suite ajouter des fichiers.

- a. Si votre logiciel n'est pas déjà ouvert en français, cliquez sur le menu **LANGUAGE**, et choisissez l'option **FRANÇAIS** (vous devrez redémarrer TextStat pour que le changement soit effectif),
- b. À partir du menu **CORPUS**, choisissez l'option **NOUVEAU CORPUS** (ou cliquez sur le bouton **NOUVEAU CORPUS** ).
- c. Dans la boîte de dialogue qui s'ouvre, tapez un nom pour ce nouveau corpus et choisissez l'emplacement qui vous convient (le sous-répertoire que vous avez créé au début des exercices).
- d. Cliquez sur le bouton **ENREGISTRER LE CORPUS** () pour sauvegarder le nouveau corpus vide. Une nouvelle boîte de dialogue s'ouvrira pour vous aviser que vous pouvez maintenant ajouter des fichiers à votre corpus. Cliquez sur le bouton **OK** pour revenir à la fenêtre principale de TextStat.

2. Spécifiez les fichiers à inclure dans le corpus :

- a. À partir du menu **CORPUS**, choisissez l'option **AJOUTER UN FICHIER** (ou cliquez sur le bouton **AJOUTER UN FICHIER** ).
- b. Dans la boîte de dialogue qui s'ouvre, sélectionnez l'ensemble des fichiers de décisions de la Cour suprême que vous avez extraits du dossier compressé et cliquez sur le bouton **OUVRIER** pour les ajouter à la liste de fichiers du corpus. (Vous pouvez sélectionner l'ensemble des fichiers en enfonçant les touches **CTRL + A** de votre clavier.)

3. Sauvegardez la définition du corpus (c'est-à-dire, la liste des fichiers) :

- a. À partir du menu **CORPUS**, choisissez l'option **ENREGISTRER LE CORPUS**.

Faire des recherches simples avec TextStat

1. Cliquez sur l'onglet **OCCURRENCES**.

2. Faites une recherche simple pour le terme *légal* dans le corpus :

- a. Dans le champ de recherche qui apparaît en haut de la fenêtre, tapez *légal*.
- b. Cliquez sur le bouton **RECHERCHER**
- c. Une liste des occurrences de *légal* apparaît dans la partie gauche de la fenêtre. La barre d'état en bas de la fenêtre de TextSTAT affiche le nombre total d'occurrences. Familiarisez-vous avec les résultats. *Que remarquez-vous à propos des occurrences trouvées? Se limitent-elles au terme *légal*?*

Pour visualiser un plus long contexte entourant une des occurrences, cliquez sur la ligne en question pour d'abord la sélectionner. Avec le bouton droit de la souris et à partir du menu contextuel qui apparaît, choisissez ensuite l'option **MONTREZ PLUS DE CONTEXTES**. Un segment plus long du texte est affiché dans l'onglet *Contexte* . Une fois que vous aurez parcouru le contexte et les autres données fournies, cliquez de nouveau sur l'onglet **OCURRENCES** pour retourner aux résultats.

3. Ajustez les paramètres de recherche :


- a. Dans la section droite de la fenêtre, cochez la case **RECHERCHER MOTS ENTIERS SEULEMENT** afin de chercher la chaîne de mots telle que vous l'avez écrite exactement (au lieu de chaînes non-définies, c'est à dire de chaînes qui contiennent celle que vous avez entrée soit au début, soit au milieu ou à la fin);
- b. Cochez la case **RECHERCHER SANS DISTINCTION MAJUSCULES/MINUSCULES** afin de chercher la chaîne spécifiée sans tenir en compte la casse (les minuscules et majuscules) ;
- c. Cliquez sur le bouton **RECHERCHE** pour voir la différence dans les résultats.
- d. Évaluez cette différence. *Comment ces fonctions peuvent-elles aider un traducteur ou un autre professionnel langagier à cibler plus facilement de l'information pertinente? Dans quelles conditions?*
- e. *Peut-il y avoir des inconvénients à utiliser ces fonctions?*

Ajuster la présentation des résultats

1. Dans la partie droite de la fenêtre, ajustez les différents paramètres de l'affichage :

- a. L'option **MARQUER LA CHAÎNE RECHERCHÉE** permet d'afficher la chaîne recherchée toute en majuscules;
- b. Les paramètres du contexte (*contexte gauche* et *contexte droit*) permettent de spécifier le nombre de caractères à afficher à gauche et à droite de la chaîne recherchée;
- c. Les options en dessous des paramètres du contexte permettent d'afficher les occurrences de la chaîne recherchée en les triant à partir du premier mot à sa gauche, ou à sa droite (ordre alphabétique).
- d. Modifiez certains de ces paramètres et cliquez sur le bouton **ACTUALISER** en dessous des paramètres d'affichage pour mettre à jour les données selon les options que vous avez choisies.
- e. Évaluez le résultat de ces ajustements. *Comment pourraient-ils aider un traducteur ou un autre langagier à cibler différentes informations pertinentes?*

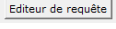
2. Copiez une ligne de la concordance :

- a. Toujours à partir de l'onglet *Occurrences*, sélectionnez une ligne en cliquant dessus dans la fenêtre des résultats.
- b. Cliquez sur le bouton **COPIER DANS PRESSE-PAPIER** ().

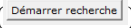
- c. Collez les données dans un fichier de traitement de texte, dans une fiche terminologique, etc. (généralement à l'aide du raccourci **CTRL + V**).

Faire des recherches multicritères avec des opérateurs de proximité

1. Faites une recherche multicritère avec des opérateurs de proximité (une recherche pour des contextes qui contiennent deux chaînes à une certaine distance l'une de l'autre) :

- a. Cliquez sur le bouton **ÉDITEUR DE REQUETE** () à droite du champ de recherche.
- b. Dans le premier champ de recherche de la boîte de dialogue qui s'ouvre, tapez *jugement*.

NOTE : Si vous le désirez, vous pouvez utiliser la troncature (des caractères de remplacement): ? remplace 0 ou 1 caractère et * peut remplacer un ou plusieurs caractères.

- c. Dans le deuxième champ de recherche de la boîte de dialogue qui s'ouvre, tapez *prononc**.
- d. À partir des listes en dessous des champs, choisissez 0 comme le nombre minimal et 5 comme le nombre maximal de mots qui peuvent apparaître entre les deux chaînes spécifiées : ceci permet de chercher les deux chaînes quand elles apparaissent de manière contiguë (0 mot entre les chaînes) et quand elles sont séparées par 1 à 5 mots.
- e. En bas de la boîte de dialogue, cliquez sur le bouton **DEMARRER RECHERCHE** () , et puis sur le bouton **QUITTER** pour revenir à la fenêtre des résultats.

2. Dans la fenêtre des résultats, regardez comment TextSTAT a transformé automatiquement votre recherche en expression régulière (un formalisme qui permet de faire des représentations formelles de chaînes). **Pouvez-vous identifier comment les divers éléments de la requête que vous avez construite dans l'éditeur de requête ont été convertis en expressions régulières?**

- a. Retournez dans l'**ÉDITEUR DE REQUETE** et inversez l'ordre des chaînes avant de répéter la requête. **En quoi les résultats ont-ils changé? Selon vous, quelles sont les conclusions à tirer du potentiel de TextSTAT à faire des recherches complexes? Quels sont les points forts et faibles de l'ÉDITEUR DE REQUETE?**
- b. Pour en savoir plus sur les options d'interrogation à l'aide d'expressions régulières, cliquez sur le bouton d'aide (?) dans la barre de menu, puis choisissez l'option **EXPRESSIONS REGULIERES – AIDE**. Une page Web s'affiche décrivant les options disponibles (en anglais seulement).

3. Comparez les paramètres par défaut pour les différents types de recherches :

- a. Identifiez une unité lexicale qui apparaît sous diverses formes dans le corpus (ex. singulier et pluriel, masculin et féminin, avec et sans majuscules). Essayez *pourvoi*.
- b. Faites une recherche simple pour cette unité lexicale dans sa forme de base, avec les options **RECHERCHER MOTS ENTIERS SEULEMENT** et **RECHERCHER SANS DISTINCTION MAJUSCULE/MINUSCULE** cochées. Regardez les résultats.
- c. Refaites la même recherche, mais cette fois avec ces deux options décochées. Comparez les résultats aux premiers. **En quoi diffèrent-ils?**
- d. Laissez les options décochées. Cliquez sur le bouton **ÉDITEUR DE REQUETE** et entrez la même chaîne dans le premier champ de recherche. Lancez la recherche (**DEMARRER RECHERCHE, QUITTER**).

- e. Regardez les résultats ainsi que les options qui sont maintenant sélectionnées. Quelles sont les restrictions imposées sur des recherches à l'aide de l'ÉDITEUR DE REQUETE? Quelles sont les implications de ceci quant aux possibilités de recherche à l'aide de cette fonction?

Exporter les résultats de vos recherches en les exportant vers un document Word

1. Exportez le produit de vos recherches, si désiré :
 - a. À partir du menu **EXPORTER**, choisissez l'option **LISTE D'OCCURRENCES > MS WORD**.
 - b. Un fichier Word s'ouvre automatiquement contenant une copie de la liste complète des contextes affichés dans la fenêtre de vos résultats. Sauvegardez ce fichier dans le sous-répertoire que vous avez créé dans votre U : (**FICHIER > ENREGISTRER SOUS...**) en lui attribuant le nom de votre choix.

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 COMPRESSÉ**)).
 - c. Copiez le dossier compressé sur une clé USB ou, s'il fait moins de 2 Mo, envoyez-le en pièce jointe à votre courriel.

Questions pour réflexion

1. D'après vous, TextSTAT est-il facile à utiliser?
2. Que pensez-vous des fonctions de base disponibles?
3. Comment TextSTAT peut-il aider à analyser des banques de textes/corpus? Quelles tâches peut-il faciliter?
4. Quels critères peut-on utiliser pour comparer ce concordancier à d'autres (par exemple, le concordancier unilingue de LogiTerm ou WordSmith Tools)? À l'aide de ces critères, comparez TextSTAT aux autres concordanciers que vous connaissez.