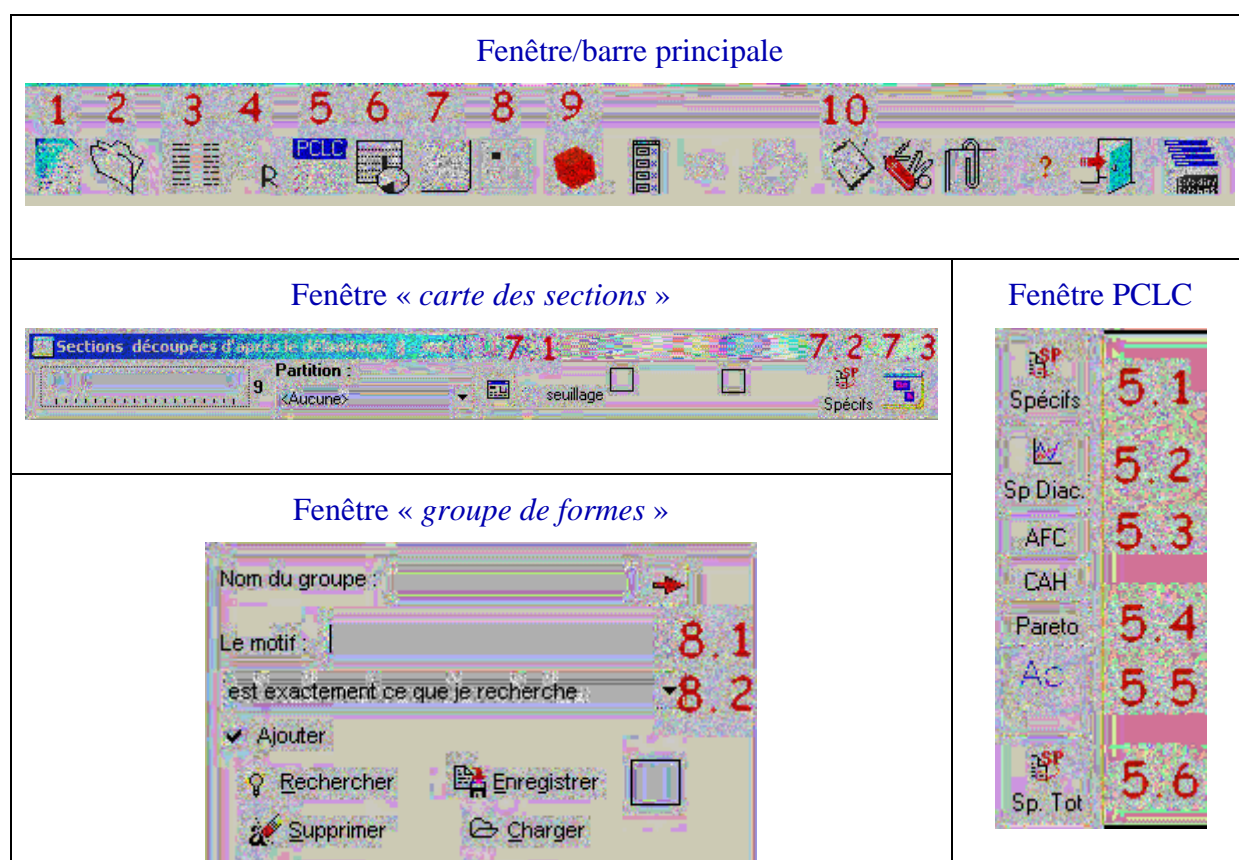











Fonctionnalités de Lexico3

Tableau des fonctionnalités

Pour présenter les fonctionnalités mises en œuvre dans les sections suivantes, nous avons réunis ci-dessous les différentes icônes associées aux fonctionnalités visées dans leur contexte d'utilisation :



Le tableau ci-contre rassemble, nomme et présente brièvement l'ensemble des fonctionnalités du logiciel *Lexico3* utilisées pour mener à bien l'exploration textométrique proposée dans les sections suivantes. On pourra aussi se reporter aux différents manuels du logiciel disponibles en ligne.

| <i>N°</i> | <i>Nom</i> | <i>Paramètres</i> | <i>Localisation</i> | <i>icône</i> |
|-----------|------------------------------------|--|--|---|
| 1 | SEGMENTATION | Liste de délimiteurs <i>Par défaut :</i> . , ; ! ? / _ - \ " ' () [] { } § \$ | Barre principale |  |
| 3 | CONCORDANCE | Forme (ou Type Généralisé) | Barre principale |  |
| 4 | SEGMENTS REPETES | | Barre principale |  |
| 5 | PCLC | Une fois la partition construite, on peut accéder au tableau présentant les Principales Caractéristiques lexicométriques de la partition. | Barre principale |  |
| 6 | PARTITION | Une clé définissant une partition dans le corpus original est du type : <CLE= « valeur »> C'est le nom de la clé qui est donné ici pour construire la partition visée | Barre principale |  |
| 6 | VENTILATION | Forme ou groupe de formes | | |
| 7 | CARTE DES SECTIONS | délimiteur de section | Barre principale |  |
| 8 | GROUPE DE FORMES | Cette fonctionnalité produit des listes de formes qu'il est possible de mémoriser, d'exporter ou de « projeter » sur les graphiques construits par Lexico3 . Elle permet surtout de faire des recherches de formes ou de groupes de formes en utilisant la notion d'expression régulière. | Barre principale |  |
| 5.3 | AFC | | Fenêtre des PCLC |  |
| 5.1, 7.2 | SPECIFICITES (POSITIVES NEGATIVES) | Partie ou section du corpus | Fenêtre des PCLC Carte des sections |  |

Lexico3, Tableau des Fonctionnalités

Glossaire

segmentation - opération qui consiste à délimiter des unités minimales dans un texte. Les **unités minimales** (pour un type de segmentation) - unités que l'on ne décompose pas en unités plus petites pouvant entrer dans leur composition (ex : dans la segmentation en formes graphiques les formes ne sont pas décomposées en fonction des caractères qui les composent)

caractères délimiteurs / non-délimiteurs : distinction opérée sur l'ensemble des caractères qui entrent dans la composition du texte, permettant aux procédures informatisées de segmenter le texte en occurrences (suite de caractères non-délimiteurs bornée à ses extrémités par des caractères délimiteurs).

On distingue parmi les caractères délimiteurs:

- les caractères délimiteurs d'occurrence (encore appelés "**délimiteurs de forme**") qui sont en général : le blanc, les signes de ponctuation usuels, les signes de préanalyse éventuellement contenus dans le texte.
- les caractères **délimiteurs de séquences** : sous-ensemble des délimiteurs d'occurrence correspondant, en général, aux ponctuations faibles et fortes contenues dans la police des caractères.
- les caractères **séparateurs de phrase** : (sous-ensemble des délimiteurs de séquence) qui correspondent, en général, aux seules ponctuations fortes.

forme ou "**forme graphique**" : archétype correspondant aux occurrences identiques dans un corpus de textes, c'est-à-dire aux occurrences composées strictement des mêmes caractères non-délimiteurs d'occurrence

partition (d'un corpus de textes) : division d'un corpus en *parties* constituées par des fragments de texte consécutifs, n'ayant pas d'intersection commune et dont la réunion est égale au corpus.

ventilation (des occurrences d'une unité dans les parties du corpus) : La suite des n nombres (n = nombre de parties du corpus) constituée par la succession des sous-fréquences de cette unité dans chacune des parties, prises dans l'ordre des parties

motif : un ensemble d'objets possédant une propriété reconnaissable.

analyse factorielle : famille de méthodes statistiques d'analyse multidimensionnelle, s'appliquant à des tableaux de nombres, qui visent à extraire des "facteurs" résumant approximativement par quelques séries de nombres l'ensemble des informations contenues dans le tableau de départ.

spécificité positive : pour un seuil de spécificité fixé, une forme i et une partie j données, la forme i est dite spécifique positive de la partie j (ou forme caractéristique* de cette partie) si sa sous-fréquence est "anormalement élevée" dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou supérieures à la sous-fréquence constatée est inférieure au seuil fixé au départ

spécificité négative : pour un seuil de spécificité fixé, une forme i et une partie j données, la forme i est dite spécifique négative de la partie j si sa sous-fréquence est anormalement faible dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou inférieures à la sous-fréquence constatée est inférieure au seuil fixé au départ

partie (d'un corpus de textes) : fragment de texte correspondant aux divisions naturelles de ce corpus ou à un regroupement de ces dernières.

section : portion de texte comprise entre deux délimiteurs de section (exemple : le paragraphe, etc.).

segment répété (ou polyforme répétée) : suite de forme dont la fréquence est supérieure ou égale à 2 dans le corpus.

Les expressions régulières avec Lexico3

Dans les sections qui suivent on utilisera à plusieurs reprises la notion d'expression régulière en particulier à travers la fonction «GROUPE DE FORMES». Nous rappelons ci-dessous brièvement cette notion et les différents opérateurs disponibles avec **Lexico3** pour écrire de telles expressions. Les expressions régulières permettent de représenter de manière générique des motifs textuels : un *motif* est un ensemble d'objets possédant une propriété reconnaissable, par exemple tous les mots terminés par le suffixe « able » ou commençant par le préfixe « pré ». Les expressions régulières permettent ainsi de décrire des portions de texte à l'aide d'opérateurs particuliers. Le tableau suivant rassemble l'ensemble des opérateurs disponibles avec **Lexico3** pour écrire des motifs sous la forme d'expression régulière :

| Opérateur | Fonction | Application |
|------------------|---|--|
| . (le point) | Représente n'importe quel caractère | L'expression "m.l" représente des séquences comme : mal, mol... |
| * | 0 ou n occurrences du caractère qui précède | L'expression "com*e" représente des séquences comme : coe, come, comme, comme... |
| + | 1 ou n occurrences du caractère qui précède | L'expression "com+e" représente des séquences comme : comme, comme,... |
| \b | Représente un début de mot | L'expression "\bcapital" représente des séquences comme : capital, capitale, capitalisme... |
| \b | Représente une fin de mot | L'expression ".*isme\b" représente des séquences comme : syndicalisme, capitalisme... |
| [] | Représente un ensemble de caractères | L'expression "[aeiou]" représente des séquences comme : un des caractères de l'ensemble des voyelles minuscules. L'expression "[a-z]" représente un des caractères minuscules compris entre a et z. |
| [^] | Représente la négation du contenu de l'ensemble de caractères | L'expression "[^aeiou]" représente un des caractères parmi ceux qui ne sont pas ceux de l'ensemble des voyelles minuscules |