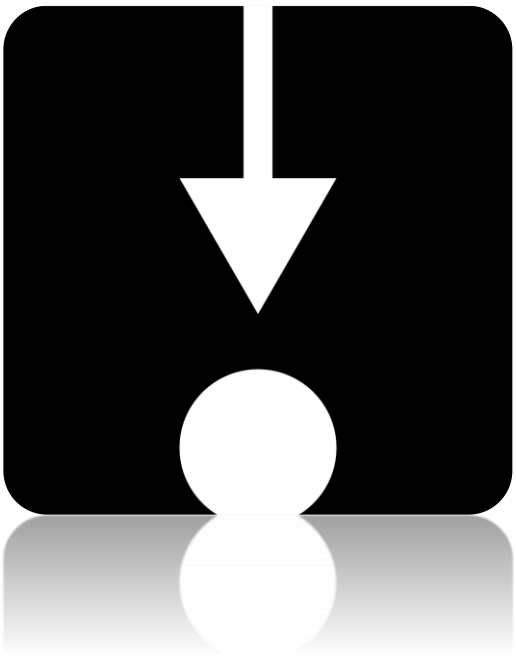
Guide de l'utilisateur #LancsBox 4.5



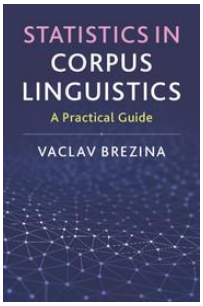
Traduction: Dr. Caroline Rossi (Université Grenoble Alpes) et Mignot Lorène, Bordat Charline, Da Cunha Belves Claire, Livion Marcia, Velado Pamela, Dolmazon Gaëlle, Dezulier Diane, Nguyen Rémy, Simon Romain, Van Essen Maaike, Roubaud Laura, Malin Gael, Sorli Liselotte, Yermenko Anna, Shindina Kristina, Cebo Jean-Paul, Esclaine Alexia, Degrave Emmanuelle.

#LancsBox v.4.5: licence

Lorsque vous publiez des études réalisées à l'aide de #LancsBox, veuillez citer l'une des deux sources suivantes :

- ☐ Brezina, V., McEnery, T. & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139-173
- ☐ Brezina, V., Timperley, M., & McEnery, A. (2018). #LancsBox v. 4.x. [software package]

Aide statistique



Brezina, V. (2018). *Statistics for corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.
Pour en savoir plus sur les procédures statistiques utilisées en linguistique de corpus, consultez Brezina (2018), ou le site Lancaster Stats Tools http://corpora.lancs.ac.uk/stats

Lectures et documents complémentaires

Brezina, V. (2016). Collocation Networks. In Baker, P. & Egbert, J. (eds.) *Triangulating Methodological Approaches in Corpus Linguistic Research*. Routledge: London.

Brezina, V. (2018). Statistical choices in corpus-based discourse analysis. In Taylor, Ch. & Marchi, A. (eds.) Corpus approaches to discourse: a critical review. Routledge: London.

Brezina, V. & Gablasova, D. (2017). The corpus method. In: Culpeper, J, Kerswill, P., Wodak, R., McEnery, T. & Katamba, F. (eds). *English Language (2nd edition).* Palgrave.

Brezina, V., McEnery, T. & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139-173.

Brezina, V., & Meyerhoff, M. (2014). Significant or random. *A critical review of sociolinguistic generalisations based on large corpora. International Journal of Corpus Linguistics*, *19*(1), 1-28.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, *67 (S1),* 155–179.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring learner language through corpora: comparing and interpreting corpus frequency information. *Language Learning, 67 (*S1), 130–154.

D'autres documents (vidéos de conférences, exercices, diapositives, etc.) sont disponibles sur le site Web #LancsBox: http://corpora.lancs.ac.uk/lancsbox/materials.php

1 Télécharger et lancer #LancsBox version 4.5

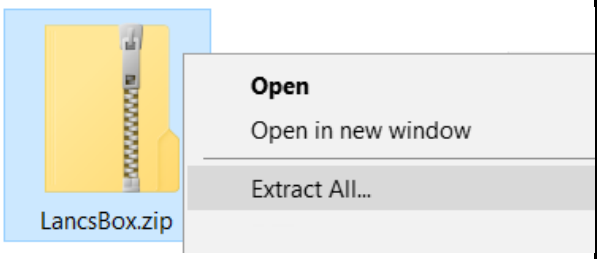
#LancsBox est un outil d'analyse de corpus nouvelle génération. La version 4.5 a été conçue principalement pour les systèmes d'exploitation 64 bits (Windows 64 bits, Mac et Linux) qui permettent un fonctionnement optimal. #LancsBox est aussi compatible avec les systèmes 32 bits mais ses performances seront limitées. La procédure de téléchargement et d'exécution est simple. Elle comporte trois étapes : 1) téléchargement, 2) extraction et 3) lancement.

1) Sélectionnez la version correspondant à votre système d'exploitation et téléchargez-la sur votre ordinateur.

2) Extrayez les fichiers du dossier «LancsBox» (dézippez)

Windows :

Faites un clic droit sur le fichier 'LancsBox.zip' que vous avez téléchargé [notez que l'extension .zip peut être cachée sous Windows], et sélectionnez " Extraire les fichiers".



Mac :

Double-cliquez sur LancsBox.app.tar

Remarque: assurez-vous d'avoir correctement dézippé «LancsBox». #LancsBox ne se lancera pas si le fichier .zip a été ouvert par un simple double-clic ou via le bouton 'Ouvrir'.

3. **Lancez #LancsBox** : Selon votre système d'exploitation, effectuez les opérations suivantes.
Windows (toutes versions) :
> Double-cliquez sur « LancsBox.bat » [notez que l'extension .bat peut être cachée sous Windows].
Mac :
> Copiez l'application LancsBox des Téléchargements vers les Applications
> Double-cliquez sur l'application LancsBox
> Autorisez #LancsBox à s'exécuter en donnant les droits d'accès de système de sécurité appropriés.

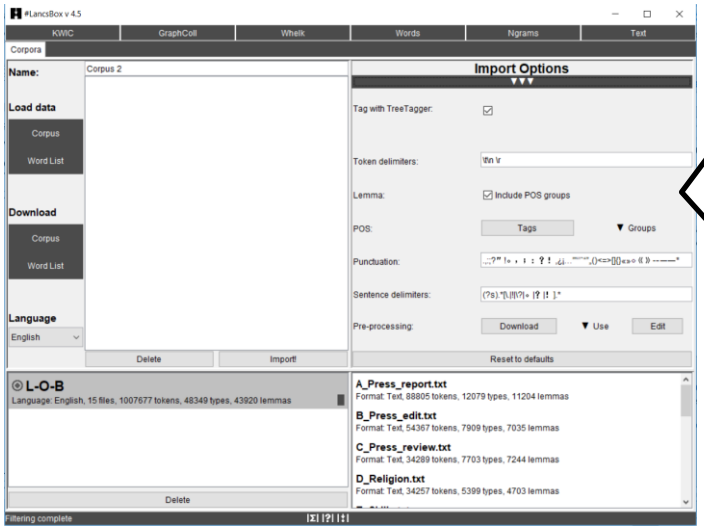
. Cliquez sur l'icône [ajoutez l'icône ici]
. Sélectionnez « Préférences système » > « Sécurité et confidentialité »
. Autorisez l'exécution de #LancsBox

Linux :
> Assurez-vous d'avoir installé oracle JDK/JVM (n'utilisez pas OpenJDK/JVM)
> Autorisez l'exécution de Lancsbox.jar et des fichiers contenus dans resources/tagger/bin
> Lancez 'LancsBox.jar'

2. Charger et importer des données

L'onglet corpus permet de charger et d'importer des données dans #LancsBox. Il s'ouvre automatiquement lorsque vous lancez #LancsBox. #LancsBox fonctionne avec des corpus de différents format (.txt, .xml, .doc, .docx, .pdf, .odt, .xls, .xlsx, .zip etc.) ainsi qu'avec des lexiques (.csv). Deux solutions sont possibles pour charger corpus et lexiques : 1) chargez vos propres données et 2) téléchargez des corpus et des lexiques disponibles sur #LancsBox.

2.1 Résumé visuel de l'onglet *Corpora*



Panneau supérieur : Importer des textes et listes de mots

Vous pouvez:

§ Charger votre corpus ou liste de mots.
§ Télécharger un corpus et listes de mots disponibles sur #LancsBox.
§ Choisir la langue.
§ Examiner les étiquettes morpho-syntaxiques.
§ Réviser les signes de ponctuation et les délimiteurs de phrases.
§ Définir les catégories de base (occurence, lemme, partie du discours, ponctuation).

Panneau inférieur: Travailler avec des corpus et listes de mots.

Vous pouvez:

§ Activer ou supprimer les corpus ou listes de mots importés.
§ Visualiser le corpus et le contenu de vos textes (occurrences, types, lemmes).
§ Prévisualiser des textes.

2.2 Charger vos corpus et listes de mots

#LancsBox vous permet de travailler facilement avec vos propres corpus et listes de mots. Ces corpus sont ceux qui sont stockés sur votre ordinateur ou à un endroit accessible depuis celui-ci (clé USB, disque partagé, Dropbox, cloud, etc.).

2.3 Types de fichiers pris en charge

#LancsBox prend en charge différents formats de fichier (txt, .xml, .doc, .docx, .pdf, .odt, .xls, .xlsx, .zip et beaucoup d'autres). #LancsBox extrait et traite le texte disponible dans les fichiers de corpus de façon automatique. Pour les listes de mots, #LancsBox est compatible avec le format texte séparé par des virgules (.csv).

2.4 Télécharger le corpus et les listes de mots #LancsBox

#LancsBox vous permet de travailler avec les corpus existants, disponibles gratuitement sur #LancsBox sous licence spécifique. Nous ajoutons régulièrement de nouveaux corpus.

2.5 Travailler avec le corpus et les listes de mots

Tous les corpus et les lexiques importés dans #LancsBox s'affichent sur le panneau en bas dans l'onglet « Corpora ». Le panneau permet de visualiser le corpus, de prévisualiser les fichiers et de recharger le corpus et les listes de mots rapidement lorsqu'on ferme ou rouvre le corpus.

4 L'outil KWIC (key word in context)

L'outil Mot-clé en contexte (KWIC) génère une liste de toutes les occurrences du terme recherché dans le corpus, sous la forme d'un concordancier.

Il peut par exemple être utilisé pour:

- Trouver la fréquence d'utilisation d'un mot ou d'une phrase dans un corpus
- Trouver les fréquences d'utilisation de différentes catégories lexicales, comme les noms, les verbes ou les adjectifs.
- Trouver des structures linguistiques complexes comme les structures passives ou des structures infinitives dans lesquelles un adverbe est intercalé entre "to" et le verbe, à l'aide de la fonction "smart searches" (recherche intelligente)
- Trier, filtrer ou présenter de manière aléatoire des lignes de concordance
- Réaliser une analyse statistique de l'usage d'un terme dans deux corpus

4.1 Résumé visuel de l'onglet KWIC

Recherche simple

Vous pouvez:
rechercher un mot ou un phrase
rechercher des nombres dans une plage de valeurs (par ex: >1930&<=1945)
utiliser le caractère générique * (par ex: nouveau*)
utiliser des expression régulières sensibles aux majuscules (par ex: /[abc].*/)
utiliser des expression régulières qui ne sont pas sensibles aux majuscules (ex: /dog|cat/i)
rechercher des marques de ponctuation (par ex: /.*\./p)
utiliser la fonction recherche intelligente (par ex: PASSIVES, NOUNS)

Recherche avancée

Vous pouvez:
effectuer des recherches sur différents niveaux d'annotation
effectuer des recherches combinées de termes, à plusieurs niveaux
utiliser des expressions régulières (ex: /N.*/)

4.2 Rechercher et afficher des résultats

#LancsBox permet d'effectuer des requêtes complexes. La barre de recherche peut être utilisée pour effectuer des recherches simples et plus avancées sur différents niveaux d'annotation.

4.3 Paramètres et fenêtre plein texte

Les paramètres de l'outil KWIC intègrent les options suivantes : Corpus (Corpora), Contexte (Context) et Affichage (Display). L'outil KWIC permet également d'ouvrir des fenêtres contextuelles en plein texte.

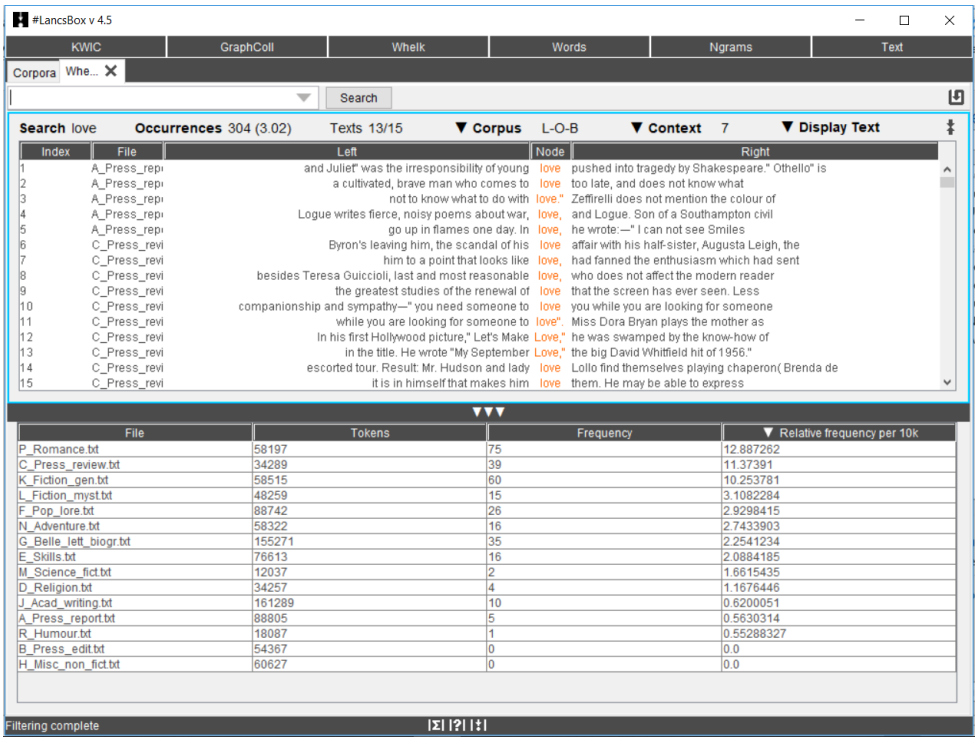
5 Outil Whelk

L'outil Whelk fournit des informations sur la façon dont le terme de recherche est réparti entre les fichiers de corpus.

Il peut être utilisé, par exemple, pour :

- trouver les fréquences absolues et relatives du terme recherché dans les fichiers du corpus,
- filtrer les résultats selon différents critères,
- trier les fichiers selon les fréquences absolues et relatives du terme recherché.

5.1 Résumé visuel de l'onglet Whelk



Panneau supérieur : Recherche de corpus

Vous pouvez :

§ Rechercher, trier et filtrer.

§ Utiliser les fonctions de recherche simple et avancée.

§ Effectuer des recherches « intelligentes ».

Panneau inférieur : Affichage de la répartition

Vous pouvez :

§ Visualiser la distribution du terme recherché dans les fichiers individuels.

§ Trier, filtrer et copier/coller.

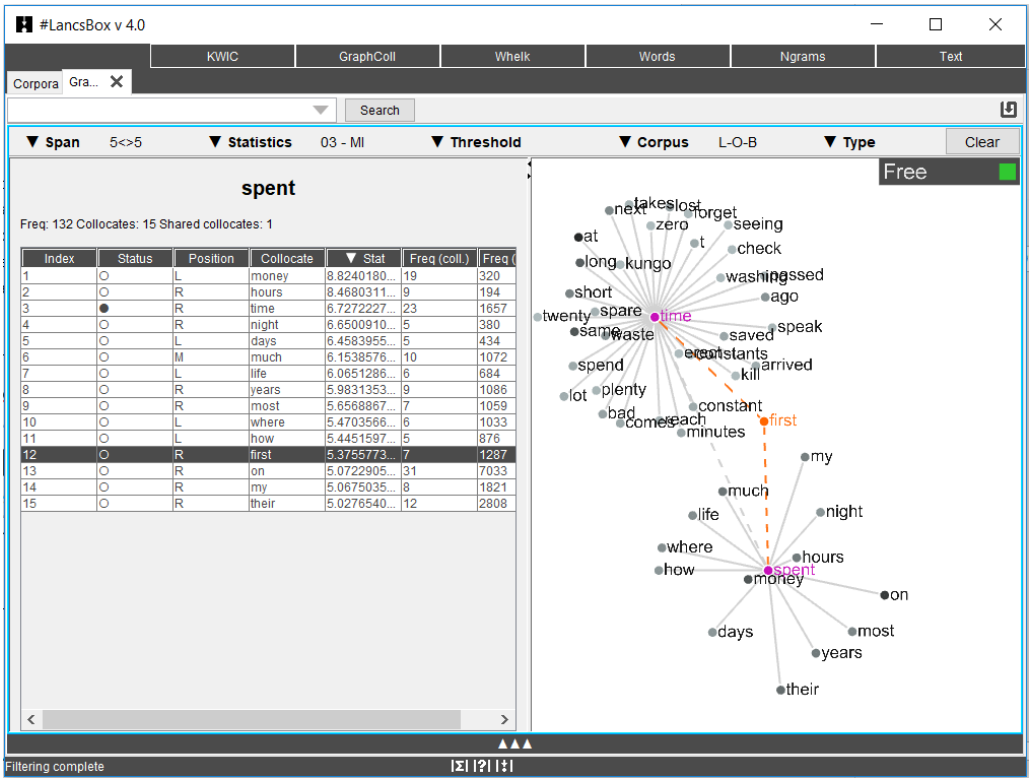
6. GraphColl

L'outil intitulé GraphColl permet d'identifier les collocations, de les afficher dans un tableau et sous forme de graphique ou de réseau de collocations.

Il peut par exemple être utilisé pour :

- trouver les collocations de certains mots ou groupes de mots
- trouver des colligations (cooccurrences de catégories grammaticales)
- visualiser les collocations et les colligations
- identifier les collocations communes de certains mots ou groupes de mots
- résumer des productions discursives sur la base des contenus et des thèmes abordés ("aboutness").

6.1 Résumé visuel de l'onglet GraphColl



6.2 Produire un graphique de collocations

Graphcoll produit des graphiques de collocations à la volée. Après avoir sélectionné les paramètres appropriés, vous pouvez commencer à rechercher le noeud de votre choix et ses collocations.

6.4 Lecture du graphique des collocations

Le graphique affiche trois catégories : I) force d'association, II) fréquence de collocation et III) position des collocations. Pour obtenir plus d'informations sur une collocation, un clic droit vous donne accès aux lignes de concordance (KWIC), dans lesquelles les collocations et le noeud sont cooccurrents.

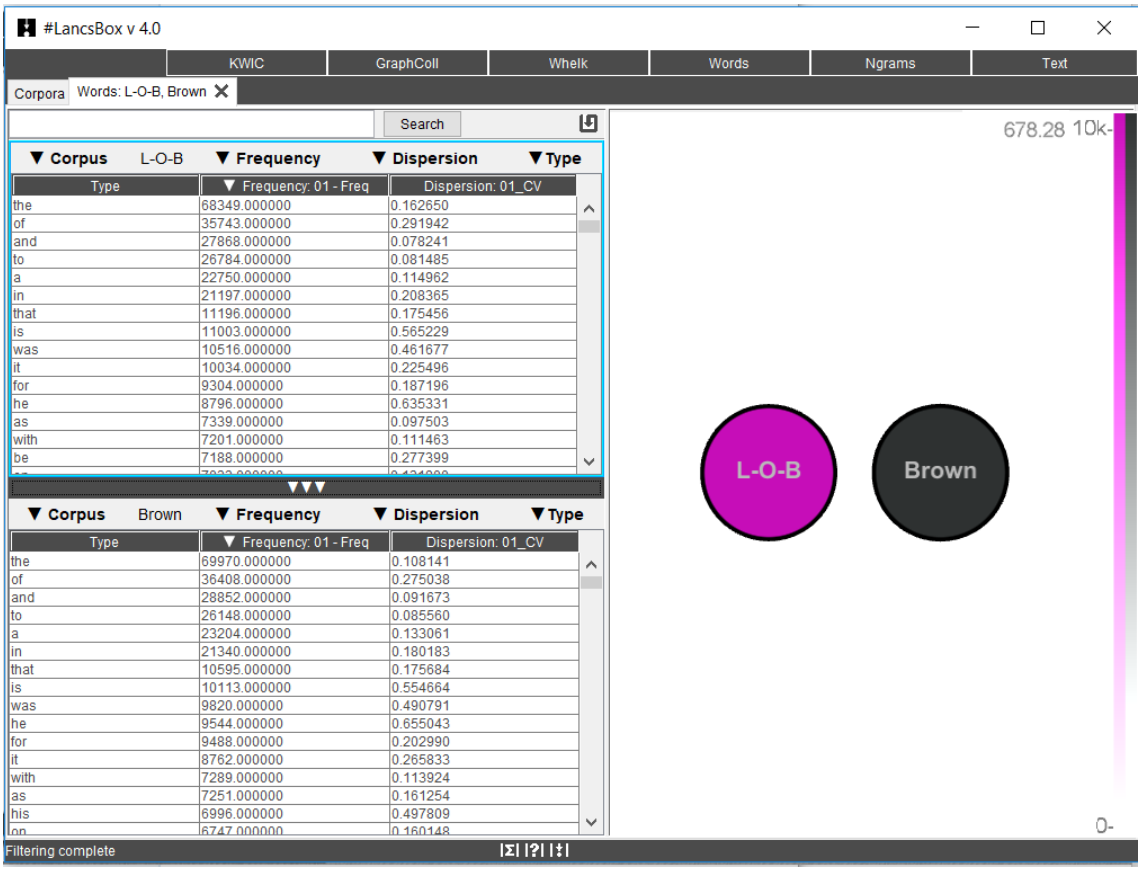
7 L'outil Words

L'outil Words permet une analyse approfondie des fréquences des types, lemmes et parties du discours, ainsi qu'une comparaison des corpus à partir du calcul des mots-clés.

Par exemple, l'outil peut être utilisé pour :

- calculer des mesures de fréquence et de dispersion pour les types, les lemmes et les étiquettes morpho-syntaxiques,
- visualiser la fréquence et la dispersion dans plusieurs corpus.
- Comparer des corpus en utilisant les mots-clés.
- Visualiser les mots-clés.

7.1 Résumé visuel



Partie gauche : création de listes de fréquence, calcul de la dispersion et des mots-clés.

Partie droite : visualisation des fréquences, des dispersions et des mots-clés.

8 L'outil Ngram

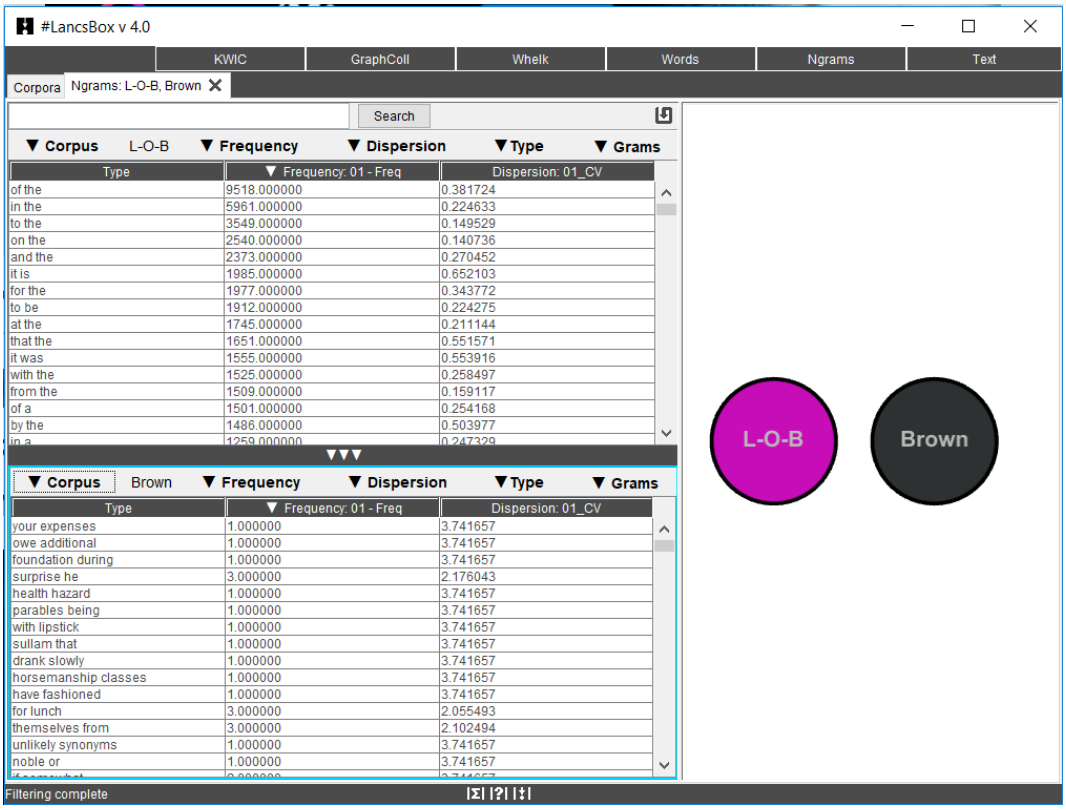
L'outil Ngram permet une analyse en profondeur des fréquences des « n-grammes » (bigrammes, trigrammes, etc.), qui peuvent être définis comme des combinaisons contiguës, ou à partir de lemmes et d'étiquettes morpho-syntaxiques.

L'outil produit également des n-grammes clés en comparant deux corpus, grâce à un calcul similaire à celui des mots-clés.

L'outil peut notamment être utilisé pour :

- identifier n-grammes, ensembles lexicaux et *p-frames* (ou skip-grammes)
- calculer des mesures de fréquence et de dispersion pour des n-grammes de types, de lemmes et d'étiquettes morpho-syntaxiques
- visualiser la fréquence et la dispersion des n-grammes dans les corpus
- comparer les n-grammes de deux corpus en utilisant le calcul des mots-clés
- visualiser les n-grammes clés.

8.1 Résumé visuel



Volet gauche : Création des listes de fréquence, calcul de la dispersion et des n-grammes clés.

Volet droit : Visualisation des fréquences, des dispersions et des n-grammes clés.

9 Texte

L'outil Texte vous donne des indications détaillées quant au contexte dans lequel on utilise un mot ou une phrase.

L'outil peut notamment être utilisé pour :

- voir un terme dans son contexte intégral
- prévisualiser un texte
- prévisualiser tous les textes d'un corpus dans un seul document
- voir les différents niveaux d'annotations d'un texte ou d'un corpus.

9.1 Résumé visuel

Le surlignage permet de repérer toutes les occurrences du terme recherché.
La navigation se fait avec les flèches verticales.