

Oxford WordSmith Tools

Version 4.0

© 2004-2006 Mike Scott

Oxford WordSmith Tools

version 4.0

par Mike Scott

© 2004-2006 Mike Scott

WordSmith Tools

© 2004-2006 Mike Scott

Tous droits réservés. La reproduction de cette oeuvre, complète ou partielle, sous quelque forme que ce soit, graphique, électronique ou mécanique, y compris le photocopiage, l'enregistrement ou le stockage ne peut se faire sans l'accord écrit de l'éditeur.

Les produits mentionnés dans ce document peuvent être des marques commerciales et/ou des marques déposées appartenant à leurs propriétaires respectifs. L'éditeur et l'auteur ne revendiquent en aucun cas ces marques déposées.

Bien que toutes les précautions aient été prises lors de la préparation de ce document, l'éditeur et l'auteur ne peuvent être tenus responsables en cas d'erreurs ou d'omissions, ni pour les dommages résultant de l'utilisation des informations contenues dans ce document ni de l'utilisation des programmes et du code source l'accompagnant. En aucun cas l'éditeur et l'auteur ne peuvent être tenus responsables pour la perte de bénéfice, ou tout autre dommage commercial, causé ou déclaré comme ayant été causé directement ou indirectement par ce document.

Imprimé en mars 2006.

Éditeur

Oxford University Press

Mes reconnaissances en particulier à:

Audrey Spina et Élodie Guthmann qui l'ont traduit.

Toutes les personnes ayant contribué à la création de ce document en testant WordSmith Tools dans ses diverses versions. Merci en particulier à ceux qui ont signalé des problèmes et m'ont envoyé leurs suggestions.

Sommaire

	0
Partie I WordSmith Tools	2
Partie II Présentation générale	4
1 Nouveautés de la version 4	4
2 Contrôleur	4
3 Concord	4
4 KeyWords	5
5 WordList	5
6 Utilitaires	5
Sélection des langues	5
File Utilities	6
Minimal Pairs	6
Splitter	6
Text Converter	6
Version Checker	7
Viewer	8
Webgetter	9
Partie III Premiers pas	11
1 premiers pas avec Concord	11
2 premiers pas avec KeyWords	12
3 premiers pas avec WordList	13
Partie IV Installation et mise à jour	15
1 installation de WordSmith Tools	15
2 paramètres réseau par défaut	15
3 vérification de la version	16
Partie V Contrôleur	18
1 modifier les paramètres	18
2 accents	18
3 ajouter des notes	18
4 traitement par lots	18
5 dossiers contenant les lots	21
6 modifier la police	21
7 sélectionner les textes favoris	22
8 sélectionner la langue	23
9 sélectionner les textes	23

10	sélection des fichiers à partir d'une boîte de dialogue standard	26
11	couleurs	26
12	totaux des colonnes	28
13	calculer une nouvelle colonne de données	29
14	calculer la fréquence des données	30
15	copier vos résultats	32
16	traitement personnalisé	33
17	personnalisation des menus contextuels	36
18	édition d'une liste de données	38
19	chercher des fichiers pertinents	39
20	instructions de cours ou de session	40
21	mise en page et format	40
22	faire correspondre des mots dans une liste	42
23	imprimer et aperçu avant impression	44
24	première utilisation de WordSmith	45
25	listes précédentes	45
26	quitter WordSmith	45
27	réduire les données à n entrées	46
28	enregistrer au format texte	46
29	enregistrer les valeurs par défaut	48
30	enregistrer les résultats	49
31	rechercher et remplacer	50
32	recherche par saisie	50
33	rechercher un mot ou un fragment de mot	50
34	voir les noms de fichiers	51
35	listes d'arrêt	51
36	arrêter le traitement	52
37	texte et langues	53
38	gestion de la fenêtre	55
39	ignorer les lignes indésirables	56

Partie VI Concord 58

1	but	58
2	qu'est-ce qu'une concordance ?	58
3	masquage	58
4	catégories	59
5	voisinage des co-occurents	60
6	paramètres des co-occurents	60
7	affichage des co-occurents	61
8	relation de collocation	63

9	collocation	63
10	Concord : clusters	64
11	Concord : dispersion	66
12	Concord : enregistrement et impression	67
13	Concord : affichage des options	68
14	Concord : gestion des fichiers son et vidéo	68
15	Concord : ce que vous pouvez voir et faire	68
16	paramètres de concordance	70
17	effectuer une concordance sur des balises	73
18	terme de contexte	74
19	modifier des concordances	75
20	termes de recherche issus du fichier	75
21	balise la plus proche	76
22	schémas	79
23	re-trier :	79
24	re-trier : co-occurents	80
25	re-trier : schéma de dispersion	80
26	segments de texte dans Concord	80
27	syntaxe du terme de recherche	81
28	contrôleur WordSmith : Concord : paramètres	82

Partie VII KeyWords 86

1	but	86
2	Analyse de deux listes de mots	86
3	définition des associés	86
4	associés	87
5	sélection des fichiers	87
6	groupes	87
7	Clusters KeyWords	88
8	concordance	89
9	Création d'une base de données	89
10	exemple de mots-clés	90
11	définition des mots-clés clés	90
12	définition de pertinence	90
13	base de données KeyWords	91
14	KeyWords : conseils	91
15	KeyWords : calcul	92
16	KeyWords : liens	92
17	créer une liste de mots à partir de données de mots-clés	93
18	p valeur	93

19 calcul des schémas	93
20 affichage des schémas	94
21 effectuer un regroupement de synonymes	95
22 re-trier : KeyWords	95
23 l'écran des mots-clés	96
24 contrôleur WordSmith : Paramètres de KeyWords	97
Partie VIII WordList	100
1 but	100
2 regroupement automatique de lemmes	100
3 sélection d'un fichier de lemmes	101
4 comparaison de listes de mots	102
5 fusion des listes de mots	103
6 affichage de comparaison	103
7 analyse de cohérence (détaillée)	105
8 analyse de cohérence (simple)	106
9 lemmes	107
10 listes d'index : utilisation	108
11 listes d'index : affichage	109
12 création d'un index WordList	110
13 clusters d'index	111
14 recherche par menu	114
15 scores d'informations mutuelles	115
16 informations mutuelles : calcul	117
17 Affichage des informations mutuelles	117
18 re-trier : listes de cohérence	119
19 statistiques	120
20 Importer des mots à partir d'une liste au format texte	120
21 ratios type/token	122
22 sensibilité à la casse	123
23 paramètres minimum et maximum	124
24 tri	124
25 Affichage WordList	125
26 contrôleur WordSmith : Paramètres de WordList	127
Partie IX Viewer et Aligner	130
1 but	130
2 alignement avec Viewer	130
3 alignement et déplacement	130
4 édition	131

5	numérotation des phrases et des paragraphes	131
6	options	131
7	regroupement et dissociation des phrases	132
8	paramètres	132
9	aspects techniques	133
10	non correspondances de traduction	133
11	résolution des problèmes	134
12	phrases inhabituelles	134

Partie X Utilitaires 136

1	Convertir les données provenant de versions antérieures	136
	Convertir les données provenant de versions antérieures	136
2	WebGetter	136
	présentation générale	136
	paramètres	137
	affichage	138
	limites	139
3	Sélecteur de langue	140
	Présentation générale	140
	Langue	141
	Police	142
	Tri	142
	Autres langues	143
	Enregistrement de vos préférences	144
4	Minimal Pairs	144
	fonction	144
	critères	144
	sélectionner vos fichiers	145
	affichage	145
	règles et paramètres	145
	exécution du programme	146
5	File Utilities	147
	index	147
	Splitter	147
	fonction de Splitter.....	147
	Splitter : noms de fichiers.....	148
	Splitter : caractères de remplacement.....	148
	regrouper des fichiers texte	149
	comparer deux fichiers	150
	découpeur de fichiers	151
	rechercher des doublons	151
	renommer	152
6	Text Converter	152
	but	152
	Text Converter : paramètres	153
	Text Converter : syntaxe	155
	Text Converter : déplacer si...	156
	Fichier de conversion Text Converter	157
	Text Converter : exemple de fichier de conversion	158

Partie XI Balises et notes	160
1 présentation générale	160
2 types de balises	160
3 gestion des balises	161
4 balises multimédia	162
5 balises utilisées comme sélecteurs	163
6 uniquement si les éléments contenus sont... ..	164
7 sélection au sein de textes	165
8 création d'un fichier de balises	166
9 début et fin des segments de texte	168
Partie XII Référence	170
1 file types	170
2 international versions	170
3 version 32 bits	171
4 remerciements	171
5 bibliographie	172
6 bugs	173
7 Jeux de caractères	174
présentation générale	174
accents et symboles	174
ansi et ascii	175
DOS	175
Windows	175
Unicode	176
8 presse-papiers	176
9 adresses de contact	178
10 format de date	179
11 Définitions	179
définitions	179
séparateurs de mots	180
12 version de démonstration	180
13 mode édition et mode saisie	181
14 rechercher des textes source	181
15 folders\;directories	182
16 HistoryList	182
17 HTML, SGML et XML	182
18 traits d'union	183
19 Internet	183
20 limites	183
21 limites spécifiques des outils	183

22	liens entre les outils	184
23	liste des options des menus et des boutons	185
24	raccourcis clavier	188
25	noms de fichiers longs	189
26	configuration minimale requise	189
27	manuel de WordSmith Tools	189
28	nombres	189
29	valeur de dispersion du schéma	189
30	paramètres d'impression	190
31	mémoire RAM disponible	191
32	corpus de référence	191
33	restaurer le dernier fichier	192
34	mots uniques/clusters	192
35	vitesse	193
36	barre d'état	193
37	outils pour le repérage de schémas	193
38	informations sur la version	194
39	fichiers zip	195

Partie XIII Résolution des problèmes 198

1	liste des FAQ	198
2	apostrophes introuvables	198
3	espace entre les colonnes	198
4	problème de balises Concord	198
5	Incohérence entre Concord et WordList	199
6	en panne	199
7	limite de la version de démonstration	199
8	symboles inhabituels	199
9	couleurs illisibles	200
10	les touches ne répondent pas	200
11	pineapple-slicing	200
12	l'imprimante n'a pas imprimé	200
13	trop lent	201
14	ne démarre pas	201
15	la liste de mots ne fonctionne pas	201

Partie XIV Messages d'erreur 203

1	accès au fichier refusé	203
2	associés introuvables	203
3	aucun co-occurent trouvé	203

4	aucun élément activé	203
5	aucun groupe identifié	203
6	aucun mot-clé à schématiser	203
7	aucun mot-clé trouvé	203
8	aucun terme de liste d'arrêt	204
9	aucun terme de liste d'arrêt de concordance	204
10	aucun terme de liste d'arrêt KeyWords	204
11	aucun terme de liste de balises	204
12	aucun terme de liste de correspondance	204
13	aucun terme de liste de lemmes	204
14	aucune entrée de concordance	204
15	aucune entrée présente dans la base de données KeyWords	205
16	aucune ligne supprimée à ignorer	205
17	aucune liste de mots sélectionnée	205
18	calcul KeyWords non valide	205
19	ce programme nécessite Windows 98 ou une version supérieure	205
20	c'est une version de démonstration	205
21	clusters introuvables	205
22	détails de la liste de correspondance non spécifiés	205
23	deux fichiers sont nécessaires	206
24	dossier cible introuvable	206
25	échec lors de la lecture du fichier	206
26	échec lors de l'enregistrement du fichier	206
27	erreur de liste de base	206
28	espace insuffisant pour la variable calculée	206
29	fichier *.ini introuvable	206
30	fichier base de données KeyWords non valide	207
31	fichier de balises introuvable	207
32	fichier de balises non lu	207
33	fichier de comparaison WordList non valide	207
34	fichier de concordance non valide	207
35	fichier de conversion introuvable	207
36	fichier de liste d'arrêt de concordance introuvable	208
37	fichier de liste d'arrêt introuvable	208
38	fichier de liste d'arrêt non lu	208
39	fichier de liste de mots introuvable	208
40	fichier introuvable	208
41	fichier spécifié introuvable	208
42	fichier texte original nécessaire mais introuvable	208

43	fichier WordList non valide	209
44	fichier(s) source introuvable(s)	209
45	format du code d'enregistrement erroné	209
46	formulaire incomplet	209
47	il doit y avoir un nombre	209
48	il n'est possible d'enregistrer les mots qu'au format ASCII	209
49	impossible d'accéder à Internet	209
50	impossible d'accéder au fichier texte - fichier utilisé ailleurs ?	210
51	impossible d'appeler un autre outil	210
52	impossible de calculer les mots-clés si les langues diffèrent	210
53	impossible de créer le dossier car un autre dossier porte le même nom	210
54	impossible de créer un nouveau nom de dossier	210
55	impossible de fusionner la liste avec elle-même	210
56	impossible de fusionner les bases de données KeyWords	210
57	impossible de lire le fichier	211
58	imprimante nécessaire	211
59	informations mutuelles incompatibles	211
60	inscription non valide	211
61	inscription réseau utilisée ailleurs	211
62	jeu de caractères réinitialisé sur <x> pour convenir à : <language>	212
63	la fonction n'est pas encore disponible	212
64	la fonction n'est pas encore prête	212
65	le fichier comporte des "trous"	212
66	le fichier de base de données KeyWords est défectueux	212
67	le fichier de comparaison WordList est défectueux	213
68	le fichier de concordance n'est pas valide	213
69	le fichier de liste de mots est défectueux	213
70	le fichier KeyWords est défectueux	213
71	le fichier ne contient aucune des balises spécifiées	213
72	lecteur non valide	213
73	les dispersions vont de paire avec les concordances	214
74	les liens entre les outils sont défailants	214
75	les noms de fichiers doivent être différents	214
76	limite de regroupement atteinte	214
77	limite de termes de recherche des fichiers atteinte	214
78	liste des messages d'erreur	214
79	mémoire insuffisante	216
80	messages de confirmation : acceptez-vous une relecture ?	216
81	nom de fichier non valide	216

82	nombre de phrases trop important	216
83	nombre non valide	216
84	noms complets du lecteur et du dossier nécessaires	217
85	Non correspondance de version de WordSmith	217
86	pas un fichier WordSmith	217
87	pas un fichier WordSmith actuel	217
88	pour que ce message ne s'affiche plus...	217
89	pourquoi la recherche a-t-elle échoué ?	217
90	problème de disque – le fichier n'a pas été enregistré	218
91	statistiques non disponibles	218
92	troncature à xx mots – le fichier de liste de balises en comporte davantage	218
93	trop de requêtes pour ignorer les groupes correspondants	218
94	utilisation sur l'ordinateur X uniquement	218
95	WordSmith Tools a expiré	218
96	WordSmith Tools est déjà en cours d'utilisation	218
97	XX jours restants	218
	Index	219

WordSmith Tools

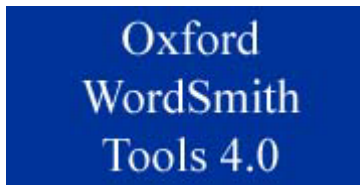
WordSmith Tools

Section



I

1 WordSmith Tools




Oxford WordSmith Tools est une suite intégrée de programmes permettant d'étudier le "comportement" des mots dans les textes. Ces outils vous permettront de voir comment sont utilisés les mots dans vos propres textes ou dans ceux d'autres personnes.

L'outil **WordList** vous permet d'afficher une liste de tous les mots ou clusters de mots d'un texte, classés par ordre alphabétique ou de fréquence. Le concordancier, **Concord**, vous permet d'afficher un mot ou un syntagme en contexte afin de voir quels sont les autres mots qui l'entourent. **KeyWords** vous permet de rechercher les mots-clés d'un texte.

Les outils sont utilisés par Oxford University Press pour leur propre travail lexicographique (préparation de dictionnaires), par les professeurs de langues et leurs étudiants et par les chercheurs qui étudient les schémas linguistiques dans de nombreuses langues différentes.

Obtenir de l'aide

[Si vous disposez d'une connexion Internet, cliquez ici pour afficher un guide pas-à-pas avec des captures d'écran montrant les fonctionnalités de WordSmith.](#)

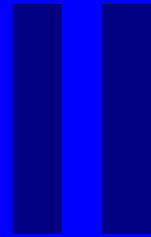
La plupart des menus et boîtes de dialogue comportent des options d'aide. La plupart du temps, il suffit d'appuyer sur F1 ou de cliquer sur , pour obtenir de l'aide. Vous pouvez aussi cliquer sur Aide (option située dans la partie droite de la plupart des menus). Au sein d'un fichier d'aide (comme celui-ci), il peut s'avérer plus facile de cliquer sur le bouton Rechercher et de regarder l'index affiché ou de parcourir les écrans d'aide.

Cliquez ici pour commencer à utiliser [WordList](#), [Concord](#), ou [KeyWords](#).

WordSmith Tools

Présentation générale

Section



2 Présentation générale

2.1 Nouveautés de la version 4

La version 4 présente un programme entièrement réécrit.

Nouvelles fonctionnalités :

- [guide détaillé en ligne \(avec captures d'écran\) expliquant comment créer des listes de mots, des concordances, etc.](#)
- [version en ligne complète de ce fichier d'aide](#)
- nombre [quasi illimité](#) de concordances et listes de mots
- gestion améliorée des [clusters](#) de listes de mots
- [Clusters](#) Concord de tailles variables
- Concord montre où apparaît chaque occurrence dans les [phrases, paragraphes, en-têtes et sections](#)
- Gestion du format [Unicode](#) permettant ainsi le traitement de nombreuses [langues](#)
- possibilité d'effectuer une concordance et de jouer le [fichier son](#) correspondant
- [gestion des balises](#) améliorée
- création de votre propre [corpus de textes à partir d'Internet](#)
- fonctions statistiques avancées pour les [collocations](#)
- lemmatisation avancée en utilisant vos [fichiers *.dll personnalisés](#)
- utilisation de [fichiers *.zip](#)
- plus d'utilitaires
- vérification de la [version](#)
- plus de langues dans [Aligner](#)

2.2 Contrôleur



Ce programme accède aux différents outils. Il permet d'afficher et de modifier les paramètres actuels, gère la sélection des fichiers texte, et appelle les différents outils.

Le contrôleur apparaît en haut à gauche de votre écran.

Si pouvez le réduire si vous estimez que votre écran est trop [encombré](#).

Pour consulter le guide détaillé pas à pas avec des captures d'écran, cliquez [ici](#) pour visiter le site de WordSmith.

2.3 Concord



Concord est un programme effectuant une [concordance](#) en utilisant des fichiers texte [DOS](#), [Texte brut](#), [ASCII](#) ou [ANSI](#).

Pour l'utiliser, il faut spécifier un terme de recherche, que Concord doit rechercher dans tous les fichiers texte que vous avez sélectionnés. Le programme affiche ensuite les résultats de

concordance et vous donne accès aux informations concernant les co-occurents du terme de recherche.

Les listes peuvent être [enregistrées](#) pour une réutilisation ultérieure telle que l'édition, l'impression ou la copie dans un traitement de texte. Il est également possible de les enregistrer comme fichiers texte.

Rubriques connexes : Sommaire de l'aide Concord, [Les boutons](#)

2.4 KeyWords



Ce programme permet de localiser et d'identifier les mots-clés dans un texte donné. Il compare les mots du texte avec un ensemble de mots de référence généralement tirés d'un grand corpus de textes. Tout mot dont la fréquence dans le texte est notoire est considéré comme étant un mot-clé. Les mots-clés sont affichés par ordre de pertinence.

La répartition des mots-clés peut être [schématisée](#).

Les listes peuvent être [enregistrées](#) pour une réutilisation ultérieure telle que l'édition, l'impression ou la copie dans un traitement de texte. Il est également possible de les enregistrer comme fichiers texte.

Ce programme doit avoir accès à au moins deux listes de mots, lesquelles doivent d'abord être créées avec le programme [Word List](#).

Rubriques connexes : Sommaire de l'aide KeyWords, [Les boutons](#)

2.5 WordList



Ce programme génère des listes de mots à partir d'un ou plusieurs fichiers texte [ANSI](#) ou [ASCII](#). Les listes de mots peuvent être affichées par ordre alphabétique ou par ordre de fréquence. Vous pouvez les [enregistrer](#), éditer, imprimer, copier dans votre logiciel de traitement de texte ou les enregistrer en tant que fichiers texte.

Rubriques connexes : Sommaire de l'aide WordList, [Les boutons](#)

2.6 Utilitaires

2.6.1 Sélection des langues



C'est un outil permettant de sélectionner les langues à utiliser.

En règle générale, cette sélection n'est effectuée qu'une fois, lors de votre première utilisation de WordSmith Tools.

Rubriques connexes : [Outil de sélection des langues](#)

2.6.2 File Utilities



Programmes pour

- [comparer deux fichiers](#)
- [découper de gros fichiers en plusieurs parties](#)
- [rechercher des doublons](#)
- [renommer](#) plusieurs fichiers
- rechercher des "[trous](#)" présents dans des fichiers texte
- scinder de gros fichiers en leurs composants textuels
- [joindre](#) de nombreux petits fichiers texte à des fichiers texte fusionnés

2.6.3 Minimal Pairs



Ce programme sert à rechercher les fautes de frappe et les paires de mots minimales.

Rubriques connexes : [fonction](#), [critères](#), [sélectionner vos fichiers](#), [affichage](#), [règles et paramètres](#), [exécution du programme](#).

2.6.4 Splitter



Splitter est un utilitaire qui scinde de gros fichiers en de plus petits fichiers pour l'analyse de texte. Vous pouvez spécifier un symbole pour représenter la fin d'un texte (par exemple </**Text**>) et Splitter parcourt un gros fichier en copiant le texte vers un nouveau fichier texte chaque fois qu'il trouve le symbole.

Rubriques connexes : Sommaire de l'aide de Splitter, [Les boutons](#)

2.6.5 Text Converter



Text Converter est un utilitaire général permettant d'effectuer les trois tâches principales suivantes : éditer des textes, renommer des fichiers texte, modifier les attributs des fichiers et déplacer des fichiers dans un nouveau dossier s'ils contiennent certains mots ou syntagmes. La fonction la plus utilisée est le remplacement de chaînes dans les fichiers texte. Il permet d'effectuer un "rechercher et remplacer" comme le font les logiciels de traitement de texte, mais il effectue cette opération dans un grand nombre de fichiers, les uns après les autres. Au cours

de cette opération, il peut également remplacer n'importe quel nombre de chaînes, et pas seulement une.

Il est très utile pour parcourir un grand nombre de fichier et les reformater comme vous le souhaitez, comme par exemple supprimer les espaces superflus, s'assurer que seuls les paragraphes comportent <Entrée> à la fin, ou encore modifier les caractères accentués.

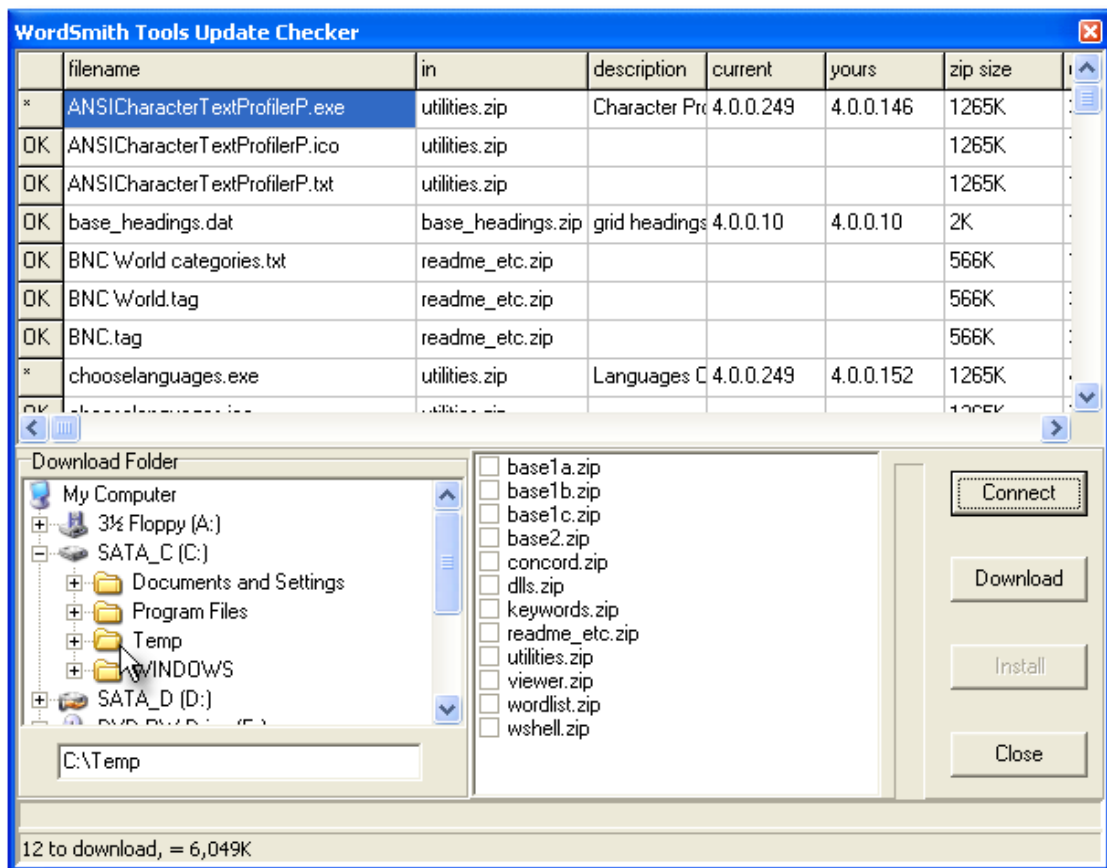
Rubriques connexes : Sommaire de l'aide de Text Converter, [Options des menus et des boutons](#)

2.6.6 Version Checker

4.0

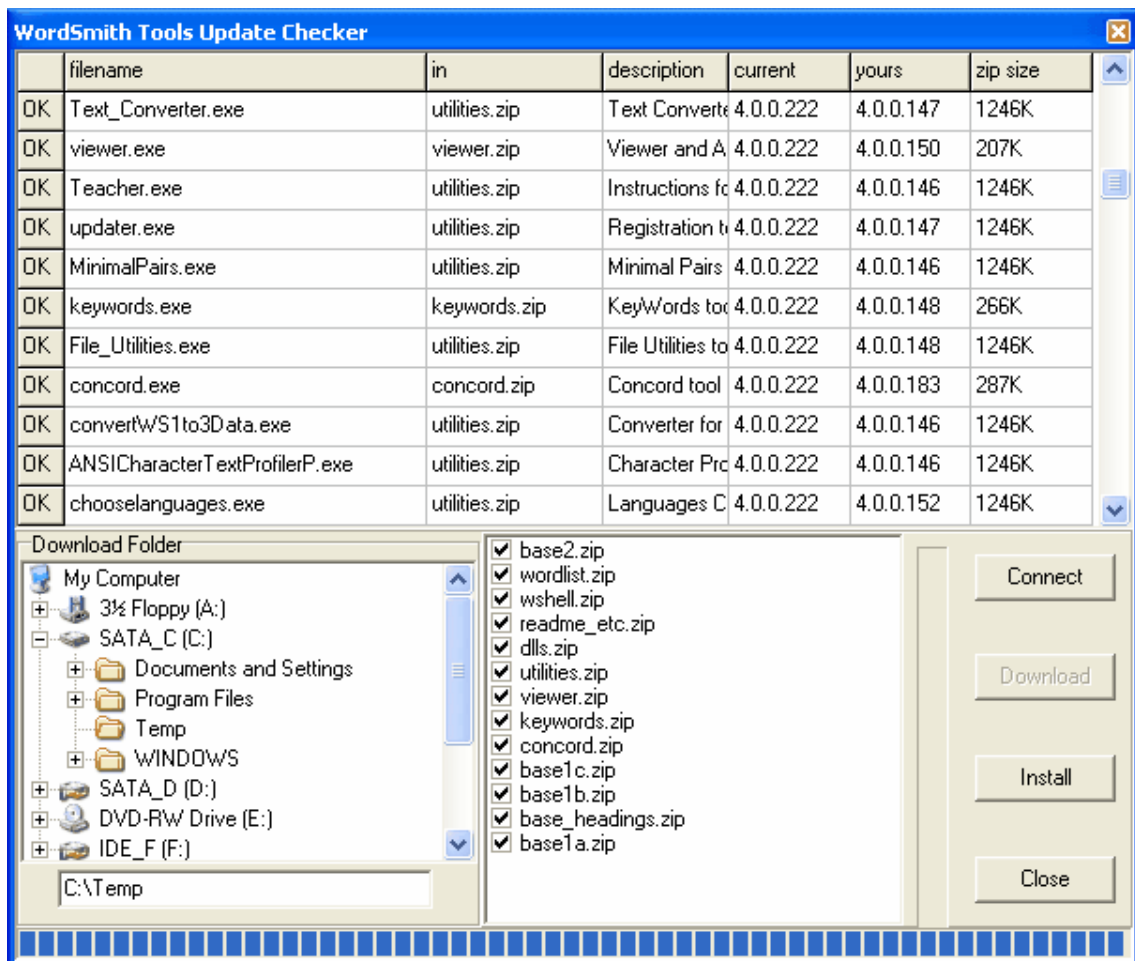
C'est un outil permettant de vérifier si l'un des composants de votre version actuelle a besoin d'une mise à jour. Si c'est le cas, le programme télécharge le composant nécessaire. Vous pouvez y accéder par le menu du contrôleur principal, en cliquant sur *Fichier | Vérification de la version via Internet*.

Une fois le programme lancé, après avoir cliqué sur *Connecter*, vous verrez ceci s'afficher :



Les différents composants de WordSmith sont listés dans la fenêtre supérieure, et la version actuelle est comparée avec celle que vous avez. S'il y a une différence, tous les fichiers du fichier *.zip correspondant seront marqués d'un astérisque dans la marge de gauche.

Par défaut, le téléchargement s'effectue vers l'emplacement où WordSmith est déjà installé, mais vous êtes libre de modifier cela. Par exemple, sur la capture d'écran, c'est le dossier `c:\temp` qui a été spécifié. Cliquez sur *Télécharger* si vous souhaitez récupérer les fichiers mis à jour.



Une fois le téléchargement effectué, les différents fichiers *.zip sont cochés (en bas à droite de la fenêtre) si le téléchargement est réussi, et le bouton Installer apparaît. Cliquer sur ce bouton permet de décompresser tous les fichiers cochés.

2.6.7 Viewer



Viewer est un utilitaire qui vous permet d'analyser vos fichiers en différents formats. Cet outil peut être appelé à partir d'autres outils lorsque vous souhaitez afficher le texte source.

Viewer peut également être utilisé pour copier un fichier texte avec des [phrases ou paragraphes numérotés](#) ou pour [aligner](#) deux versions d'un texte et afficher alternativement les phrases et paragraphes de chacun.

Rubriques connexes : Sommaire de l'aide Viewer, [Les boutons](#)

2.6.8 Webgetter



Outil permettant de récupérer des textes à partir d'Internet.

Le but

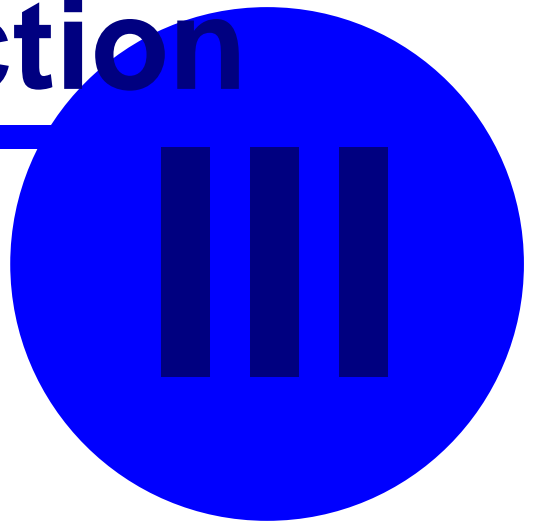
Télécharger des pages web à l'aide d'un moteur de recherche vous permet de construire votre propre corpus de textes.

Rubriques connexes : [Explications plus détaillées](#), [Paramètres](#), [Affichage](#), [Limites](#)

WordSmith Tools

Premiers pas

Section



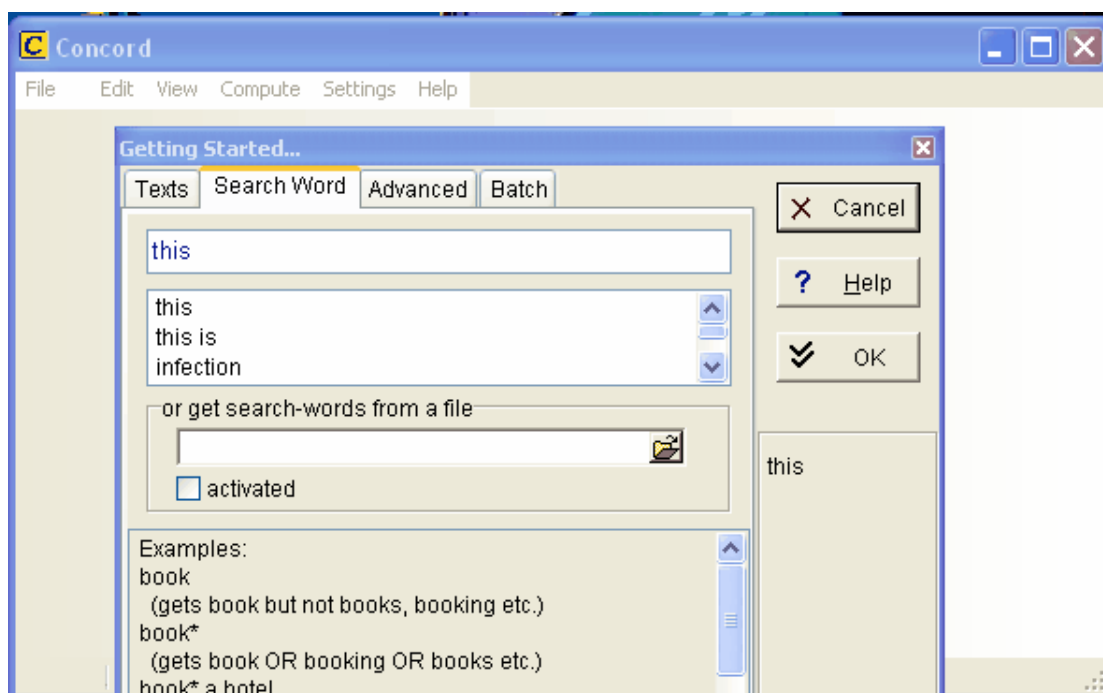
3 Premiers pas

3.1 premiers pas avec Concord

Pour consulter le guide détaillé pas à pas avec des captures d'écran, cliquez [ici pour visiter le site de WordSmith](#).

Dans la fenêtre principale Oxford WordSmith Tools (celle comportant le [contrôleur](#) Oxford WordSmith Tools dans sa barre de titre), sélectionnez l'option Outils, et une fois ouvert, le bouton Concord apparaîtra. Cliquez pour que Concord s'ouvre.

À présent, une boîte de dialogue s'affiche, vous permettant de [sélectionner vos textes](#) ou de modifier votre choix, et d'effectuer une nouvelle concordance. Ce qui s'affiche ressemble globalement à ceci :

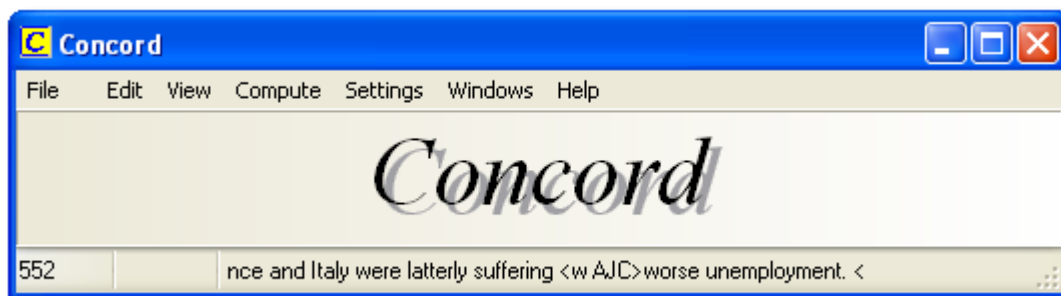


Si vous ne voyez que la fenêtre comportant Concord dans sa barre de titre, cliquez sur Fichier | Nouveau (🟢) et la fenêtre de démarrage s'affichera.

Si vous n'avez [jamais utilisé WordSmith](#) auparavant, vous verrez qu'un texte a été automatiquement sélectionné pour vous aider à démarrer avec le programme.


Vous devez ensuite spécifier un [terme ou un syntagme de recherche](#) puis cliquer sur OK (🟢).

Pendant que Concord effectue la tâche demandée, vous pouvez voir un indicateur d'avancement tel que celui ci :



Jusqu'ici, il y a 552 entrées, et la dernière montre le contexte d'utilisation du mot **worse**, le terme de recherche.

Si vous souhaitez modifier d'autres paramètres, cliquez sur [Avancés](#), mais vous pouvez sans doute conserver les paramètres par défaut. Concord parcourt à présent vos textes et recherche la [balise](#) ou le terme demandé.

N'oubliez pas d'[enregistrer les résultats](#) (appuyez sur F2 ou sur ) si vous souhaitez conserver la concordance pour une réutilisation ultérieure.

Rubriques connexes : Sommaire de l'aide de Concord.

3.2 premiers pas avec KeyWords

Pour consulter le guide détaillé pas à pas avec des captures d'écran, cliquez [ici pour visiter le site de WordSmith](#).

Dans la fenêtre principale Oxford WordSmith Tools (celle comportant le [contrôleur](#) Oxford WordSmith Tools dans la barre de titre), sélectionnez l'option Outils, et une fois ouverte, KeyWords apparaîtra. Cliquez pour que KeyWords s'ouvre.

Démarrage

Une boîte de dialogue vous permet de [sélectionner vos listes de mots](#). Il vous faudra sélectionner deux listes de mots à partir desquelles sera constituée la liste de mots-clés : la première est basée sur un texte unique et l'autre est basée sur plusieurs textes qui doivent être présents en quantité suffisante pour constituer un bon corpus de référence. Vous verrez deux listes des fichiers de listes de mots dans le dossier actif. Si aucune liste ne s'affiche, retournez dans WordList et créez quelques listes de mots. Sélectionnez une petite liste de mots au-dessus ainsi qu'une liste de [corpus de référence](#) avec lequel la comparer. Lorsque vos textes sont sélectionnés, vous pouvez procéder à une analyse de mots-clés. Cliquez sur *créer une liste de mots-clés maintenant*.

À ce moment-là, KeyWords démarre le traitement de votre fichier et une fenêtre de [progression](#) s'affiche dans le contrôleur principal. Une fois que KeyWords a terminé, une liste de mots-clés s'affichera. Ceux situés dans la partie supérieure sont des mots-clés plus pertinents que ceux situés en bas.

N'oubliez pas d'[enregistrer les résultats](#) (appuyez sur F2) si vous souhaitez conserver la liste de mots-clés pour une réutilisation ultérieure.

Rubriques connexes : Sommaire de l'aide de KeyWords, [À quoi sert-il ?](#)

3.3 premiers pas avec WordList

Pour consulter le guide détaillé pas à pas avec des captures d'écran, cliquez [ici pour visiter le site de WordSmith](#).

Le mieux est de commencer par WordList. Dans la fenêtre principale de la suite Oxford WordSmith Tools (celle comportant "[contrôleur](#) Oxford WordSmith Tools" dans sa barre de titre), cliquez sur l'option Outils, puis le bouton WordList apparaîtra. Cliquez pour que WordList s'ouvre.


Démarrage

Une boîte de dialogue s'affiche. Elle permet de [sélectionner vos textes](#) ou de modifier votre sélection, et de créer une nouvelle liste de mots.

Si vous n'avez [jamais utilisé WordSmith](#) auparavant, vous verrez qu'un texte a été automatiquement sélectionné pour vous aider à démarrer avec le programme.

D'autres paramètres peuvent être modifiés dans le menu, mais ce n'est pas vraiment nécessaire pour la création d'une nouvelle liste de mots, qu'elle soit unique ou sous forme de [lot](#).

À ce moment-là, WordList démarre le traitement de votre fichier et une fenêtre de [progression](#) s'affiche dans le contrôleur principal. Une fois la création de la liste terminée, trois fenêtres apparaissent et affichent les mots de votre fichier texte par ordre alphabétique et par ordre de fréquence ainsi que des statistiques.

N'oubliez pas d'[enregistrer les résultats](#) (appuyez sur F2 ou sur ) si vous souhaitez conserver la liste de mots pour une réutilisation ultérieure.

Rubriques connexes : Sommaire de l'aide de WordList.

WordSmith Tools

Installation et mise à jour

Section

IV

4 Installation et mise à jour

4.1 installation de WordSmith Tools

20 Mo d'espace disque disponible sont nécessaires à l'exécution des programmes. Le double est nécessaire durant l'installation. De toute manière, Windows ne peut pas fonctionner correctement sans qu'il y ait *au moins* 20 Mo d'espace disponible.

1. Vous avez reçu ou téléchargé un ou plusieurs fichiers *.exe. Placez-les dans un dossier temporaire, par exemple `c:\temp`. Le processus est plus facile si le dossier ne comporte aucun fichier.
2. Exécutez ces fichiers. Les fichiers nécessaires à **Oxford WordSmith Tools** sont ensuite extraits dans le dossier de votre choix (`c:\wsmith4` par défaut).
3. Exécutez `c:\wsmith4\wshell.exe` pour démarrer. Une boîte de dialogue s'affichera, vous demandant d'effectuer une mise à niveau. Si vous disposez d'une clé d'enregistrement, vous pouvez la saisir dans l'espace prévu à cet effet, sinon **WordSmith** continuera en [version de démonstration](#).
4. Si vous ne disposez que de peu d'espace disque, vous pouvez tout à fait supprimer les fichiers *.exe situés dans le dossier `c:\temp`.

Vous pouvez obtenir une version plus récente sur la [page d'accueil de WordSmith](#).

Pour désinstaller le programme, supprimez tous les fichiers du dossier `\wsmith4`.

Rubriques connexes : [Définition des options par défaut](#), [Adresses de contact](#).

4.2 paramètres réseau par défaut

Si vous avez acheté une licence de site, il est plus aisé d'installer une copie de WordSmith sur un serveur accessible à tous les utilisateurs. Il est bien sûr souhaitable de protéger cet emplacement principal et d'empêcher que des résultats y soient enregistrés ou que la copie originale de WordSmith ne soit modifiée. Vous pouvez alors ouvrir le fichier `wshell.ini` : vous verrez qu'il comporte une section vous permettant de spécifier précisément l'emplacement sur lequel chaque utilisateur doit enregistrer ses préférences.

Vous verrez les termes suivants :

```
lecteurs interdits
dossier limité
dossier d'instructions
dossier écriture/lecture-réseau
```

par exemple :

```
[NETWORK]
network-read/write folder=m:\wsmith4
```

Le lecteur M: sera utilisé sur le réseau car tout utilisateur peut y écrire des données.

```
prohibited drives=xyz
```

(X: Y: et Z: sont les lecteurs auxquels vous ne voulez pas que les utilisateurs aient accès lorsqu'ils sélectionnent les fichiers.

```
limited folder=v:\texts
```

V:\TEXTS (ainsi que tout sous-répertoire) est l'emplacement par défaut où les utilisateurs sélectionnent les corpus sur le réseau. Ils peuvent bien entendu rechercher à d'autres emplacements auxquels ils ont accès.

```
instructions folder=L:\English\WSmith instructions
```

Lorsque vous utilisez le logiciel lors d'un cours, vous pourrez placer les instructions dans ce dossier.

Lorsqu'une personne utilise WordSmith pour la première fois, le logiciel détectera qu'il s'agit d'une version réseau et lira les informations du dossier écriture/lecture-réseau ci-dessus. Il tentera ensuite de créer automatiquement le dossier spécifié ci-dessus (en théorie, vous ne devriez pas avoir à le faire vous-même) et de copier les fichiers *.ini ainsi que les autres fichiers de paramètres à partir du dossier de votre serveur où se trouve WordSmith vers le dossier automatiquement créé. Prenez garde à ce que le lecteur et le dossier que vous spécifiez soient véritablement ceux auxquels les utilisateurs peuvent accéder.

Rubriques connexes : [Instructions de cours](#)

4.3 vérification de la version

WordSmith est livré avec un fichier nommé `wordsmith_version_check.exe` qui vous permet de vérifier si votre version est à jour et, dans le cas contraire, de télécharger les mises à jour et patches nécessaires. Pour pouvoir les installer, WordSmith doit être fermé.

Rubriques connexes : [informations sur la version](#), [mise à jour de la version](#).

WordSmith Tools

Contrôleur

Section

V

5 Contrôleur

5.1 modifier les paramètres

La fenêtre principale des paramètres est accessible via le [contrôleur](#). Pour y accéder, cliquez sur *Paramètres | Modifier les paramètres* à partir du contrôleur.

Vous pouvez sélectionner et [enregistrer](#) les paramètres relatifs aux éléments suivants :

- [police](#)
- [couleurs](#)
- [dossiers](#)
- [balises](#)
- [paramètres généraux](#)
- [listes de correspondance](#)
- [listes d'arrêt](#)
- [listes de lemmes](#)
- [paramètres du texte et de la langue](#)
- [Paramètres de Concord](#)
- [Paramètres de KeyWords](#)
- [Paramètres de WordList](#)
- [Paramètres utilisateur spécifiques avancés](#)
- [Paramètres du fichier d'index](#)

5.2 accents

Cette fenêtre affiche les caractères accentués disponibles pour la [langue](#) sélectionnée.

5.3 ajouter des notes

Cela vous permet de [sauvegarder](#) des notes avec vos données.

Par exemple, si vous effectuez une concordance et la triez en utilisant vos propres catégories, c'est-à-dire les [catégories définies par l'utilisateur](#), il peut être utile de les lister et de sauvegarder les informations pour une réutilisation ultérieure.

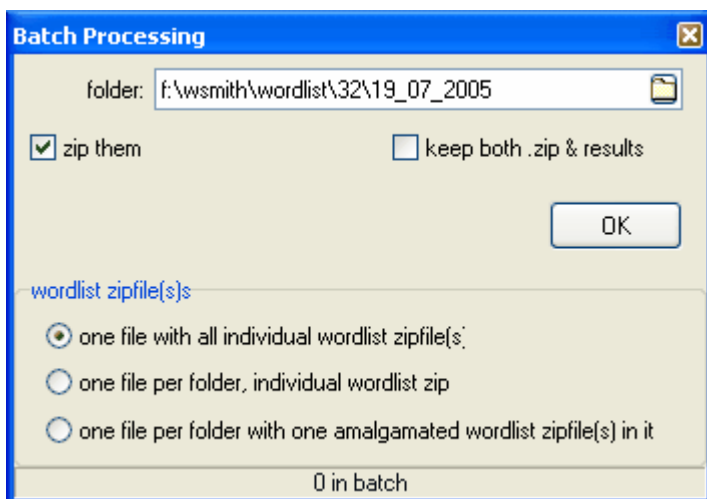
Pour accéder à ces notes en-dehors de la suite **Oxford WordSmith Tools**, sélectionnez le texte en utilisant la touche Maj et les touches directionnelles ou en cliquant avec la souris, puis copiez-le dans le [presse-papiers](#) en utilisant *Ctrl+Inser* et collez-le dans un logiciel de traitement de texte comme le **bloc-notes**.

5.4 traitement par lots

Le but

Le traitement est utile lorsque vous souhaitez créer des listes séparées, mais voulez éviter la tâche fastidieuse de les créer une par une, de sélectionner manuellement chaque fichier texte, créer la liste de mots ou effectuer la concordance, l'enregistrer et ainsi de suite.

Si vous avez sélectionné plusieurs fichiers texte, WordList, Concord et KeyWords peuvent les traiter en tant que lot.



Dossier de destination

Le nom suggéré est la [date](#) du jour. Vous pouvez la modifier si vous le souhaitez. Quel que soit le nom choisi, le dossier sera créé au démarrage du traitement du lot.

Les résultats sont ensuite stockés dans des dossiers dont le nom dérive de celui du dossier principal. Plus précisément, si vous commencez à créer des listes de mots dans

`c:\wsmith\wordlist\05_07_19_12_00`, elles auront cette forme :

`c:\wsmith\wordlist\05_07_19_12_00\0\fred1.lst`

`c:\wsmith\wordlist\05_07_19_12_00\0\jim2.lst`

..

`c:\wsmith\wordlist\05_07_19_12_00\0\mary512.lst`

puis

`c:\wsmith\wordlist\05_07_19_12_00\1\joanna513.lst`

etc.

Le nom du fichier sera le nom du fichier texte source, complété par l'extension standard (*.
`lst`, `*.cnc`, `*.kws`).

Les compresser

Si l'option est cochée, les résultats sont stockés dans un fichier `*.zip` standard. Vous pouvez les extraire avec votre utilitaire de compression habituel, tel que Winzip, ou utiliser WordSmith pour ce faire. Les fichiers contenus dans le zip sont parfaitement identiques aux fichiers non compressés, simplement, ils occupent moins d'espace disque. Par ailleurs, il est préférable pour le système qu'il n'y ait pas des centaines de fichiers au sein du même dossier.

Si vous les compressez, vous obtiendrez :

`c:\wsmith\wordlist\05_07_19_12_00\batch.zip`

et tous les sous-fichiers seront supprimés sauf si vous cochez "Conserver le fichier zip et les résultats".

Un fichier / Un fichier par dossier ?

La première alternative (par défaut) permet de créer un fichier zip contenant toutes vos listes de mots individuelles. Chaque liste de mots, concordance ou liste de mots-clés correspond à un texte source.

Mais que se passe-t-il si vos fichiers sont structurés comme ceci ?

`\..\BNC`

`\..\BNC\written`

`\..\BNC\written\humanities`

`\..\BNC\written\medicine`

`\..\BNC\written\science`

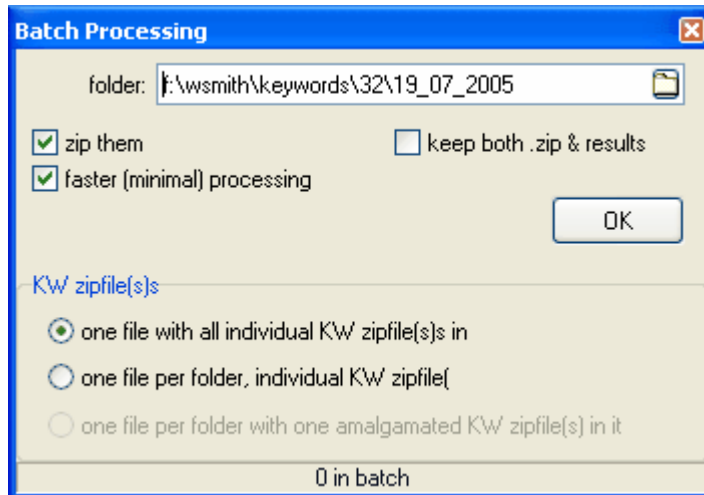
`\..\BNC\spoken`

etc.

L'option *Un fichier par dossier contenant des fichiers zip individuels* permet de créer un fichier zip séparé pour chaque dossier de fichiers texte (par ex. un pour humanities, un autre pour medicine, etc.), avec une liste pour chaque texte source.

L'option *Un fichier par dossier contenant un fichier zip fusionné* permet de créer un fichier zip séparé pour chaque dossier de fichiers, mais qu'une seule liste de mots ou concordance pour tout ce dossier de fichiers.

Traitement rapide (minimal)



Cette case à cocher n'est activée que si vous vous préparez à démarrer un traitement de fichiers où plusieurs types de résultats peuvent être calculés simultanément. Si par exemple vous effectuez une concordance, les [co-occurents](#), les [schémas](#) ainsi que les [schémas de dispersion](#) sont calculés une fois chaque concordance effectuée. De manière analogue, dans KeyWords, les [schémas de dispersion](#), les [liens](#) et les calculs sont traités à mesure que les mots-clés sont calculés.

Si l'option est cochée, seul un calcul minimal est effectué (traitement des mots-clés dans *KeyWords* et concordances dans *Concord*). La tâche est donc effectuée plus rapidement, et vous pouvez toujours recalculer les schémas ultérieurement, tant que les [textes source](#) ne sont ni supprimés ni déplacés.

Exemple : vous voulez créer des listes de mots et avez sélectionné 1 200 fichiers texte provenant d'un magazine appelé "The Elephant".

Vous indiquez que

C:\WSMITH\WORDLIST\ELEPHANT est le nom de votre dossier.

S'il y a déjà un dossier nommé C:\WSMITH\WORDLIST\ELEPHANT, une boîte de dialogue s'affiche, vous demandant la permission d'écraser ce dossier ainsi que tous les sous-dossiers qu'il contient.

Une fois que vous avez cliqué sur OK,

1 200 nouvelles listes de mots sont créées, nommées *trunk.LST*, *tail.LST* ... *digestive system.LST*. Elles se situent toutes dans des [sous-dossiers numérotés](#) d'un dossier nommé C:\WSMITH\WORDLIST\ELEPHANT.

Si vous n'avez pas coché l'option "les compresser en 1 fichier zip", vous trouverez les fichiers dans C:\WSMITH\WORDLIST\ELEPHANT\0.

Si vous avez coché l'option, il y aura donc un fichier C:\WSMITH\WORDLIST\ELEPHANT.ZIP contenant tous vos résultats. Les 1 200 fichiers *.LST créés auront été effacés mais le fichier *.ZIP contient toujours toutes vos listes.

L'avantage d'un fichier *.zip est qu'il occupe beaucoup moins d'espace disque et qu'il est facile de l'envoyer par courrier électronique. WordSmith peut accéder aux résultats à l'intérieur d'un fichier *.zip, vous permettant de sélectionner la liste de mots, concordance, etc., que vous souhaitez afficher.

Obtenir les résultats dans WordSmith

Cliquez sur Fichier | Ouvrir puis modifiez le type de fichier en "fichier lot *.zip". Lorsque vous sélectionnez un fichier zip, une fenêtre listant son contenu s'affiche à l'écran. Double-cliquez sur l'un des fichiers pour l'ouvrir.

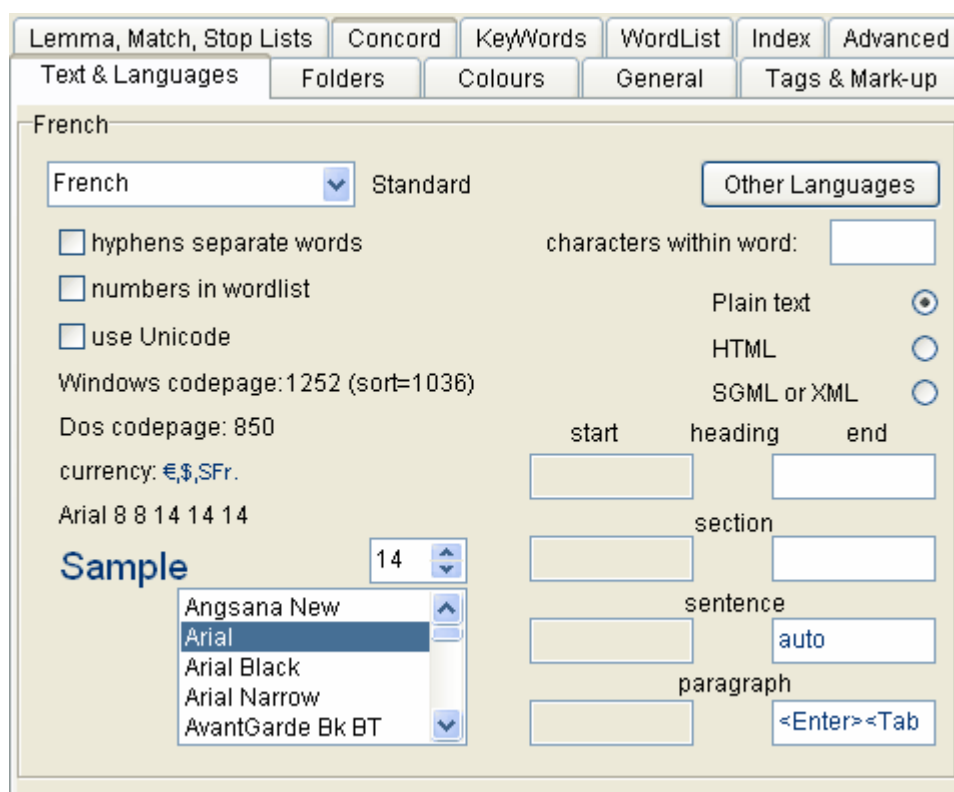
Remarque : il va de soi que Concord ne peut qu'ouvrir une concordance et KeyWords un fichier de liste de mots-clés. Si vous sélectionnez un fichier zip contenant un autre type de fichier, un message d'erreur s'affichera.

5.5 dossiers contenant les lots

The operating system gets unhappy if there are too many files in a folder. So WordSmith batch processing creates a numbered sub-folder ...\\0 which receives the first 500 or so files; if you have chosen to work with more, it then makes another folder ...\\1 for another 500 or so, until all your lists have been made.

5.6 modifier la police

Elles sont accessibles dans le menu principal Paramètres ou en cliquant sur *Modifier les paramètres | Texte et langues* dans le [contrôleur](#). Vous pouvez sélectionner la police Windows et la taille d'affichage que vous préférez, ainsi que la police d'[impression](#), et ce pour toute la suite Oxford WordSmith Tools. Remarque : chaque [langue](#) peut posséder sa propre police par défaut.



Si des données sont visibles, la police change automatiquement. Si vous ne souhaitez qu'aucune modification ne soit apportée aux fenêtres de données, commencez par les minimiser.

Pour définir une colonne de données en gras, italique, souligné etc., utilisez l'option [mise en page](#).

Oxford WordSmith Tools propose des polices convenant à la [langue](#) sélectionnée dans la boîte supérieure gauche. Chaque langue peut posséder sa propre police par défaut. Les paramètres de choix de langues peuvent être affichés (et modifiés avec prudence) dans le fichier `\wsmith4\language_choices.ini`.

5.7 sélectionner les textes favoris

enregistrer les favoris


Cette option permet d'enregistrer la sélection de textes. Elle est utile si l'ensemble des textes est complexe, lorsqu'il inclut plusieurs dossiers par exemple.

Vous pouvez ainsi enregistrer une liste de fichiers texte dont le statut est inconnu ou correspond à vos critères lors d'une [sélection de fichiers en fonction de leur contenu](#). Les fichiers ne correspondant pas aux critères définis sont ignorés.

recupérer les favoris

Cette option permet de lire à partir du disque une sélection précédemment enregistrée.

Par défaut, le [nom de fichier](#) est identique au nom de l'outil pour lequel vous sélectionnez des textes, auquel s'ajoute la portion `recent_chosen_text_files.dat`, dans le dossier principal WordSmith.

Pour charger un ensemble de préférences () , vous pouvez utiliser un fichier texte que vous avez édité avec le Bloc-notes, mais chaque fichier nécessaire doit être spécifié : les caractères de remplacement ne peuvent pas être utilisés et il faut indiquer un chemin d'accès complet lecteur:\dossier

Rubriques connexes : [Sélection de textes](#)

5.8 sélectionner la langue

En règle générale, cette sélection n'est effectuée qu'une fois, lors de votre première utilisation de WordSmith Tools.

Sélectionnez la langue pour le texte que vous analysez dans le [contrôleur](#) en cliquant sur Modifier les paramètres | Texte et langues. La langue et le [jeu de caractères](#) doivent être compatibles avec Windows Western (1 252) et DOS Multilingual (850).

Oxford WordSmith Tools prend en charge une large gamme de langues allant de l'albanais à l'ukrainien. Chinois, japonais, arabe, etc. sont pris en charge avec [Unicode](#). Vous pouvez afficher des listes de mots et des concordances dans différentes langues en même temps.

Le but

L'ordre alphabétique varie considérablement selon les langues. En espagnol, par exemple, l'ordre est le suivant : A,B,C,CH,D,E,F,G,H,I,J,K,L,LL,M,N,Ñ,O,P,Q,R,S,T,U,V,W,X,Y,Z. Par défaut, dans certaines langues, les caractères accentués sont considérés comme leurs équivalents non accentués. En français par exemple, on a *donne*, *donné*, *données*, *donner*, *donnez*, etc. En revanche, dans d'autres langues, les caractères accentués ne sont pas considérés comme étant liés à leur forme non accentuée. En tchèque, on a *cesta* .. *cas* .. *hre* .. *chodník*.

Le tri est effectué selon les routines Microsoft. Si vous traitez des textes dans une langue mal gérée par Microsoft, vous devriez tout de même obtenir des listes de mots triées dans un ordre cohérent.

Remarque : "sensible à la casse" signifie que *Mère* s'affichera après *mère* (et non pas avant *allumette* ni après *zèbre*).

Il est important de comprendre que la comparaison de deux listes de mots (dans KeyWords par exemple) se base sur l'ordre de tri pour obtenir des résultats satisfaisants. Vous obtiendrez des résultats incongrus si vous comparez deux listes de mots déclarées dans des langues différentes.

Comment les langues sont sélectionnées

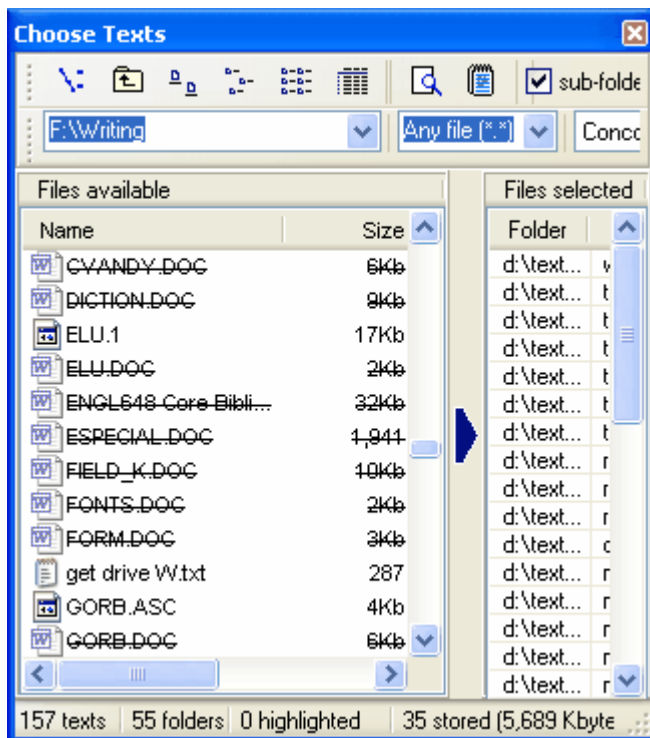
Exécutez le sélecteur de langue à partir du contrôleur principal

Rubriques connexes : [Sélectionner des accents et des symboles](#), [Caractères accentués](#), [Traitement de texte Chinois](#) etc.

5.9 sélectionner les textes

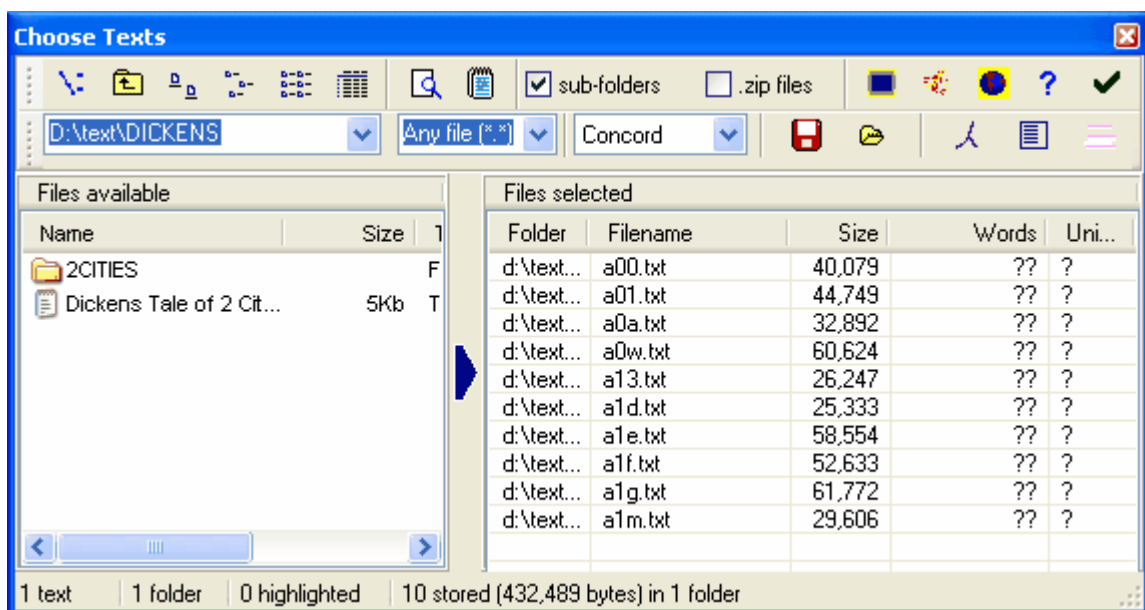
Types de fichiers texte

WordSmith nécessite des [fichiers texte brut](#), comme ceux que vous obtiendriez en enregistrant un fichier *.doc en *.txt. Tout fichier Word *.doc apparaîtra barré dans la liste.



Comment y accéder

Cette fonction est accessible à partir du menu *Fichier* du [contrôleur](#) et du menu *Paramètres* ou encore de l'élément de menu *Nouveau* (🟢) dans les différents outils.



Les deux zones principales situées à gauche et à droite sont :

- fichiers à sélectionner (à gauche)
- fichiers déjà sélectionnés (à droite)

En cliquant sur la flèche bleue, vous pouvez déplacer n'importe quel fichier sélectionné dans la

partie gauche vers la zone des fichiers sélectionnés, à droite. Vous pouvez également les faire glisser de gauche à droite.

La liste affichée à droite donne les détails complets du fichier, c'est-à-dire le nom, la date, la taille, le nombre de mots (marqué ici par ?? car WordSmith ne connaît pas le nombre tant que la concordance n'est pas effectuée ou la liste de mots créée) et indique si le texte est encodé en [Unicode](#) (marqué ici par ? pour la même raison que celle évoquée ci-dessus). À droite de la colonne Unicode se trouve une autre colonne indiquant si chaque fichier texte [correspond à vos critères](#).

Les boutons de la partie supérieure gauche vous permettent de modifier l'affichage des fichiers : icônes, liste ou détails (par défaut).

Si vous n'avez encore jamais utilisé WordSmith (plus précisément si aucune concordance ni liste de mots n'est enregistrée), vous pouvez utiliser un chapitre extrait de *Tale of 2 Cities* de Charles Dickens (ce fichier est sélectionné pour votre première utilisation). Pour éviter que ce texte ne soit automatiquement sélectionné, assurez-vous d'enregistrer au moins une liste de mots ou une concordance. Reportez-vous à la rubrique [listes précédentes](#).


Types de fichiers

La spécification de fichier par défaut est *.* (c'est-à-dire tous les fichiers) mais vous pouvez modifier cela dans la fenêtre située au-dessus de la flèche bleue ou dans le fichier [wshell.ini](#).

Outil

Sur la capture d'écran précédente on peut voir Concord, et on y sélectionne des textes. Ceci dit, il est tout à fait possible d'utiliser d'autres outils (WordList, KeyWords, etc.)



Tri

Pour re-trier la liste, cliquez sur *Nom*, *Taille*, *Type*, *Mots*, *Unicode* ou *Modifié*. Le bouton rouge et le bouton jaune () reclassent les fichiers des deux côtés dans un ordre aléatoire.

Sélectionner tout

Permet de sélectionner tous les fichiers du dossier actuel.

Lecteurs et dossiers

Double-cliquez sur l'un des dossiers pour l'ouvrir. Vous pouvez réaccéder à un dossier si son nom se trouve dans la liste d'historique de la fenêtre des dossiers, et aisément revenir en arrière avec le bouton Windows standard "précédent" . Vous pouvez également cliquer sur le bouton  pour sélectionner un nouveau lecteur ou dossier.

Sous-dossiers

Si l'option est cochée, lorsque vous sélectionnez un lecteur ou un dossier dans la partie gauche, vous sélectionnez également tous les sous-dossiers (et leur contenu) de ce lecteur ou dossier.

Affichage

Ce bouton vous permet de parcourir le fichier sélectionné afin de décider de l'inclure ou pas. Les caractères accentués (par ex. æ, é), symboles monétaires tels que £, ¥, ¢ ainsi que les [balises](#) apparaissent selon les [paramètres des caractéristiques de texte](#). Vous pouvez modifier ces paramètres tout en [affichant](#) le fichier.

Afficher dans Bloc-notes

Grâce à cette option, vous pouvez afficher le contenu du texte dans *Bloc-Notes*, l'éditeur de fichiers texte de Windows.



Récupérer à partir d'Internet

Cette option vous donne accès à [WebGetter](#) pour télécharger du texte à partir d'Internet.

Fichiers zip

Si vous cliquez deux fois sur un [fichier zip](#), vous pouvez accéder à son contenu comme s'il s'agissait d'un dossier. Ces derniers peuvent également être affichés.

Favoris

Deux boutons situés à droite ( et ) vous permettent [d'enregistrer ou d'accéder à une précédente sélection de fichiers](#), vous évitant ainsi des choix fastidieux.

Test Unicode

Ce bouton permet de tester les fichiers sélectionnés. Sur la capture d'écran ci-dessus, aucun test n'a été effectué, c'est pourquoi un point d'interrogation apparaît pour chaque fichier. Si votre fichier texte est au format [Unicode](#), un U s'affiche, un A s'affiche si le texte est au format ASCII ou [Ansi](#), et enfin, un D s'affiche s'il s'agit d'un document Word *.doc.

Effacer

Comme son nom l'indique, cette option vous permet de changer d'avis et de recommencer à zéro. Si des noms de fichiers sont sélectionnés, seuls ces derniers seront effacés.

OK

Ce bouton permet de stocker la sélection actuelle de fichiers. Si l'option "sous-dossiers" est cochée, tous les fichiers du type spécifié, présents dans les sous-dossiers, seront sélectionnés. Vous pouvez vérifier les fichiers sélectionnés dans la section *Tous les paramètres actuels*.

Rubriques connexes : [Rechercher des textes sources](#).

5.10 sélection des fichiers à partir d'une boîte de dialogue standard

La boîte de dialogue est très semblable à celle utilisée pour [sélectionner des fichiers texte](#) ; elle vous permet également de sélectionner des éléments à partir d'un [fichier zip](#). Pour examiner un fichier, vous pouvez utiliser [Viewer](#) : cela ne présente aucun intérêt dans le cas d'une liste de mots, une liste de mots-clés ou une concordance. Cela peut en revanche être utile pour examiner un fichier texte connexe, par exemple un fichier *readme.txt*, se trouvant dans le même fichier zip que votre concordance ou vos listes de mots. Pour sélectionner plusieurs fichiers de manière discontinue, maintenez la touche Ctrl enfoncée tout en cliquant sur les fichiers avec la souris. Pour une sélection continue, maintenez la touche Maj enfoncée tout en sélectionnant le bloc de fichiers.

5.11 couleurs

Cette option est disponible à partir du menu principal des Paramètres dans tous les outils, et à partir du menu Modifier les paramètres dans le [contrôleur](#). Vous pouvez ainsi sélectionner vos couleurs par défaut pour tous les outils. Les couleurs disponibles peuvent être définies pour

texte brut	couleur par défaut
texte en surbrillance	comme ci-dessus lorsque le texte est sélectionné
balises	balisage
terme de recherche	terme de recherche de concordance ; mots dans les listes de mots ou de mots-clés

[terme de tri principal](#)

indique les préférences de tri principal ; utilisé pour le pourcentage de données dans les listes de mots ou de mots-clés

terme de tri secondaire

indique la couleur du tri du premier départageur

terme de contexte

terme de contexte

mots supprimés

toute ligne de données supprimées

ligne non numérotée

toute ligne n'ayant pas été triée par l'utilisateur

terme de recherche

terme de recherche de concordance lorsque l'élément est sélectionné

mis en surbrillance

terme de tri principal

tri primaire lorsque l'élément est sélectionné

mis en surbrillance

terme de tri

tri primaire avec départageur lorsque l'élément est sélectionné

secondaire mis en

surbrillance

terme de contexte

terme de contexte lorsque l'élément est sélectionné

mis en surbrillance

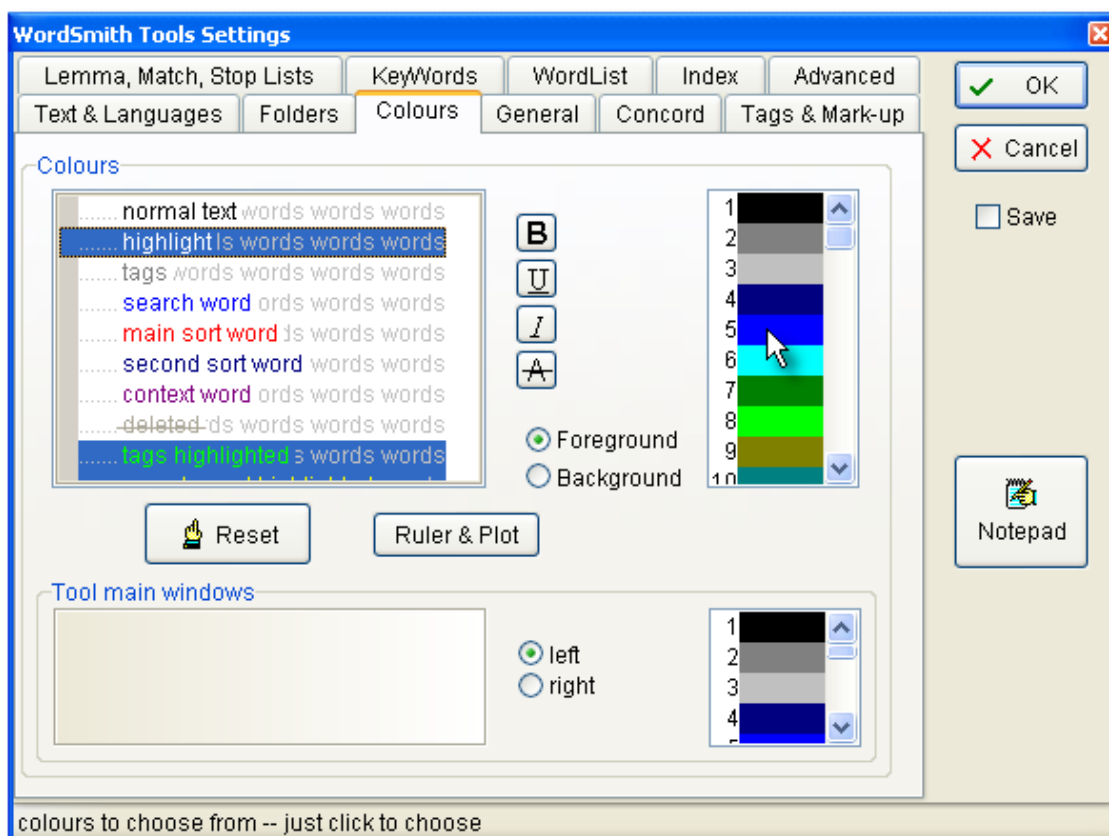
co-occurent le plus

co-occurent le plus fréquent ou terme de cohérence détaillée, [p valeur](#) des mots-clés

fréquent

affichage des textes

dans [Viewer](#)



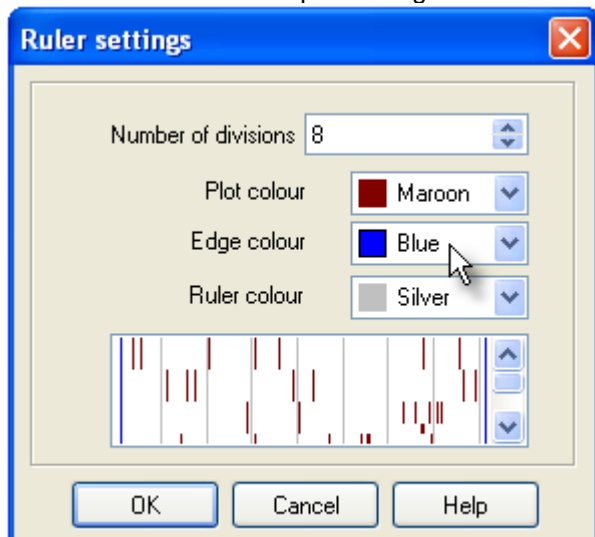
Pour modifier les couleurs, cliquez d'abord sur l'élément que vous souhaitez modifier (vous verrez une différence dans la marge de gauche : ici, "surbrillance" a été sélectionné), puis cliquez sur une couleur dans la boîte de couleurs (là où se trouve le curseur sur la capture d'écran). Les boutons d'options situés en dessous des couleurs déterminent si vous modifiez les couleurs d'arrière-plan ou de premier plan. Vous pouvez cliquer sur le bouton Réinitialiser si vous souhaitez revenir aux valeurs par défaut.

Les mêmes couleurs, ou des dégradés de gris équivalents, apparaissent sur les sorties papier.

Sinon, vous pouvez [paramétrer l'imprimante](#) sur noir et blanc. Dans ce cas, toute colonne n'utilisant pas la couleur dédiée au texte brut apparaîtra en italique (ou en gras ou souligné si vous avez déjà défini la colonne sur italique).

Règle et schéma

Ceci permet d'ouvrir une autre fenêtre de dialogue dans laquelle vous pouvez définir des couleurs et des divisions pour la règle :



Rubriques connexes : [mise en page](#) pour modifier les couleurs de chaque colonne de données.

5.12 totaux des colonnes

Le but

Cette fonction vous permet d'afficher un total ainsi que des statistiques de base pour chaque colonne de données si ces dernières sont numériques.

Procédure

Lorsqu'une liste de mots, de concordance ou de mots-clés est visible, cliquez sur le menu *Affichage | Totaux des colonnes* pour afficher ou masquer les totaux des colonnes.

	Word	Total	Texts	as	Set	at ends well	d cleopatra	s you like it	cimbeline	dy of e
Total		847,645	27,772			24,497	26,601	22,877	28,910	16,...
Max		26,442	35			707	790	694	854	
Min		1	1							
Mean		33.45	5.04			0.97	1.05	0.90	1.14	
Sd.		411.92	7.37			12.58	12.58	12.23	13.90	
2	A	14,209	35	0		1.89	1.28	1.96	1.65	
3	ABOUT	396	35	0		0.03	0.03	0.03	0.03	
4	ACT	371	35	0		0.11	0.16	0.02	0.04	
5	AFTER	400	35	0		0.07	0.07	0.07	0.07	
6	AGAIN	737	35	0		0.11	0.08	0.07	0.09	
7	AGAINST	561	35	0		0.08	0.06	0.06	0.06	
8	AGE	175	35	0		0.02	0.01	0.04	0.01	
9	ALL	3,644	35	0		0.40	0.44	0.40	0.41	

Ici vous pouvez voir les totaux des colonnes dans une liste de cohérence détaillée basée sur des pièces de Shakespeare. La liste est triée en fonction de la colonne Textes : les éléments du haut apparaissent dans l'intégralité des 35 pièces utilisées pour constituer la liste. Dans le cas de la pièce "Anthony and Cleopatra", l'article A représente 1,28 % des mots dans cette colonne, c'est-à-dire 1,28 % des mots dans la pièce Anthony and Cleopatra. Dans le cas du mot ACT, il représente le pourcentage le plus élevé dans sa ligne (ce mot est plus utilisé en terme de pourcentage dans cette pièce que dans les autres).

5.13 calculer une nouvelle colonne de données

Le but

Cette fonction fait apparaître une calculatrice vous permettant de choisir des fonctions pour calculer les valeurs que vous souhaitez. Par exemple, une liste de mots est généralement accompagnée de la fréquence de chaque type, et cette fréquence est donnée en pourcentage de l'ensemble des tokens des textes. Vous pouvez insérer une colonne supplémentaire affichant la fréquence comme pourcentage du nombre de types de mots, ou une colonne affichant la fréquence comme pourcentage du nombre de fichiers texte à partir desquels la liste de mots a été créée.

Procédure

Cliquez sur *Calculer | Nouvelle colonne* et créez votre propre formule. Les boutons standard d'une calculatrice s'affichent : les nombres de 0 à 9, la décimale, les parenthèses et les 4 fonctions de base. Dans la partie droite, il y a une liste des fonctions mathématiques standard pouvant être utilisées (pi, racine carrée, etc.). Pour y accéder, cliquez deux fois dessus. En dessous se trouve une partie permettant l'accès à vos données de la liste active, répertoriant tous les en-têtes de colonnes chiffrés. Vous pouvez également les faire glisser ou cliquer deux fois dessus.

Absolu et relatif

Il existe deux manières d'accéder à vos propres données. Un accès relatif (par défaut) signifie que la nouvelle colonne doit accéder aux données à partir d'une autre colonne, mais sur la même ligne. C'est le processus des feuilles de calcul. Absolu désigne un accès à partir d'une colonne et d'une ligne fixes.

Exemples

- Rel(2) ÷ 5** pour chaque ligne de données, la nouvelle colonne contient les données de la colonne 2 de la même ligne, divise la valeur par 5 et inscrit le résultat dans votre nouvelle colonne.
- RelC(2)** pour chaque ligne de données, la nouvelle colonne contient les données de la colonne 2 de la même ligne, ajoute la valeur au total actuel et inscrit le résultat dans votre nouvelle colonne.
- Rel(3) + (Rel(2) ÷ 5)** pour chaque ligne de données, la nouvelle colonne contient les données de la colonne 2 de la même ligne, divise la valeur par 5, l'ajoute aux données de la colonne 3 de la même ligne, et inscrit le résultat dans votre nouvelle colonne.
- Abs(2;1) ÷ 5** pour chaque ligne de données, la nouvelle colonne contient les données de la colonne 2 de la ligne 1, divise la valeur par 5 et inscrit le résultat dans votre nouvelle colonne. Cet exemple n'était qu'une illustration. En effet, il donne le même résultat à chaque ligne.
- Rel(2) ÷ Abs(2;1) × 100** pour chaque ligne de données, la nouvelle colonne contient les données de la colonne 2 de la même ligne, divise la valeur par la valeur de la colonne 2 de la ligne 1, multiplie le résultat obtenu par 100 et inscrit le résultat total dans votre nouvelle colonne. Ainsi, la colonne 3 donne un pourcentage des résultats du haut de la colonne 2. La valeur de la première ligne serait 100 %. Cependant, à mesure que les fréquences diminuent, le pourcentage de l'élément le plus fréquent diminue aussi.

Vous pouvez modifier le format des variables ainsi calculées ou même les supprimer : Reportez-vous à la section [mise en page](#).

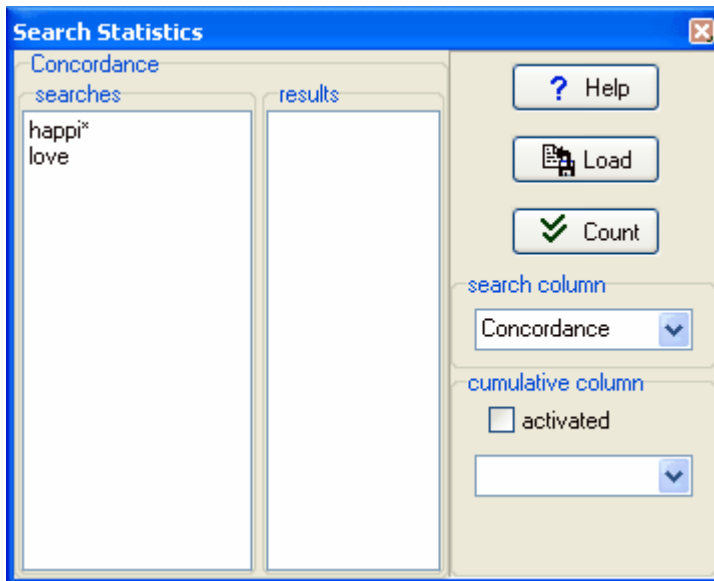
Rubriques connexes : [calculer la fréquence des données](#), totaux des colonnes

5.14 calculer la fréquence des données

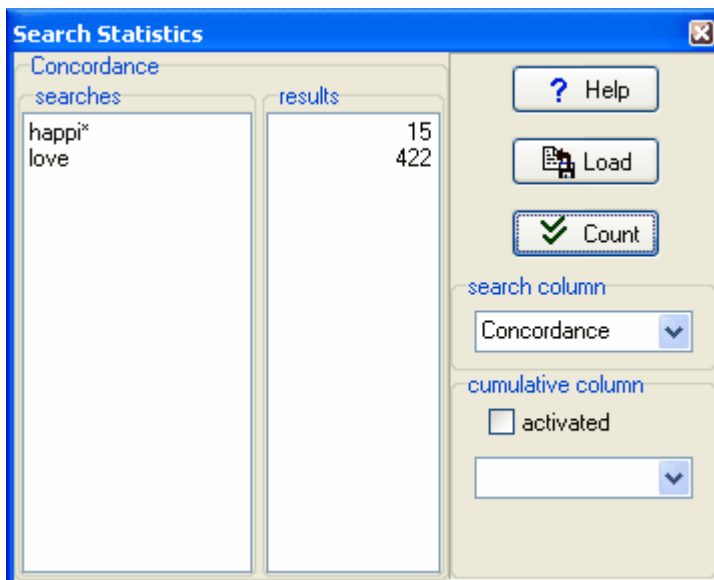
Il est possible de savoir combien de lignes de concordance comportent "happi*" ou combien d'éléments d'une liste de mots se terminent par *1y. Pour ce faire, sélectionnez *Statistiques de résumé* dans le menu *Calculer*.

Exemple

Une concordance a déjà été calculée. Sélectionnez n'importe quel endroit des lignes de concordance puis cliquez sur *Calculer : Statistiques de résumé*. Saisissez **happi*** et **love** dans la fenêtre de recherche.



Cliquez sur Compter, et ceci devrait apparaître :



Une fois l'opération traitée, il s'avère que 15 de vos lignes de concordance comportent **happi*** et 422 comportent le mot entier **love** (mais pas **loved**, **loves**).

Colonne de recherche

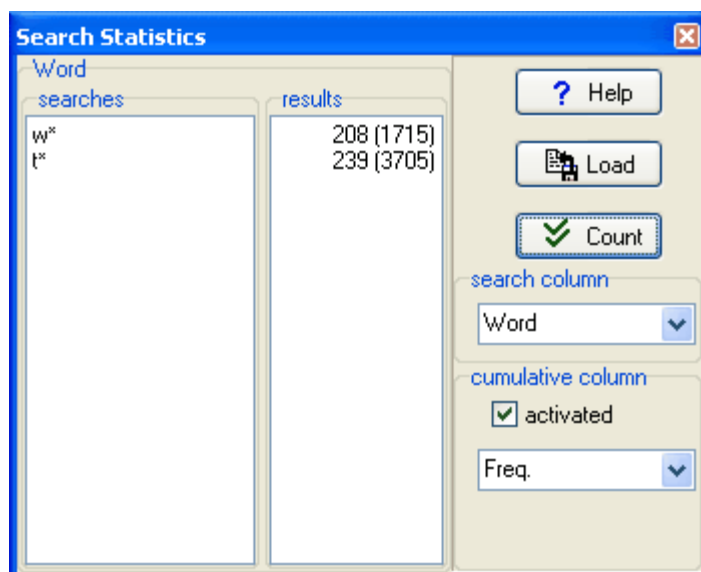
Cette zone combinée vous permet de sélectionner la colonne de données à inclure dans le décompte.

Colonne cumulative

Un décompte cumulatif inscrit les résultats dans une colonne de données autre que celle traitée lors de votre recherche. Les colonnes de cette zone combinée ne comportent que des données numériques. Sélectionnez-en une, et assurez-vous que l'option *activée* est cochée.

Dans cet exemple, une liste de mots a été traitée, et la recherche portait sur les mots commençant par **w** ou **t**. Il y en a 208 qui commencent par **w**. Entre parenthèses se trouve le

nombre 1715. Cela signifie que le décompte cumulé des mots commençant par w en terme de fréquence (colonne *Freq.*) est 1715. En d'autres termes, la fréquence moyenne est d'environ 8 (1715/208). Cependant, pour les mots commençant par t, bien que le nombre absolu soit similaire (239), la fréquence moyenne cumulée est d'environ 15. Cela s'explique par le fait qu'en anglais, il y a de nombreux mots à haute fréquence commençant par t.



Charger

Cette fonction vous permet de charger dans une fenêtre de recherche n'importe quel fichier au format [texte brut](#) que vous avez précédemment préparé.

Rubriques connexes : [calculer une nouvelle colonne de données](#).

5.15 copier vos résultats

La méthode la plus rapide et la plus simple pour copier vos données dans votre traitement de texte, par exemple, est de sélectionner avec les touches directionnelles et d'appuyer sur Ctrl +Ins. Ceci place les données dans le presse-papiers.

Si vous cliquez sur Copier (mise à jour nécessaire), plusieurs possibilités s'offrent à vous :

[presse-papiers](#)

[enregistrer en tant que fichier texte](#)

[impression](#)

[enregistrer](#) en tant que données (méthode différente de l'enregistrement en tant que fichier texte. Cela vous permet d'enregistrer de façon à pouvoir accéder à vos données ultérieurement)

tous: copier toutes les lignes et colonnes de données

sélectionnées: copier uniquement les lignes et les colonnes que vous avez déjà mises en surbrillance

spécifier: copier en fonction des critères de l'outil :

Dans **Concord**, vous pouvez [enregistrer](#) des lignes de concordance que vous avez classées [en fonction de vos propres catégories](#). Vous pouvez spécifier une catégorie à enregistrer (entre **a** et **z** ou entre **A** et **Z**).

5.16 traitement personnalisé

Cette fonctionnalité, d'un niveau avancé, est disponible dans le menu Modifier les paramètres | Avancé.

Le but

Je ne peux pas connaître les critères que vous avez définis pour traiter vos textes, à l'exception des critères prédéfinis tels que le choix des textes, du terme de recherche, etc. Vous serez amené à effectuer des vérifications ou modifications spécifiques des données avant qu'elles ne soient converties au format **WordSmith**. Vous pourriez par exemple avoir besoin de lemmatiser un mot en fonction des exigences spécifiques de votre langue.

Cette fonction le permet. Si vous avez choisi de filtrer les concordances, **Concord** appelle le fichier *.dll chaque fois qu'il trouve un résultat pour votre mot de recherche au cours du traitement de vos fichiers texte. Votre fichier *.dll reçoit des informations sur les éléments trouvés, et le fichier peut ainsi modifier le résultat ou envoyer des informations à **Concord** pour qu'il l'ignore.

Procédure

Sélectionnez votre fichier *.dll (il peut avoir n'importe quel nom) et cochez une ou plusieurs options dans la page Avancé. Pour cela, il faut appeler des fonctions standard et connaître leurs noms et leurs formats. C'est à vous d'écrire votre propre programme *.dll qui effectuera la tâche désirée. Il peut être écrit dans n'importe quel langage (C++, Java, Pascal, etc).

Exemple de lemmatisation d'un mot dans WordList

Le fichier *.dll est fourni lors de l'installation, compilé et prêt à l'utilisation.

Votre fichier *.dll doit contenir une fonction comportant les spécifications suivantes :

```
function WordlistChangeWord(  
    original : pointer;  
    language_identifiant : DWORD;  
    is_Unicode : WordBool) : pointer; stdcall;
```

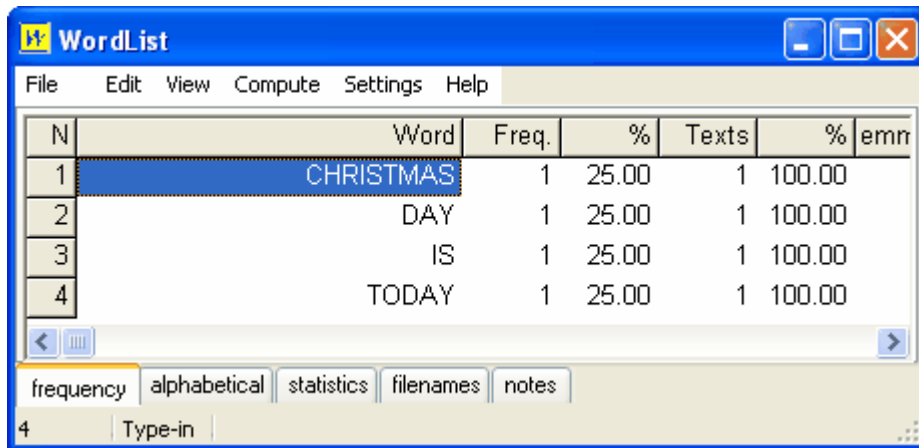
L'élément language_identifiant est un nombre correspondant à la langue avec laquelle vous travaillez. Reportez-vous aux [valeurs LCID \(List of Locale ID\) affectées par Microsoft](#) .

Ainsi, l'original (envoyé par WordSmith) peut être PCHAR (7 ou 8 bits) ou PWIDECHAR (Unicode 16 bits), et le résultat fourni par votre fichier *.dll peut mener vers :

- a) néant (si vous ne souhaitez pas que le terme d'origine apparaisse dans votre liste)
- b) le même PCHAR/PWIDECHAR s'il ne doit y avoir aucune modification
- c) une forme de remplacement

Voici un exemple où le texte source était

```
Today is Easter Day.
```



Le code source dans Delphi pour ce fichier *.dll est le suivant :

```
*****
```

```
library WordSmith4CustomDLL;
```

```
uses
```

```
  Windows, SysUtils;
```

```
{
```

```
  This example uses a very straightforward Windows routine for comparing
  strings, CompareStringA and CompareStringW which are in a Windows .
  dll.
```

```
  The function does a case-insensitive comparison because
  NORM_IGNORECASE (=1) is used. If it was replaced by 0, the comparison
  would be case-sensitive.
```

```
  In this example, EASTER gets changed to CHRISTMAS.
```

```
}
```

```
function WordlistChangeWord(
```

```
  original : pointer;
```

```
  language_identifier : DWORD;
```

```
  is_Unicode : WordBool) : pointer; stdcall;
```

```
begin
```

```
  Result := original;
```

```
  if is_Unicode then begin
```

```
    if CompareStringW(
```

```
      language_identifier,
```

```
      NORM_IGNORECASE,
```

```
      PWideChar(original), -1,
```

```
      PWideChar(widestring('EASTER')), -1) - 2 = 0
```

```
    then
```

```
      Result := pwidechar(widestring('CHRISTMAS'));
```

```
  end else begin
```

```
    if CompareStringA(
```

```
      language_identifier,
```

```
      NORM_IGNORECASE,
```

```
      PChar(original), -1,
```

```
      PChar('EASTER'), -1) - 2 = 0
```

```
    then
```

```
      Result := pchar('CHRISTMAS');
```

```
  end;
```

```
end;

function ConcordChangeWord(
  original : pointer;
  language_identifier : DWORD;
  is_Unicode : WordBool) : pointer; stdcall;
begin
  Result := WordlistChangeWord(original,language_identifier,
is_Unicode);
end;

function KeyWordsChangeWord(
  original : pointer;
  language_identifier : DWORD;
  is_Unicode : WordBool) : pointer; stdcall;
begin
  Result := WordlistChangeWord(original,language_identifier,
is_Unicode);
end;

function HandleConcordanceLine
(source_line : pointer;
 hit_position,
 hit_length : word;
 byte_position_in_file,
 language_id : DWORD;
 is_Unicode : WordBool;
 filename : pchar) : pointer; stdcall;

function extrasA : string;
begin
  Result := #9+pchar(filename)+
           #9+ IntToStr(byte_position_in_file)+
           #9+ IntToStr(hit_position)+
           #9+ IntToStr(hit_length);
end;

function extrasW : widestring;
begin
  Result := extrasA;
end;

var f : TextFile;

  output_file : string;
begin
  Result := source_line;
  output_file := ChangeFileExt(ParamStr(0),'')+
'_user_dll_concordance_lines.txt';
  if (not IsPathDelimiter(ExpandUNCFileName(ParamStr(0)),1)) and
(DiskFree(Ord(UpCase(output_file[1]))-64) > 1024*2000) then
  try
    if FileExists(output_file) then begin
      AssignFile(f,output_file);
      Append(f);
    end else begin
      AssignFile(f,output_file);
      Rewrite(f);
    end;
    if is_Unicode then
```



```
        Writeln(f,pwidechar(source_line)+extrasW)
    else
        Writeln(f,pchar(source_line)+extrasA);
    Flush(f);
    CloseFile(f);
except
end;
end;

exports

    ConcordChangeWord,
    KeyWordsChangeWord,
    WordlistChangeWord,
    HandleConcordanceLine;

begin
end.
```

5.17 personnalisation des menus contextuels

Connexion

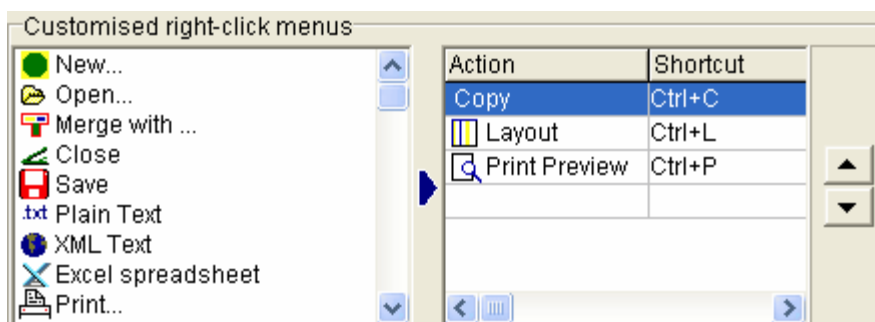
La connexion peut s'avérer utile si vous obtenez des résultats étranges ou que vous souhaitez afficher des détails sur la manière dont ils ont été obtenus. Si cette option est activée, WordSmith enregistre quelques informations sur la progression des résultats dans le fichier journal.

Dans le contrôleur principal, cliquez sur *Modifier les paramètres | Avancés*. Ici, vous pouvez activer ou désactiver la connexion et sélectionner un nom de fichier approprié. Si vous l'activez, vous aurez la possibilité d'effacer le fichier journal précédent. Rubriques connexes : [rapports d'erreur envoyés par courrier électronique](#).

Personnalisation des menus contextuels

Toutes les grilles de données possèdent des "menus contextuels" qui apparaissent lorsque vous cliquez avec le bouton droit de la souris.

Pour personnaliser ceci, cliquez sur *Modifier les paramètres | Avancés* dans le contrôleur principal.



Une liste d'options de menu s'affiche à gauche. Vous pouvez les ajouter à la liste de droite en sélectionnant celles de votre choix et en cliquant sur le bouton du milieu. Pour supprimer une option de la liste de droite, sélectionnez-la et cliquez sur *Suppr.* Pour réorganiser les options, cliquez sur les flèches.

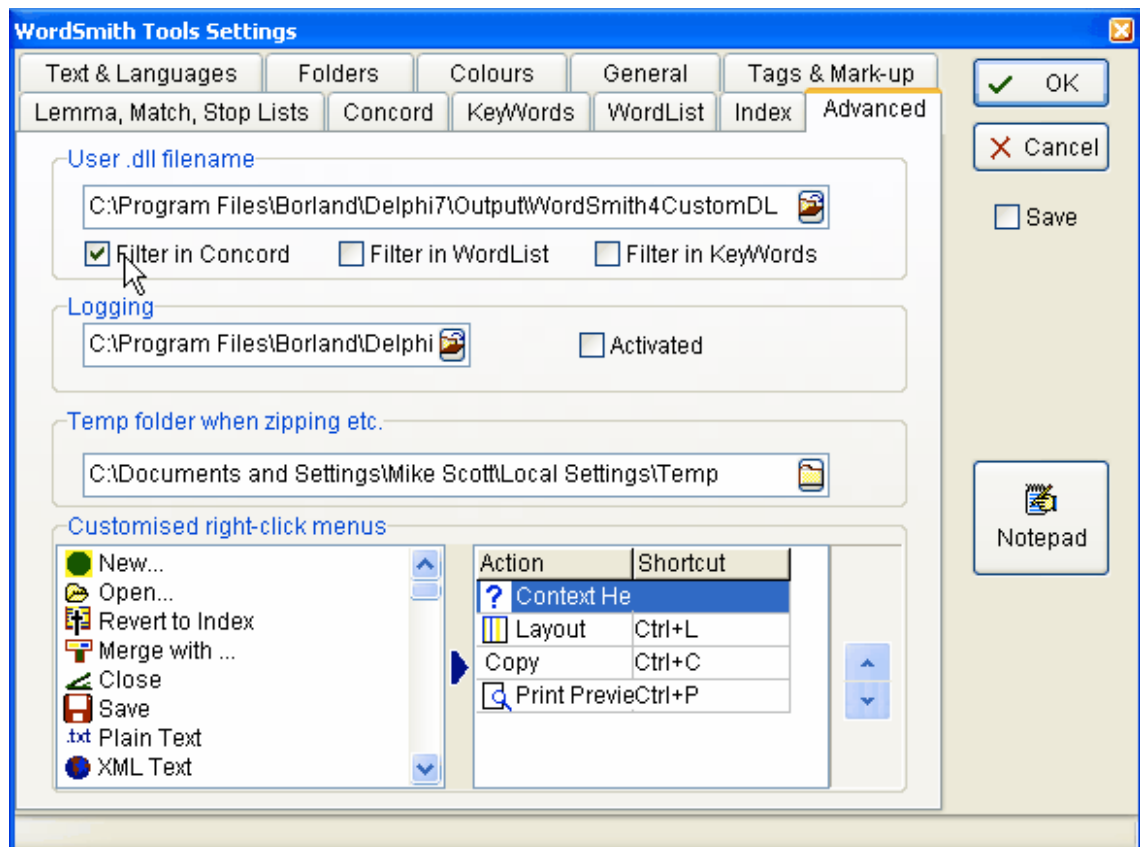
Toutes les options présentes dans votre menu contextuel apparaissent également dans la [barre](#)

[d'outils.](#)

Pour enregistrer les options de manière permanente, reportez-vous à la section [Enregistrer les paramètres par défaut.](#)

Fichier *.dll utilisateur

Si vous avez un fichier DLL que vous souhaitez utiliser pour récupérer les résultats de WordSmith, vous pouvez le sélectionner ici. Le fichier sélectionné par cet utilisateur, `WordSmithCustomDLL.dll`, est fourni lors de l'installation du programme et peut être utilisé si vous le souhaitez. Si l'option "Filtrer dans Concord" est cochée, ce fichier DLL ajoute toutes les lignes de concordance trouvées dans un fichier texte brut à un fichier nommé `Concord_user_dll_concordance_lines.txt` dans votre dossier `\wsmith4`, si l'espace disque est suffisant.



Rubriques connexes : [options des menus et des boutons.](#)

5.18 édition d'une liste de données

Lorsqu'une liste de mots est affichée à l'écran, vous voyez ceci :

The screenshot shows a window titled '480.lst' with a menu bar (File, Edit, View, Compute, Settings, Help) and a toolbar. Below the toolbar is a table with the following data:

N	Word	Freq.	%	Texts	% emmas	Set
4	AA	6		6	1.25	
5	AAA	5		4	0.83	
6	AAC	1		1	0.21	
7	AACHEN	1		1	0.21	
8	AACUTE	1		1	0.21	
9	AAH	3		3	0.63	

Below the table is a toolbar with buttons for 'frequency', 'alphabetical', 'statistics', 'filenames', and 'notes'. At the bottom of the window, there is a status bar showing '72,028 Type-in AA'.

Dans la barre d'outils située dans la partie inférieure,

72,028 Type-in AA

le nombre affiché dans la première cellule est le nombre de mots de la liste de mots actuelle et **AA**, situé dans la troisième cellule, est le mot sélectionné.

Lorsque l'utilisateur saisit n'importe quelle expression, WordList tentera de trouver ce qui a été saisi dans la liste.

Si vous cliquez avec le bouton droit sur la deuxième cellule, ceci s'affiche :

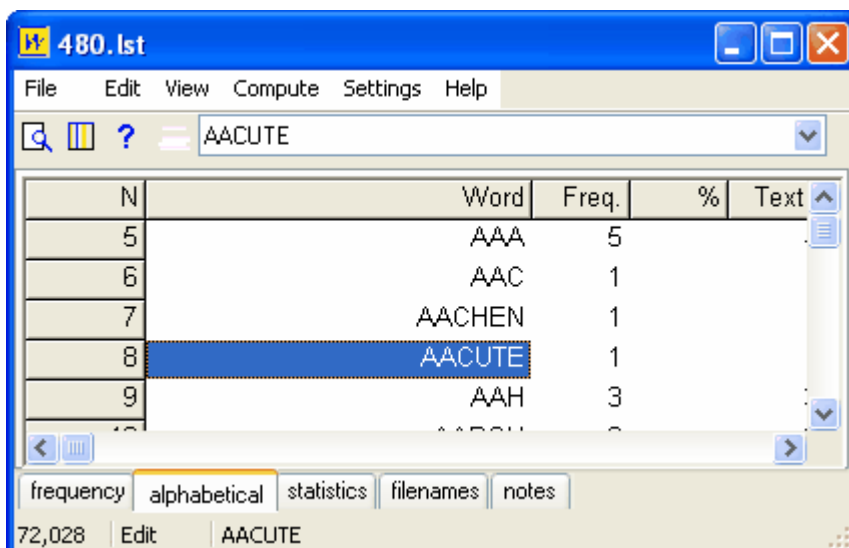


Vous pouvez modifier les options de cette liste sur *Classer* (pour classer vos mots par type, par exemple adjectifs, noms, etc.) ou *Éditer*, pour les modifier. Remarque : certaines données sont calculées en utilisant d'autres données et ne peuvent donc pas être éditées. Par exemple, les données relatives au pourcentage indicateur de fréquence sont basées sur la fréquence des mots et sur le nombre total de mots. Vous pouvez éditer la fréquence des mots, mais pas le pourcentage de fréquence des mots.

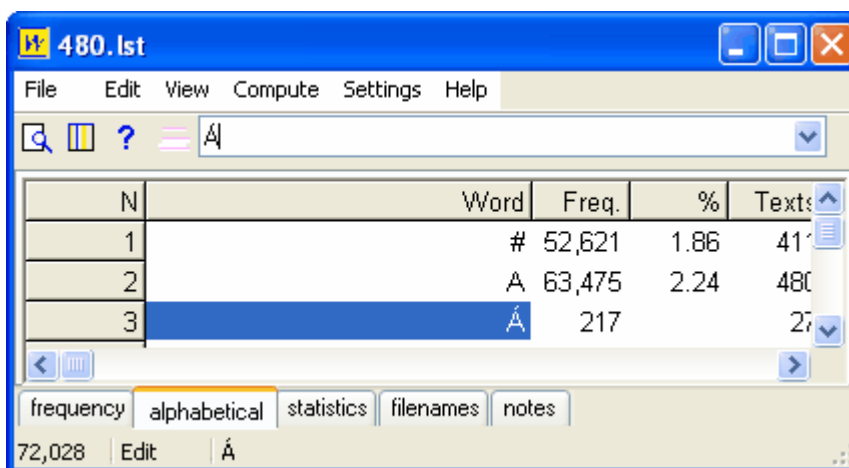
Cliquez sur *Éditer*.

À présent, saisissez n'importe quelle lettre dans la colonne que vous souhaitez éditer. Cette action affiche la barre d'outils (si elle n'était pas déjà visible avant) pour que vous puissiez modifier la forme du mot ou sa fréquence. Si vous tapez le mot de telle manière qu'il correspond à un autre mot de la liste, cette dernière sera modifiée afin de refléter les modifications apportées.

Dans l'exemple, on souhaite corriger **AACUTE**, qui devrait en fait être **Ā**.



En saisissant **Á**, les résultats s'affichent immédiatement dans la fenêtre :



En cliquant sur la flèche à droite de la zone combinée d'édition, vous voyez s'afficher le mot d'origine au cas où vous décideriez de le garder en mémoire.



Une fois les modifications terminées, vous pouvez [re-trier](#) (🌀) les données, et si vous avez modifié un mot tel que **AAAAAGH** en un mot déjà existant tel que **AAGH**, vous pouvez [regrouper](#) les deux entrées.

Rubriques connexes : [regrouper des entrées](#), [rechercher des fichiers source](#).

5.19 chercher des fichiers pertinents

Le but

Supposons que vous ayez identifié *muscle*, *fibre*, *protéine* comme mots-clés d'un texte donné. Vous voudrez peut-être savoir si d'autres textes de votre corpus contiennent ces mots.

Procédure

Cette fonction est accessible dans toute fenêtre de données contenant l'option **F**, par exemple une liste de [mots-clés](#).

Vous pouvez ainsi rechercher tous les fichiers texte contenant **au moins une fois chacun** des mots que vous avez marqué (avec **m**). Avant de cliquer, [sélectionnez l'ensemble de textes](#) que vous souhaitez parcourir.

Ce que vous obtenez

Une concordance basée sur tous les mots que vous avez marqués, indiquant les [fichiers texte](#) dans lesquels ils ont été trouvés. Mais c'est une concordance "approximative" : les fichiers texte dans lesquels ne figurent pas *tous* les mots sélectionnés seront ignorés.


5.20 instructions de cours ou de session

Lorsque WordSmith est exécuté lors d'une session de formation, vous pouvez rendre certaines instructions accessibles à vos étudiants.

Pour ce faire, assurez-vous qu'il y a un fichier nommé `teacher.rtf` dans votre dossier principal `\wsmith4` où se trouvent les programmes ou dans le "dossier d'instructions" dont l'utilisation est expliquée dans la rubrique [Paramètres réseau par défaut](#). Si ce fichier est trouvé, il sera automatiquement affiché au démarrage de WordSmith. Pour qu'il ne s'affiche plus, il suffit de le renommer. Vous pouvez éditer ce fichier au moyen d'un logiciel de traitement de texte RFT (Rich Text Format), tel que MS Word™, en l'enregistrant au format `*.rtf`.

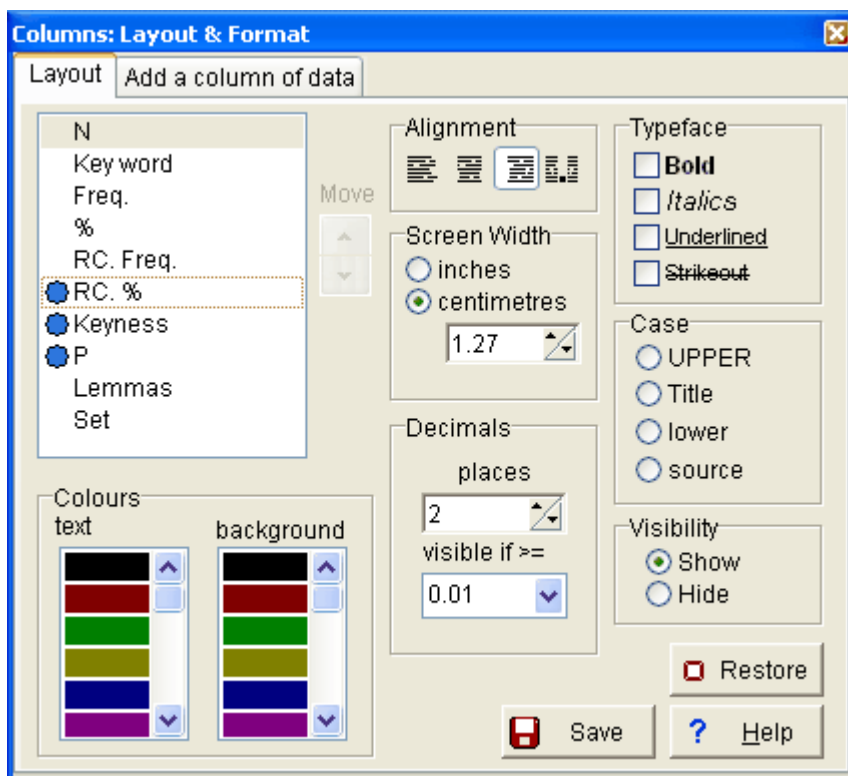
Rubriques connexes : [Paramètres réseau par défaut](#)

5.21 mise en page et format

Lorsqu'une liste est ouverte, cliquez avec le bouton droit ou cliquez sur Affichage | Mise en page  pour sélectionner vos formats d'affichage préférés pour chaque colonne de données.

Mise en page ou Ajouter des données ?

L'onglet *Mise en page* vous permet de définir la mise en page de vos données. *Ajouter une colonne de données* [permet de calculer une nouvelle variable](#).



Dans la fenêtre supérieure gauche de Mise en page s'affichent les en-têtes de colonne disponibles. 3 de ces en-têtes ont été activés (en cliquant dessus) pour que les paramètres puissent être modifiés :

Déplacer

Cliquez sur les flèches pour déplacer une colonne vers le haut ou vers le bas afin de les afficher dans un ordre différent.

Alignement

Cette section vous permet de choisir un alignement à gauche, centré, à droite ou encore un alignement basé sur les décimales, et ce pour chaque colonne.

Type de caractères

Texte normal, gras, italique et/ou souligné. Si aucune option n'est cochée, les caractères seront en style normal.

Largeur de l'écran

Vous pouvez choisir l'unité qui vous convient le mieux (cm ou pouces).

Casse

Vous avez le choix entre minuscules, MAJUSCULES, Titre ou source (tel quel dans le fichier texte d'origine). Par défaut, la plupart des données sont en majuscules.

Décimales

Vous pouvez préciser ici le nombre de décimales pour les données numériques, lorsque cela est applicable.

Visibilité

Vous avez le choix entre "afficher" ou "masquer", ou encore afficher si la valeur est plus grande qu'un nombre préalablement défini. Si *** s'affiche, alors cette option n'est pas applicable aux

données de la colonne actuellement sélectionnée.

Couleurs

La partie inférieure gauche de la fenêtre affiche les couleurs disponibles pour l'arrière-plan et le premier plan. Cliquez sur une couleur pour modifier l'affichage de la colonne d'informations actuellement sélectionnée.

Restaurer

Ce bouton permet de restaurer les paramètres tels qu'ils étaient auparavant. Cela vous permet de supprimer les mises en page personnalisées enregistrées (voir ci-dessous).

Enregistrer

Cette option d'enregistrement permet d'affecter une mise en page préférée à toutes les listes futures. Supposons que vous ayez une concordance ouverte. Si vous modifiez la mise en page et [enregistrez](#) la concordance de manière habituelle, les paramètres seront tout de même enregistrés. Cependant, lors de la prochaine concordance, la mise en page par défaut de WordSmith sera appliquée. Si vous utilisez cette option d'enregistrement, la prochaine concordance aura la même mise en page que la concordance actuelle.

De plus, une mise en page personnalisée sera enregistrée dans le dossier `\wsmith4` et portera par exemple le nom de `liste_concordance_personnalisee.dat`.

Il est également possible de masquer ou afficher certaines colonnes de données en modifiant les paramètres de votre fichier `wshell.ini`. Par exemple, dans la section [Concord] de `\wsmith4\wshell.ini`, vous pouvez masquer de façon permanente la colonne Définir. Pour ce faire, il vous suffit de modifier

```
show set column=YES
en
show set column=NO
```

Rubriques connexes : [paramètres d'enregistrement par défaut](#), définir les [couleurs](#) dans le [contrôleur](#) principal de la suite Oxford WordSmith Tools.

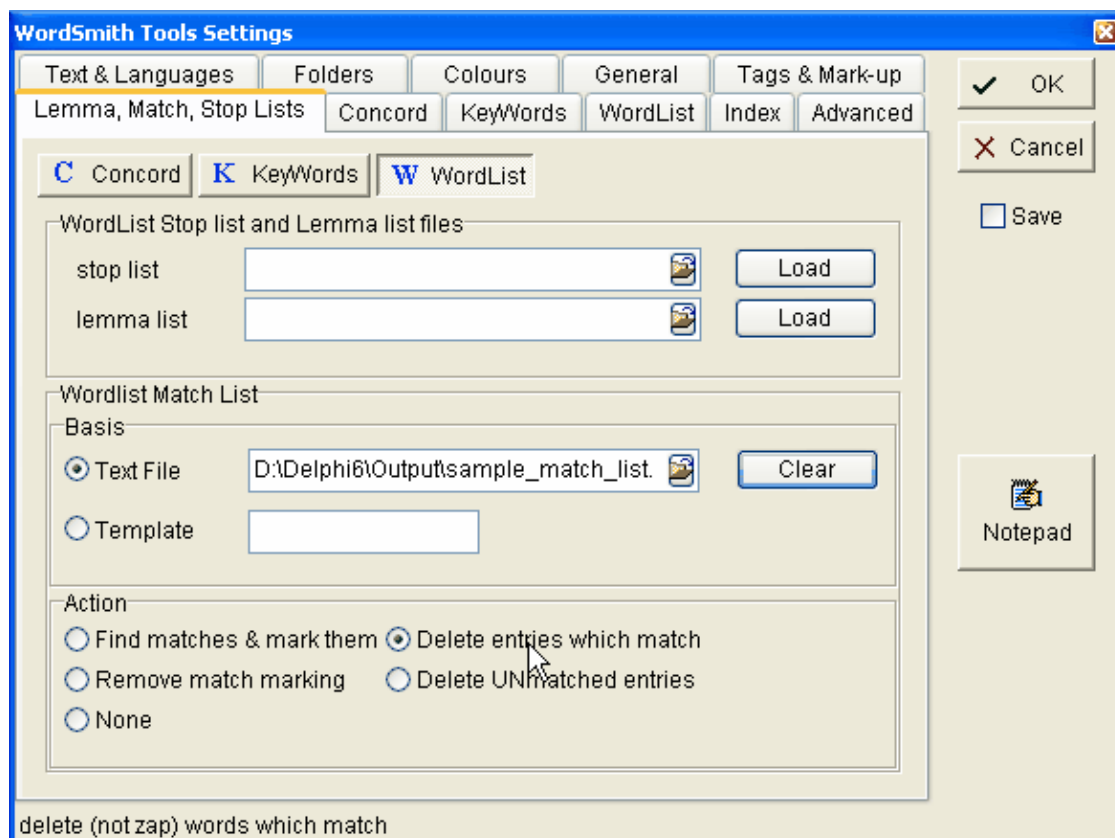
5.22 faire correspondre des mots dans une liste

Le but

Cette fonction permet de filtrer une liste. Vous pouvez mettre en relation les entrées d'une concordance ou d'une liste de mots (liste de mots WordList, liste de co-occurents etc.) en indiquant les mots qui vous intéressent. Par exemple, pour marquer tous les mots de votre liste qui sont des mots de fonction, ou tous ceux qui finissent par **-ing**. Ceux qui correspondent sont marqués d'un (~). Lorsque les entrées sont marquées, vous avez la possibilité d'effacer toutes celles qui sont marquées ou toutes celles qui ne le sont pas.

Procédure

Cliquez sur la colonne contenant les données que vous souhaitez faire correspondre. La plupart du temps, cette colonne comporte des mots, et non des nombres. Cliquez ensuite sur le bouton *Liste de correspondance*. La boîte de dialogue des paramètres principaux du contrôleur s'affiche.



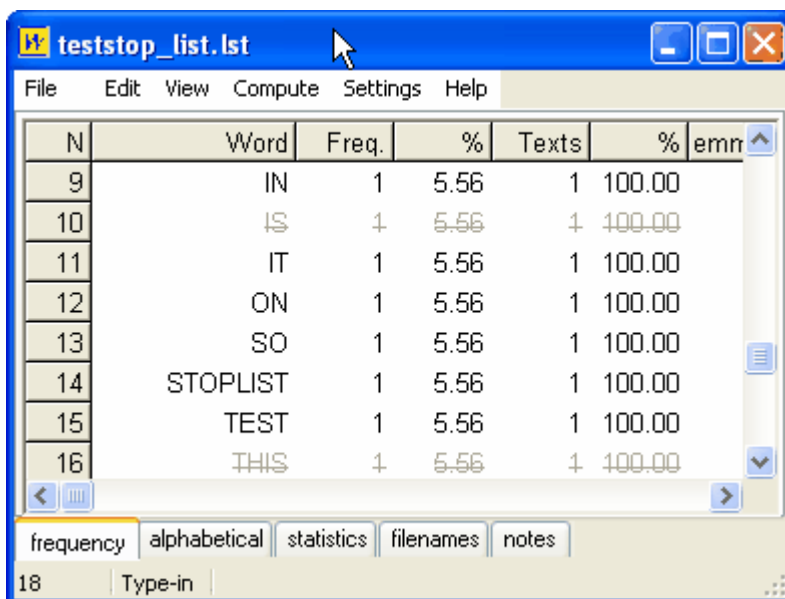
Fichier texte ou Modèle

Vous avez deux possibilités : filtrer en utilisant un fichier texte contenant tous les mots qui vous intéressent (par exemple un fichier de texte brut contenant des mots de fonctions), ou un modèle de filtre comme ***ing** (qui recherche tous les mots finissant par **ing**). Si vous sélectionnez un fichier, le contrôleur vous indique le nombre de mots qu'il contient après l'avoir lu.

Action

Il est important de bien définir l'action à effectuer. Le curseur sur la capture d'écran ci-dessus indique que l'action souhaitée est d'effacer tous les mots qui correspondent.

L'outil actuel vérifie ensuite chaque entrée de la colonne sélectionnée dans votre liste active pour déterminer si elle correspond au modèle ou à un des mots contenus dans votre fichier de texte brut. Les entrées qui correspondent sont marquées ou effacées, selon l'action demandée (comme dans l'exemple ci-dessous, où la liste de correspondance comportait **IS** et **THIS**).



N	Word	Freq.	%	Texts	%	emn
9	IN	1	5.56	1	100.00	
10	IS	1	5.56	1	100.00	
11	IT	1	5.56	1	100.00	
12	ON	1	5.56	1	100.00	
13	SO	1	5.56	1	100.00	
14	STOPLIST	1	5.56	1	100.00	
15	TEST	1	5.56	1	100.00	
16	THIS	1	5.56	1	100.00	

Vous pouvez obtenir les statistiques des correspondances grâce à l'option de menu [Statistiques de résumé](#).

Rubriques connexes : [Comparaison de listes de mots](#), [Comparer les versions](#), [Listes d'arrêt](#), [Correspondances de lemmes](#)

5.23 imprimer et aperçu avant impression

Par défaut, cette fonction permet d'afficher un aperçu avant impression, montrant l'aspect de votre page actuelle de données, et à partir duquel vous pouvez imprimer.

Augmenter et diminuer la taille

Zoom 100% (🔍) ou ajuster à la fenêtre (🖥️), ou sélectionnez un affichage dans la liste. L'affichage correspond exactement au rendu que vous aurez sur papier. Les quelques différences entre ce que vous voyez et ce que vous obtenez sont liées à des différences de police de caractères.

Page suivante (➡) et précédente (⬅)

Permet d'aller à la page suivante ou de revenir à la page précédente.

Orientation portrait (📄) ou paysage (📖)

Imprime selon la disposition souhaitée.

En-tête, Pied de page, Marges

Vous pouvez insérer un en-tête et un pied de page dans chaque page. Si vous incluez <Date>, c'est la date du jour qui sera insérée. Cliquez sur les nombres pour modifier les marges et visualiser le résultat dans l'aperçu avant impression sur la droite.

Imprimer (🖨️)

La page d'impression Windows standard apparaît et est configurée par défaut pour imprimer la page actuelle. Vous pouvez sélectionner d'autres pages dans la boîte de dialogue standard si vous le souhaitez.

Rubriques connexes : [Paramètres d'impression](#)

5.24 première utilisation de WordSmith

Pour les personnes utilisant WordSmith pour la première fois, le processus peut paraître complexe. Après tout, la première fois que vous avez utilisé un logiciel de traitement de texte, cela paraissait compliqué, mais vous saviez déjà ce qu'était un texte et comment en écrire un...

Un petit fichier texte est livré lors de l'installation de WordSmith. Le logiciel vérifie si vous l'avez déjà utilisé, et si ce n'est pas le cas, il sélectionnera automatiquement ce fichier texte lorsque vous ouvrirez Concord, WordList, etc. Le processus de vérification de WordSmith est le suivant :

- 1) Est-ce que des concordances ou des listes de mots ont été [enregistrées](#) ?
et
- 2) Est-ce qu'un ensemble de [fichiers texte favoris](#) a été enregistré pour une récupération aisée ?

5.25 listes précédentes

Cette fenêtre affiche la liste de résultats obtenus lors de précédentes utilisations de WordSmith.

Pour afficher l'une d'elles, sélectionnez-la et faites un double-clic dessus. Les données s'afficheront dans l'outil adéquat.

Le menu contextuel de la fenêtre est accessible en cliquant avec le bouton droit de la souris.

Pour supprimer une entrée, sélectionnez-la et appuyez sur *Suppr.*
Pour re-trier vos entrées par ordre alphabétique, cliquez sur *Re-trier.*

5.26 quitter WordSmith

Alt+X est la touche de raccourci.

La fermeture du [contrôleur](#) Oxford WordSmith Tools entraîne la fermeture de tous les outils.

Si vous appuyez sur *Alt+X* ou utilisez les commandes système de fermeture, une boîte s'affichera, et vous pourrez enregistrer tous les ensembles de données avant que l'outil en question ne se ferme. Si une fenêtre de données est toujours ouverte, vous devrez confirmer la fermeture.

Si vous êtes pressé, l'option "Fermer sans vérifier" permet de ne pas effectuer toutes ces vérifications.

Par défaut, toute liste de mots, concordance ou liste de mots-clés enregistrée ou extraite sera automatiquement restaurée au démarrage de la suite Oxford WordSmith Tools. Il est possible de désactiver cette fonctionnalité temporairement à partir d'une option de menu Wshell ou de la désactiver définitivement dans le fichier `\wsmith4\wshell.ini`.

5.27 réduire les données à n entrées

Vous voudrez peut-être réduire de façon aléatoire une grande liste de mots, concordance etc. (par exemple pour en extraire un échantillon). Cette option de menu vous permet de spécifier le nombre d'entrées que vous souhaitez avoir dans votre liste. Si vous réduisez les données, les entrées seront aléatoirement [supprimées](#) pour n'en garder que le nombre désiré. Cette procédure est irréversible. Rien ne sera modifié sur votre disque, mais si vous changez d'avis, vous devrez refaire le calcul ou revenir à une version sauvegardée antérieure.

Rubriques connexes : [ignorer](#), [édition d'une liste de données](#).

5.28 enregistrer au format texte






Le but

Enregistrer au format texte signifie enregistrer vos données sous forme de fichier [texte brut](#) (par opposition au format WordSmith d'extraction de données). Il est généralement plus rapide de copier le texte sélectionné dans le [presse-papiers](#), par exemple si vous souhaitez simplement insérer vos résultats dans votre logiciel de traitement de texte.

Pour copier les données en couleur ou exporter un schéma, utilisez le [presse-papiers](#).

Dans le cas d'une concordance, si vous ne souhaitez que les mots qui apparaissent dans votre ligne de concordance (pas le nombre de caractères mentionné ci-après), utilisez le presse-papiers et effectuez un collage ou un collage spécial au format graphique.


Procédure

Pour accéder à cette fonction, cliquez sur Enregistrer sous... | Texte brut () , texte au format XML () , feuille de calcul Excel () ou Imprimer dans un fichier (via F3 ou ) ou Copier () vers un fichier texte.

Les options comprennent :

en-tête	mots que vous souhaitez enregistrer au début de vos données (ne saisissez rien si vous ne souhaitez rien enregistrer) ;
numéroté(e)(s)	enregistrement des nombres qui figurent dans la colonne de gauche
séparateur de colonnes	il s'agit d'une tabulation par défaut, mais vous pouvez spécifier un autre séparateur de colonnes
lignes	toutes celles que vous avez mises en surbrillance/une sélection donnée, par exemple 1-10, 5-, -3
colonnes	toutes celles que vous avez mises en surbrillance/une sélection donnée (la colonne 1 est celle avec les nombres)

Vous pouvez ensuite extraire aisément les données dans votre feuille de calcul, base de données, logiciel de traitement de texte, etc. (si vous souhaitez les utiliser sous forme de tableau dans un logiciel de traitement de texte, enregistrez-les d'abord au format texte, puis cliquez sur l'option Convertir le texte en tableau dans votre logiciel de traitement de texte, si disponible. Séparez le texte aux tabulations).

Remarque : Une feuille de calcul Excel () peut traiter un maximum de 65 000 lignes. Ce qui s'affiche ressemble à ceci :

WordSmith Tools 4.0 -- 16/10/2004						
	Word	Freq.	%	Texts	%	Lemmas
1						
2						
3	1	THE	46	8.88	1	100
4	2	A	20	3.86	1	100
5	3	OF	16	3.09	1	100
6	4	AND	13	2.51	1	100
7	5	SPLEEN	11	2.12	1	100
8	6	TO	11	2.12	1	100
9	7	BLOOD	9	1.74	1	100
10	8	IN	8	1.54	1	100
11	9	IS	8	1.54	1	100
12	10	#	6	1.16	1	100
13	11	WAS	6	1.16	1	100
14	12	WITH	6	1.16	1	100

Enregistrer une ligne de concordance au format texte enregistrera le nombre de caractères que vous aurez spécifié (dans les paramètres du [contrôleur Concord](#)). De cette façon, vous pouvez spécifier un nombre fixe de caractères pour que vos termes de recherche s'alignent correctement lorsque vous utilisez une police de caractères non proportionnelle.

Si vos données comportent un schéma, vous obtiendrez une feuille de calcul dans le fichier Excel qui ressemble à ceci.

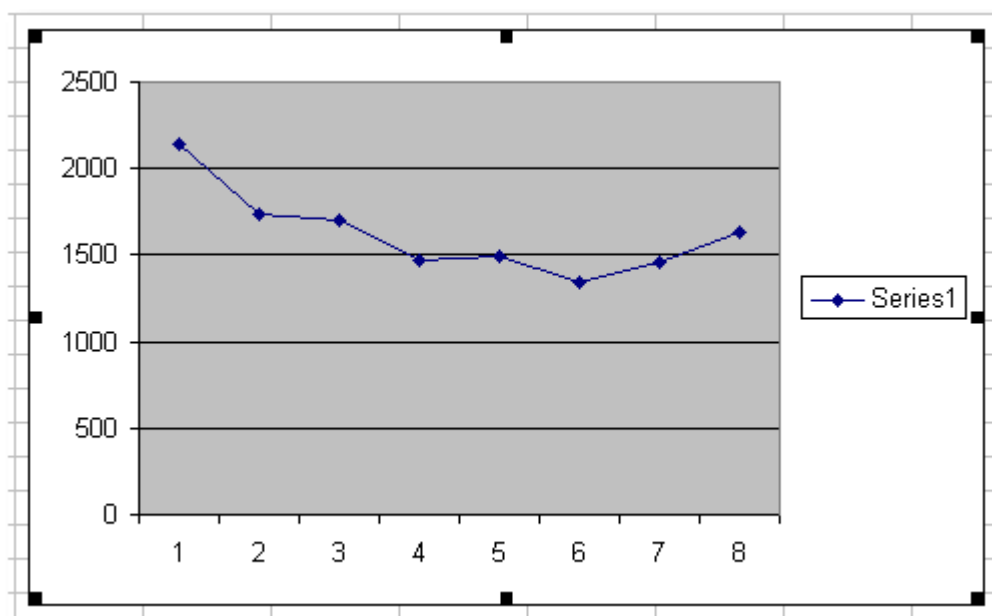
Divisions	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Raw Totals	2140	1735	1699	1466	1495	1347	1458	1632
cep.txt	0.237705	0.155738	0.262295	0.065574			0.065574	0.213115
hp6.txt	0.333333	0.166667		0.166667		0.333333		
bn4.txt	0.333333		0.166667	0.166667	0.333333			
asu.txt	0.142857	0.142857	0.571429	0.142857				
cbg.txt	0.146154	0.223077	0.207692				0.207692	0.215385
g2e.txt	0.222222	0.111111	0.277778	0.111111	0.111111	0.055556	0.055556	0.055556
cm6.txt	0.5					0.5		
a0f.txt	0.111111	0.111111		0.222222	0.111111	0.111111	0.111111	0.222222
guy.txt	1							
k5m.txt	0.336957	0.141304	0.086957			0.086957	0.163043	0.184783
hj4.txt	0.217949	0.166667	0.333333	0.064103			0.051282	0.166667
h9t.txt	1							
bme.txt	0.25					0.25	0.25	0.25
acr.txt	0.117647	0.176471	0.176471	0.058824	0.176471	0.117647	0.117647	0.058824
cmu.txt	0.2	0.2				0.2	0.2	0.2
hd7.txt	0.333333		0.666667					
jye.txt	0.571429	0.142857	0.142857			0.142857		
k9j.txt	0.333333	0.166667	0.333333				0.166667	
gxf.txt	0.333333	0.333333					0.333333	
b2w.txt	1							
edf.txt	1							

Les données du schéma sont réparties dans les segments définis pour la [règle](#) (huit dans

l'exemple de la capture d'écran) et chaque pourcentage est placé dans la bonne colonne. La case B3 indique donc que 23,7 % des données de cep.txt se trouvent dans le premier huitième du fichier texte. Vous pouvez afficher les pourcentages dans Excel, comme ceci :

2140	1735	169
23.77%	15.57%	26.
33.33%	16.67%	
33.33%		16.
14.29%	14.29%	57.
14.67%	27.31%	20.

En haut sont affichées les données brutes, qu'Excel peut représenter sous forme de graphique.



Dans le cas de texte en XML (🌐), vous obtenez un petit fichier .HTM et un gros fichier .XML. En cliquant sur le fichier .HTM, vous pouvez visualiser vos données page par page, avec des boutons pour avancer, revenir en arrière, et aller à la première ou dernière page. C'est lui qui lit les données à partir du fichier .XML.

5.29 enregistrer les valeurs par défaut

Les paramètres peuvent être modifiés en cliquant sur *Modifier les paramètres* dans le [contrôleur](#) de la suite Oxford WordSmith Tools.

Les menus des outils vous permettent d'accéder aux paramètres suivants :

Couleurs, Dossiers, Texte, Général, Balises, Listes d'arrêt, Concord, KeyWords, WordList

Ces onglets permettent de régler les paramètres qui affecteront un ou plusieurs outils.

- [couleurs](#) personnaliser les couleurs par défaut
- [dossiers](#) indiquez à WordSmith les dossiers que vous utilisez habituellement

texte	jeu de caractères , traitement des trait d'unions et nombres, extension des fichiers par défaut
général	restaurer le dernier fichier , impression
balises	balises à ignorer, fichier de balises, chargement automatique des fichiers de balises
listes d'arrêt	pour Concord, KeyWords et Wordlist
correspondan	fichiers à faire correspondre, ou fichiers de lemmes pour ce
Concord	marquer les lemmes dans une liste de mots, etc.
KeyWords	nombre d'entrées, système de tri, voisinage de co-occurrence
WordList	procédure , p valeur maximum, base de données et fréquences minimum des associés, nom de fichier du corpus de référence
	longueur de mot et fréquences, ratio type/token , paramètres de cluster

paramètres permanents et fichier wshell.ini

Pour enregistrer vos paramètres, cochez la case enregistrer après avoir réglé les paramètres. Vous pouvez également modifier le fichier `wshell.ini`, qui a été installé en même temps que la suite Oxford WordSmith Tools. Il précise tous les paramètres que vous utilisez régulièrement pour la suite de programmes, comme les [dossiers](#) où se trouvent vos textes et résultats, les [couleurs](#) de l'affichage, les [polices](#), les [colonnes](#) à afficher par défaut dans une concordance, etc.

Pour afficher le fichier `\wsmith4\wshell.ini`, cliquez sur *Paramètres | Afficher l'élément actuel*.

afficher le fichier d'aide

Dans l'onglet général de Modifier les paramètres se trouve une case "afficher le fichier d'aide". Si vous cochez la case, le fichier d'aide s'affichera à chaque démarrage de WordSmith. Cela est utile pour les utilisateurs occasionnels du logiciel, qui travaillent par exemple sur une installation en réseau.

maximes

Vous pouvez éditer `wsmith4\sayings.txt` (avec le bloc-notes), qui contient les maximes contenues dans la fenêtre principale du [contrôleur](#), si vous n'aimez pas les maximes ou souhaitez en ajouter.

réseau et paramètres de CD-ROM par défaut

Si vous exécutez WordSmith directement à partir d'un CD-ROM, vous ne pouvez pas y enregistrer vos paramètres car il se trouve en lecture seule. Windows placera ailleurs le fichier `wshell.ini`, généralement dans le dossier racine de `c:\`.

Lors de la première utilisation de WordSmith, il vous faudra modifier les paramètres, sélectionner les [dossiers](#) appropriés, les caractéristiques de [texte](#), les détails de [balises](#), etc. et activer la case enregistrer, après quoi vos paramètres seront enregistrés. Vous pouvez modifier et enregistrer vos paramètres aussi souvent que nécessaire.



De même, vous ne pouvez pas modifier de façon permanente les paramètres par défaut lorsque vous travaillez en réseau, car cela aurait des répercussions pour les autres utilisateurs. Votre administrateur réseau devrait avoir installé le programme de telle sorte que vous disposiez d'une copie du fichier `wshell.ini`, que vous pouvez éditer et modifier. Si Oxford WordSmith Tools détecte une copie de `wshell.ini` dans ce dossier, il pourra charger vos paramètres personnels.

5.30 enregistrer les résultats

Pour enregistrer les résultats corrigés cliquez sur *Enregistrer* ou appuyez sur F2. Tous les résultats sont ainsi enregistrés et vous pouvez y revenir ultérieurement. Vous pouvez supprimer les éléments supprimés, en cliquant sur [ignorer](#).

Les données enregistrées se trouvent dans un format propre à **Oxford WordSmith Tools**. Ceci permet de réutiliser ultérieurement ces données. Il est impossible de les utiliser aisément en dehors des outils. Si vous souhaitez exporter vos données vers une feuille de calcul, un logiciel d'infographie, une base de données ou un logiciel de traitement de texte etc., cliquez sur [enregistrer au format texte](#) ou copiez les données dans le [presse-papiers](#).

n'enregistrer qu'une partie des données

Par défaut,  et  enregistrent toutes les données que vous n'avez pas [ignorées](#). Pour n'enregistrer qu'une partie sans perdre le reste, cliquez sur [Copier](#).

5.31 rechercher et remplacer

Certaines listes, comme les listes de [noms de fichiers](#) permettent de rechercher et remplacer. Comme dans le cas d'une recherche normale, la recherche s'effectue sur la *colonne active* de données. Si vous sélectionnez "Sensible à la casse", la casse de chaque lettre doit correspondre. Vous pouvez effectuer une recherche "vers le bas" à partir de l'entrée suivante jusqu'à la fin, ou "vers le haut" jusqu'à la première entrée. Vous pouvez sélectionner également "Confirmer chaque modification" si vous le souhaitez.

Mot entier – ou caractère de remplacement astérisque

"Mot entier" signifie rechercher un mot comportant un [séparateur de mot](#) à ses deux extrémités. Utilisez l'astérisque pour rechercher un suffixe ou un préfixe. Ainsi, ***ed** renverra toutes les entrées se terminant par **ed** ; **un*** renverra toutes les entrées commençant par **un**. ***cherche*** trouvera toutes les entrées comportant **cherche** (**cherche**, **recherche** ou **chercher** dans une entrée de deux ou trois mots comme **elle le cherche**).

Une liste de mots-clés peut-être triée en fonction des suffixes. Reportez-vous à la section [Tri dans WordList](#).

Rubriques connexes : [Recherche par saisie](#), [Rechercher avec F12](#), [Caractères accentués et symboles](#).

5.32 recherche par saisie

Lorsque les résultats d'une colonne sont affichés par ordre alphabétique, vous pouvez rapidement trouver un mot par saisie. À mesure que vous tapez, **WordSmith** se rapproche du mot recherché. Si vous avez saisi les cinq premières lettres et que **WordSmith** a trouvé le résultat correspondant, un bip retentira et la fenêtre d'édition se fermera. Le mot recherché devrait alors s'afficher.

Rubriques connexes : [Mode édition et mode saisie](#), [Recherche d'un mot ou d'une partie d'un mot](#), [Rechercher et remplacer](#), [Tri dans WordList](#)

5.33 rechercher un mot ou un fragment de mot

Vous pouvez rechercher un mot, une partie d'un mot ou un nombre dans n'importe quelle liste. La recherche concerne la *colonne active* de données. Si vous sélectionnez "Sensible à la casse", la casse de chaque lettre doit correspondre. Vous pouvez effectuer une recherche "vers le bas" à partir de l'entrée suivante jusqu'à la fin, ou "vers le haut" jusqu'à la première entrée.

Mot entier – ou caractère de remplacement astérisque

"Mot entier" signifie rechercher un mot comportant un [séparateur de mot](#) à ses deux extrémités.

Utilisez l'astérisque pour rechercher un suffixe ou un préfixe. Si vous saisissez ***ed**, vous trouverez toutes les entrées terminant par **ed**; **un*** renverra toutes les entrées commençant par **un**. ***cherche*** renverra toutes les entrées comportant **cherche** (**cherche**, **recherche** ou **chercher** dans une entrée de deux ou trois mots telle que **elle le cherche**).

Une liste de mots-clés peut-être triée en fonction des suffixes. Reportez-vous à la section [Tri dans WordList](#).

Rubriques connexes : [Recherche par saisie](#), [Rechercher et remplacer](#), [Caractères accentués et symboles](#).

5.34 voir les noms de fichiers

Ce bouton ouvre une nouvelle fenêtre et affiche le [nom du fichier](#) d'où proviennent vos données. Si nécessaire, vous pouvez éditer ces noms (si les fichiers texte ont été déplacés ou renommés par exemple). Pour ce faire, cliquez sur remplacer (🔍).

Quand ensuite vous [sauvegardez les résultats](#), l'information est définitivement enregistrée.

Dans le cas des listes de mots-clés, les données proviennent d'une liste de mots. Si la liste de mots est basée sur un seul fichier texte, le nom du fichier texte s'affiche, si elle est basée sur plusieurs fichiers, vous voyez le nom du fichier de liste de mots lui-même : pour voir le nom des fichiers texte, ouvrez la liste de mots et cliquez sur le nom des fichiers.

Rubriques connexes : [rechercher des fichiers source](#).

5.35 listes d'arrêt

Les listes d'arrêt sont des listes de mots qui ne doivent pas être inclus dans votre analyse. Il peut s'agir par exemple de mots très courants sans grand intérêt comme *le, de, était, est, ce* que vous souhaitez exclure d'une liste de mots ou d'une analyse de mots-clés.

Pour utiliser les listes d'arrêt, vous devez avoir créé un fichier avec le **bloc-notes** ou tout autre logiciel de traitement de texte, indiquant tous les mots à ignorer. Il faut séparer les mots par des virgules ou aller à la ligne pour chaque mot. Vous pouvez utiliser des majuscules ou des minuscules. Vous pouvez utiliser un point-virgule pour les lignes de commentaires.

Vous trouverez un fichier `stoplist.stp` (dans le dossier `\wsmith4`) que vous pouvez utiliser comme base et enregistrer sous un nouveau nom.

Exemple

```
; commentaire  
LE,CE,EST  
IT  
WILL
```

Sélectionnez ensuite *Liste d'arrêt* dans le menu pour spécifier la liste d'arrêt à utiliser. Vous pouvez utiliser des listes d'arrêt différentes pour **WordList** et **KeyWords**. Si la liste d'arrêt est *activée*, les mots qui s'y trouvent ne seront plus intégrés à une liste de mots. Pour utiliser toujours la même liste d'arrêt, indiquez-le dans le fichier `wshell.ini` comme [paramètre par défaut](#).

Pour des explications plus complètes avec des captures d'écran, reportez-vous à la section [Liste de correspondance](#).

Pour créer une liste d'arrêt, vous pouvez également utiliser **WordList** sur un gros corpus de textes, en indiquant une fréquence minimum élevée si vous ne souhaitez que les mots à haute fréquence. Enregistrez ensuite au format texte. Utilisez ensuite **Text Converter** pour le formater avec **stoplist.cod** comme [fichier de conversion](#).

Rubriques connexes : [Création d'un fichier de balises](#), [Liste de correspondance](#), [Lemmatisation](#).

5.36 arrêter le traitement

À mesure que WordSmith parcourt les fichiers texte ou re-trie les données, une fenêtre affichant des barres de progression s'ouvre dans le contrôleur. Si l'action est possible, un bouton *Arrêt* s'affiche également. Quatre options s'offrent à vous si vous cliquez sur ce bouton :

Continuer

continuer comme si rien n'avait été interrompu

Terminer le fichier puis arrêter

c'est un arrêt "dans les règles de l'art". Terminer le fichier signifie que vous avez un compte-rendu des actions effectuées et de celles qui n'ont pas pu être effectuées faute de temps. Comment ? En examinant les noms de fichier de la liste de mots, de la concordance ou de l'élément que vous venez de créer.

Arrêter

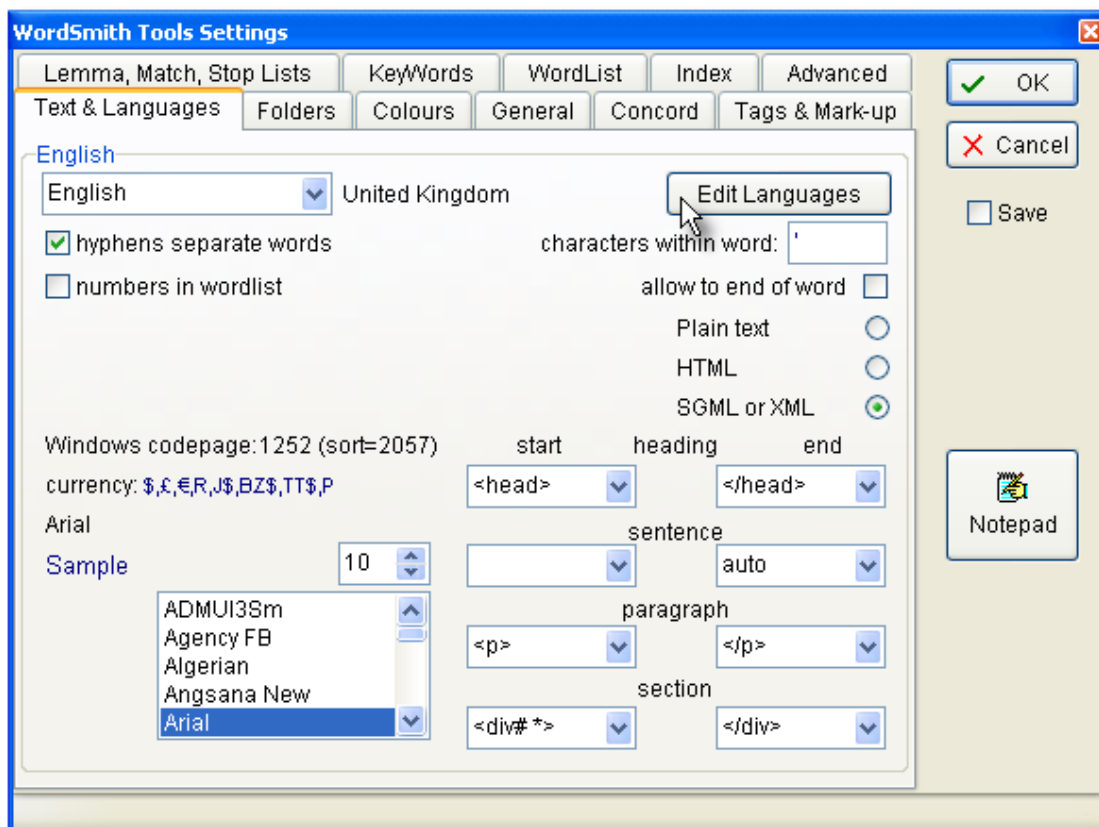
c'est un arrêt moins conventionnel, très utile si vous parcourez des fichiers très volumineux sur CD-ROM. Dans ce cas, WordSmith arrête de traiter le fichier actif mais retient les données qu'il a récoltées jusque là.

Quitter cet outil

C'est un arrêt d'urgence. L'outil entier (Concord, WordList, etc.) se fermera et une partie de la [mémoire](#) ressource pourra être perdue. Le [contrôleur](#) ne se fermera pas.

Cliquez à nouveau sur *Arrêt* pour appliquez vos choix.

5.37 texte et langues



Ces paramètres agissent sur la façon dont WordSmith va traiter vos textes. En haut se trouvent les boîtes qui permettent de choisir la famille de langue (par exemple l'anglais) et la sous-famille (Royaume-Uni, Australie, etc). Ces choix sont fixés par les préférences préalablement définies. Cela vous permet de définir vos préférences de façon permanente, puisque vous ne travaillez normalement qu'avec un nombre limité de langues. Pour ce faire, appuyez sur le bouton [Éditer les langues](#).

Les choix suivants varient en fonction de la langue :

trait d'unions et nombres

Vous pouvez spécifier s'il faut considérer les traits d'unions comme des séparateurs de mots. Si vous cochez la case trait d'union, **auto-école** sera considéré comme deux mots.

Faut-il inclure les nombres dans une liste de mots comme s'ils étaient des mots ordinaires ? Si vous ne cochez pas la case, des mots comme \$300, 50,3M ou 10ème seront ignorés dans les listes de mots, mots-clés, concordances, etc. et remplacés par un #. Ils seront inclus si vous cochez la case.

caractères dans le mot

WordSmith inclut automatiquement comme symboles alphabétiques valides tous ceux déterminés par le système d'exploitation comme étant alphabétiques pour la langue choisie. En anglais, il s'agit des lettres de A à Z et d'accents courants comme é. En arabe ou en japonais, il s'agit de tous les caractères considérés comme alphabétiques par Microsoft.

Il est utile d'autoriser des caractères supplémentaires au sein d'un mot. En anglais par exemple, l'apostrophe dans **father's** doit être incluse comme caractère valide pour pouvoir traiter le

mot entier. Cependant, il se peut que vous ne souhaitiez pas accepter les apostrophes en milieu de mot en français.

Exemples :

' (seules les apostrophes sont acceptées en milieu de mot)

'% (les apostrophes et le symbole pourcent sont acceptés en milieu de mot)

'_ (les apostrophes et le trait de soulignement sont acceptés en milieu de mot)

Vous pouvez en inclure 10 différents au total.

Pour accepter `fathers'`, cochez la case *autoriser en fin de mot*. Si vous cochez cette case, tous ces symboles seront autorisés en début et en fin de mot, tant qu'ils n'apparaissent pas seuls (comme dans " ' ").

Texte brut/HTML/SGML

Vos textes peuvent être au format [texte brut](#) : c'est la valeur par défaut. S'ils sont [codés](#) en [HTML](#), [SGML](#) ou [XML](#), sélectionnez une des options. Les outils peuvent ainsi utiliser au mieux le balisage des phrases, paragraphes et en-têtes.

Format Windows, etc.

Informations sur les [jeux de caractères](#) Windows pour votre langue de travail.

début et fin d'en-tête

Il faut que les outils puissent reconnaître les débuts et fins d'en-têtes. Si votre texte est [balisé](#) avec `<h1>` et `</h1>` par exemple, saisissez `<h#>` et `</h#>`. (# représente un chiffre, ## en représente deux, etc.). Toutes vos saisies sont sensibles à la casse : `</H#>` est différent de `</h#>`. Si dans un texte en [HTML](#) vous avez tantôt `</h1>` tantôt `</H1>`, utilisez [Text Converter](#) pour rendre vos textes cohérents à ce niveau.

début et fin de section

Si ces boîtes contiennent par exemple `<div#>` et `</div>`, les outils traiteront les sections identifiées. Attention, vos saisies sont sensibles à la casse.

début et fin de phrase

Si cet espace contient le mot `auto`, les outils traiteront les phrases comme [définies](#) (point final, point d'interrogation ou d'exclamation, suivis d'une majuscule). Si votre texte est [balisé](#) avec `<s>` et `</s>`, par exemple, indiquez-le ici. Attention, vos saisies sont sensibles à la casse.

début et fin de paragraphe

Pour que les outils puissent reconnaître les paragraphes, il faut spécifier quels sont les marqueurs de début et de fin de paragraphe, par exemple deux retours à la ligne consécutifs insérés par l'auteur du document (deux fois Entrée), ou Entrée suivi de Tab. Pour ce faire, tapez `<Enter><Tab>`. Indiquez ici si votre texte est [balisé](#) par exemple avec `<p>` et `</p>`. Attention à la casse.

Souvent, vous considérez qu'il suffit de définir une fin de paragraphe et qu'ensuite tout en découle. Souvent, les textes en HTML n'effectuent pas de distinction cohérente entre début et fin de paragraphe.

Remarque : les textes parlés du BNC utilisent `</u>` au lieu de `</p>`, mais vous pouvez conserver `</p>` dans ce cas car WordSmith utilisera `</u>` à la place si le texte ne contient pas de `</p>`.

les fichiers

. permet d'afficher tous les types de textes. Vous pouvez limiter la recherche à `*.txt`.

Rubriques connexes : [Texte balisé](#), [Listes d'arrêt](#), [Sélectionner une nouvelle langue](#), [Traitement de texte Chinois etc.](#)

5.38 gestion de la fenêtre

Le [contrôleur](#) principal Oxford WordSmith Tools se situe dans le coin supérieur gauche de votre écran et prend un quart de la taille de l'écran. À l'exception de Viewer, et Concord, la fenêtre principale de chaque outil apparaît à droite, de la même taille. La fenêtre principale de chaque outil s'affiche juste en dessous des fenêtres précédentes. Les fenêtres individuelles de résultats de chaque outil sont limitées à la fenêtre principale de chaque outil, et peuvent être superposées ou mises en cascade.

Utilisez la barre des tâches ou Alt+Tab pour passer aisément d'une fenêtre à l'autre.

"Ouvrir une nouvelle fenêtre Concord" ?

Ce message s'affiche si une fenêtre de données est déjà ouverte et que vous cliquez sur *Nouveau* pour démarrer une autre concordance. Vous pouvez ouvrir le nombre de fenêtres que vous souhaitez dans chaque outil, chacune contenant différents types de données.

réduire, déplacer et redimensionner des fenêtres

Toutes les fenêtres peuvent être agrandies ou rétrécies en tirant sur les bords avec la souris. Elles peuvent être facilement déplacées en cliquant sur la barre de titre, et en faisant glisser le pointeur de la souris. Vous pouvez réduire une fenêtre : elle devient une icône que vous pouvez restaurer en cliquant dessus. Si vous l'agrandissez, elle occupe tout l'écran de l'outil concerné. Toutes ces fonctions sont des fonctions Windows standard. Réduire la fenêtre du [contrôleur](#) principal est une bonne solution lorsque vous utilisez les outils individuels.

juxtaposition et cascade

Tous les outils principaux vous montrent quelles sont les fenêtres actives. Ces dernières sont listées sous l'élément *Fenêtre* dans le menu principal. La fenêtre active est cochée. Pour ramener une autre fenêtre devant, cliquez simplement sur son nom dans la liste.

Pour réorganiser plusieurs fenêtres différentes, vous pouvez les *juxtaper* (cela affiche des fenêtres de taille identique) s'il y en a moins de 5, sinon, vous pouvez les mettre en *cascade* verticale, les unes en dessous des autres. Vous pouvez également *juxtaper* ou *mettre les outils en cascade* à partir du programme principal **Oxford WordSmith Tools**.

encombrement de l'écran


L'écran s'encombre facilement si vous avez plusieurs concordances, chacune accompagnée d'une fenêtre de schéma, de [cluster](#), de co-occurent, etc. Remarque : toutes ces fenêtres dépendent de leur fenêtre "parent", c'est-à-dire de la concordance elle-même. De la même manière, un schéma de mots-clés est un "enfant" d'une liste de mots-clés. Vous pouvez les fermer et les rappeler quand vous le souhaitez, du moment que la fenêtre "parent" reste ouverte.

Si une concordance avec co-occurents et schémas est également ouverte, je recommande de réduire la fenêtre de concordance puis de juxtaposer les fenêtres. Cela affiche les co-occurents et les schémas tout en maintenant la concordance discrète.

restaurer le dernier fichier

Une fonctionnalité utile : le dernier fichier que vous avez enregistré ou récupéré est par défaut restauré lorsque vous relancez Oxford WordSmith Tools. Pour éviter d'encombrer l'écran, il ne s'agit que d'un seul fichier. Cette fonctionnalité peut être désactivée de façon temporaire via une option de paramètres, ou de façon permanente dans le fichier `wshell.ini` (dans votre dossier `\wsmith4`).

5.39 ignorer les lignes indésirables

Pour restaurer l'ordre correct de vos données après plusieurs éditions ou marquage des lignes à supprimer, appuyez sur le bouton Ignorer ( ou Ctrl+Z). Toutes les lignes de données que vous aurez supprimées (en appuyant sur Suppr) seront supprimées de façon permanente, sauf si vous les avez restaurées (Inser).

Dans le cas d'une liste de mots, l'ensemble du fichier sera re-trié en fonction des fréquences. Toutes les entrées supprimées sont perdues à ce stade. Les entrées désignées comme étant des lemmes de mots principaux peuvent toujours être visualisés, avant ou après enregistrement. Après suppression toutefois, vous ne pouvez plus annuler les lemmes.

Rubriques connexes : [réduire les données à N entrées](#).

WordSmith Tools

Concord

Section

VI

6 Concord

6.1 but



Concord est un programme effectuant une [concordance](#) en utilisant des fichiers texte [DOS](#), [Texte brut](#), [ASCII](#) ou [ANSI](#).

Pour l'utiliser, il faut spécifier un [terme de recherche](#), que Concord doit rechercher dans tous les fichiers texte que vous avez sélectionnés. Le programme affiche ensuite les résultats de concordance et vous donne accès aux informations concernant les co-occurents du terme de recherche, les schémas de dispersion indiquant l'endroit où apparaît le terme de recherche dans chaque fichier et les analyses de clusters qui montrent les clusters de mots (syntagmes) se répétant, etc.

Le but

Une concordance vous permet de voir de nombreux exemples d'utilisations en contexte d'un mot ou d'un syntagme. On comprend mieux l'utilisation d'un mot en examinant de nombreux exemples. De plus c'est en voyant ou en entendant les mots dans leur contexte que l'on intègre leur signification. Le contexte aide à réutiliser les mots que l'on vient d'apprendre. Un dictionnaire donne la signification d'un mot mais montre peu comment l'utiliser.

Les étudiants en langues peuvent utiliser un concordancier pour étudier l'utilisation d'un mot ou syntagme ou voir quels sont les mots associés à celui qu'ils veulent utiliser. Par exemple, c'est en utilisant un concordancier que vous pouvez constater qu'un *rapport écrit* peut *décrire*, *revendiquer*, ou *montrer*, mais qu'il ne peut pas *croire* ni *vouloir* (**le rapport veut montrer que ...*).

Les enseignants de langues peuvent utiliser un concordancier pour trouver des schémas similaires afin d'aider leurs étudiants. Ils peuvent également utiliser Concord pour créer des exercices de vocabulaire en sélectionnant deux ou trois termes de recherche, puis en les [masquant](#) et en [imprimant](#) le tout.

Les chercheurs peuvent utiliser un concordancier par exemple pour parcourir une base de données d'un hôpital recensant les accidents et ainsi voir si le mot *fracture* est associé par exemple à *chute*, *glissade*, *échelle*. C'est également un moyen d'étudier les documents historiques pour rechercher toutes les références relatives à la propriété des terrains.

6.2 qu'est-ce qu'une concordance ?

Un ensemble d'exemples d'utilisation en contexte d'un mot ou d'un syntagme donné. Une concordance du mot *donner* pourrait afficher ce qui suit :

```
... ne put me donner l'heure ...  
... Rosemary, donne m'en un autre...  
... il n'avait pas beaucoup à donner...
```

Un concordancier parcourt un texte ou un groupe de textes puis affiche les concordances. Le résultat peut être enregistré, imprimé, etc.

6.3 masquage

Dans une concordance, vous pouvez masquer les termes de recherche en utilisant des astérisques. Pour ce faire, appuyez sur la barre d'espace ou sélectionnez *Affichage | Masqué*. Répétez l'opération pour les rétablir.

Le but

Une concordance masquée s'avère pratique lorsque vous voulez créer un exercice. Dans l'exemple suivant, *give* et *put* ont été masqués :

```
... could not ***** me the time ...
... Rosemary, ***** me another ...
... would not ***** much for that ...
... could not ***** up with him ...
... so you'll ***** him a present ...
... will soon ***** up smoking ...
... he should ***** it over here ...
```

Tous les blancs ont la même taille de sorte à ne pas pouvoir repérer le mot manquant en se basant sur la taille de l'espace accordé au mot masqué.

6.4 catégories

Le but

Vous pouvez classer les entrées comme vous le souhaitez, par exemple en séparant les adjectifs des substantifs, ou suivant d'autres critères.

N	Concordance	Set	ag	word #	#	s
2	or history, and art <w VMO>may be only a small part of	B		378	0	78
3	Books and catalogues <w VMO>may contain criticism; but	C		391	0	91
4	but their writers <w VMO>may think of themselves as	T		397	0	97
5	their books <w VMO>may never be identified as	Be passive		418	0	18
6	as art criticism <w VMO>may be disputable. </p> <div2	B		505	0	05
7	that her friend <w VMO>may be right. The resources	B		1,280	0	80
8	publications. There <w VMO>may, however, be questions in	B		1,376	0	76
9	it seems that she <w VMO>may have chosen the wrong	C		1,549	0	49
10	Even though critics <w VMO>may not be as learned as	B		2,007	0	07
11	artefacts which <w VMO>may or may not be categorised	Be passive		3,103	0	03

Ici, on utilise **B** où le verbe qui suit *may* est le verbe **BE**. Cependant, la distinction est faite entre **BE** en tant que verbe et **BE** utilisé comme auxiliaire dans les constructions passives. Les autres verbes sont classés en fonction de leur première lettre.

En appuyant simplement sur une lettre ou un nombre lorsque le mode [édition, saisie, enregistrement](#) est actif (comme sur la capture d'écran ci-dessus), la ligne de concordance est marquée avec cette lettre ou ce chiffre dans la colonne Classe.

Si vous souhaitez saisir quelque chose de plus long, cliquez deux fois sur la colonne Classe et vous pourrez saisir ce que vous souhaitez.

Pour corriger une erreur, appuyez sur la touche espace.

Vous pouvez [trier](#) les lignes de concordance ultérieurement en vous basant sur ces catégories affichées ici. Il vous suffit de cliquer sur l'en-tête *Classe*.

N	Concordance	Set	ag	ord #	#	s.
1	but their writers <w VMO>may think of themselves as	T		397	0	97
2	Books and catalogues <w VMO>may contain criticism; but	C		391	0	91
3	it seems that she <w VMO>may have chosen the wrong	C		1,549	0	49
4	their books <w VMO>may never be identified as	Be passive		418	0	18
5	artefacts which <w VMO>may or may not be categorised	Be passive		3,103	0	03
6	their main topics <w VMO>may be personalities or	B		371	0	71
7	or history, and art <w VMO>may be only a small part of	B		378	0	78
8	as art criticism <w VMO>may be disputable. </p> <div2	B		505	0	05
9	that her friend <w VMO>may be right. The resources	B		1,280	0	80
10	publications. There <w VMO>may, however, be questions in	B		1,376	0	76

Rubriques connexes : mode [edit v. type-in](#).

6.5 voisinage des co-occurents

Le voisinage des co-occurents représente d'une part le nombre de co-occurents que Concord trouvera à gauche et à droite du terme de recherche, et d'autre part la distance prise en compte par **KeyWords** lors de la recherche de [liens des schémas](#). La [valeur par défaut](#) est 5,5 (5 à gauche, 5 à droite) mais elle peut aller jusqu'à 25 de chaque côté.

Pour définir le voisinage de co-occurrence ainsi que d'autres paramètres de **Concord**, cliquez sur *Modifier les paramètres* puis sur *Concord* dans le menu du [contrôleur WordSmith](#).

Rubriques connexes : [Paramètres des co-occurents](#)

6.6 paramètres des co-occurents

Pour définir le voisinage de co-occurrence ainsi que d'autres paramètres de **Concord**, cliquez sur *Modifier les paramètres* puis sur *Concord* dans le menu du [contrôleur WordSmith](#).

Spécifications minimum

La longueur minimum est 1, tout comme la fréquence minimum (la valeur par défaut est 10). Ici vous pouvez indiquer à quelle fréquence il doit apparaître dans le voisinage du terme de recherche. Les mots n'apparaissant qu'une ou deux fois ont peu de chance d'être significatifs. Ainsi, spécifier la valeur 5 n'affichera que les co-occurents apparaissant au moins 5 fois dans le contexte avoisinant.

De même, vous pouvez spécifier la longueur minimale que doit avoir un co-occurent pour être enregistré. Si le mot doit comporter au moins 3 lettres, saisissez le chiffre 3.

Voisinage

C'est ici que vous pouvez spécifier le nombre de mots à inclure dans la recherche de collocation : c'est la taille du "voisinage", mentionnée ci-dessus. La valeur maximale est 25 à gauche et 25 à droite. Les résultats affichent ensuite ces mots dans des colonnes différentes pour que vous puissiez voir exactement le nombre de fois où un co-occurent apparaît par exemple 3 mots à gauche de votre terme de recherche.

Les plus fréquents sont signalés en **rouge**.

Séparateurs

Les différents séparateurs sont :

```
no limits
stop at punctuation
stop at sentence break
stop at paragraph break
stop at heading break
stop at section break
stop at end of text
```

Ils sont disponibles dans le coin supérieur droit de la fenêtre des [paramètres du contrôleur Concord](#).

Lors du calcul des co-occurents, si le paramètre est défini sur "s'arrêter en fin de phrase", les co-occurents seront comptés dans le voisinage mentionné précédemment, mais les fins de phrase seront prises en compte.

Supposons qu'une ligne de concordance comporte ce qui suit :

```
source, per pointing integration times, respectively. However, when we
compared these two maps
```

Supposons également que le terme de recherche est **however**, seule l'expression

```
when we compared these two
```

sera considérée comme co-occurent car il y a un marqueur de fin de phrase à gauche du terme de recherche. Si le paramètre était défini sur "s'arrêter à la ponctuation", il n'y aurait rien dans la liste des co-occurents pour cette ligne car il y a un marqueur majeur d'arrêt de phrase à gauche du mot "however" et qu'il n'y a aucun élément à sa droite situé entre lui et le signe de ponctuation suivant.

6.7 affichage des co-occurents

Le but

Il s'agit de repérer les schémas lexicaux caractéristiques. Il est parfois difficile de voir les tendances dans vos lignes de concordance, surtout lorsque ces dernières sont nombreuses. Étudier les collocations de cette manière vous permet de repérer les schémas lexicaux et grammaticaux de co-occurrence. Les liens entre les co-occurrences incluant des éléments grammaticaux sont souvent nommés *colligations*.

Affichage

L'affichage de collocation se fait initialement par ordre de fréquence.

Le [degré de relation](#) entre les deux (ou 0,000 si le calcul n'a pas encore été effectué) s'affiche à côté de chaque mot et de chaque terme de recherche sur lesquels repose la concordance. S'afficheront ensuite le nombre de fois où les co-occurrences apparaissent avec le terme de recherche dans votre concordance, ainsi qu'un total pour la partie gauche et la partie droite du terme de recherche. Vous verrez ensuite des informations détaillées concernant la décomposition : combien de fois la co-occurrence apparaît dans une zone de 5 mots à gauche, ensuite dans une zone de 4 mots à gauche, et ainsi de suite jusqu'à 5 mots à droite. La position centrale (où apparaît le terme de recherche) est marquée d'un astérisque.

Le nombre de mots à gauche et à droite dépend du [voisinage de co-occurrence](#).

Ces nombres sont :

- le nombre total de fois où le mot a été trouvé dans le voisinage du terme de recherche
- le nombre total de fois où il est apparu à gauche du terme de recherche
- le nombre total de fois où il est apparu à droite du terme de recherche

un ensemble de fréquences distinctes d'apparition à gauche du terme de recherche (5L, qui signifie 5 mots à gauche, 4L... 1L)
 une colonne centrale représentant le terme de recherche
 un ensemble de fréquences distinctes d'apparition à droite du terme de recherche (1R, 2R, etc.)

Le nombre de colonnes dépend du voisinage du terme de collocation. Avec 5,5 vous obtenez 5 colonnes à gauche et 5 colonnes à droite du terme de recherche. Ainsi, vous pouvez voir précisément combien de fois chaque mot a été trouvé dans le voisinage du terme de recherche, et, par exemple, combien de fois il a été trouvé 1 mot à gauche ou 4 mots à droite. Les plus fréquents sont signalés en **rouge**. Sur la capture d'écran ci-dessous, **differences** apparaît 44 fois au total, mais seulement **39** sont en position L1.

N	Word	With	Relation	Total	Left	Right	L5	L4	L3	L2	L1	entre	R1
1	AMONG	among	12.133	621	0	0	0	0	0	0	0	621	0
2	BETWEEN	between	10.436	2,109	14	14	7	4	2	1	0	2,081	0
3	DIFFERENTIATE	between	10.151	5	5	0	0	0	0	0	5	0	0
4	SIMILARITIES	between	9.360	13	13	0	0	0	2	0	11	0	0
5	DISTINCTIONS	between	9.092	6	6	0	1	1	1	0	3	0	0
6	DISTINCTION	between	9.008	20	20	0	0	3	1	1	15	0	0
7	TENSIONS	between	8.955	8	8	0	0	1	0	1	6	0	0
8	DIFFERENCES	between	8.943	44	43	1	1	1	0	2	39	0	0
9	DISTINGUISH	between	8.919	11	11	0	0	0	0	1	10	0	0
10	TORN	between	8.884	9	9	0	0	0	0	0	9	0	0
11	DIFFERENCE	between	8.714	96	96	0	2	1	3	3	87	0	0
12	OVERLAP	between	8.677	6	6	0	0	0	0	0	6	0	0

La capture d'écran ci-dessus affiche les résultats de collocation pour une concordance de **BETWEEN/AMONG** triée en fonction de la colonne *Relation*, où les mots comme **differentiate**, **difference** apparaissent comme étant fortement liés à **between**. Quelques liens concernant **among** (**growing**, **refugees**) apparaissent plus bas dans la liste.

N	Word	With	Relation	Total	Left	Right	L5
25	GROWING	among	7.859	12	9	3	1
26	CLASHES	between	7.829	5	5	0	0
27	REFUGEES	among	7.808	5	0	5	0
28	TENSION	between	7.770	12	12	0	1

Les fréquences affichées peuvent être **re-triées** (🔍) et vous pouvez recalculer les co-occurents (🔄) si vous **ignorez** certaines entrées de la concordance ou que vous modifiez le **voisinage**.

Rubriques connexes : [Collocation](#), [Relation de collocation](#), [Informations mutuelles](#)

6.8 relation de collocation

Le but

Il s'agit ici de trouver le degré de relation entre chaque co-occurent et le terme de recherche près duquel il apparaît.

Sélectionnez une liste de mots

L'affichage des collocations présente initialement les co-occurents sans aucune information de relation (exemple 0.00), car pour calculer la relation, il est nécessaire de connaître certains éléments :

- a) la fréquence à laquelle chaque co-occurent apparaît dans le corpus utilisé
- b) la fréquence à laquelle le [terme de recherche](#) apparaît dans le corpus
- c) la fréquence à laquelle ils apparaissent ensemble dans le [voisinage](#) sélectionné.

Cependant, bien que les paramètres b) et c) soient connus au moment où la concordance est effectuée, a) n'est pas connu tant qu'une concordance ou une liste de mots n'a pas été effectuée pour chaque co-occurent.

C'est pourquoi vous devez sélectionner une liste de mot adéquate (créée par WordSmith bien sûr!) qui comporte les fréquences pour chaque mot. Il vous appartient de sélectionner une liste de mots effectivement liée à la concordance que vous avez effectuée.

Type de relation

Dans la fenêtre principale des [paramètres du contrôleur Concord](#), sélectionnez le type de relation que vous souhaitez calculer.

Rubriques connexes : [Collocation](#), [Affichage des co-occurents](#), [Informations mutuelles](#)

6.9 collocation

Qu'est-ce qu'un co-occurent ?

Les co-occurents sont des mots apparaissant dans le voisinage du terme de recherche. Les co-occurents de *lettre* peuvent inclure *poste*, *timbre*, *enveloppe*, etc. Cependant, des mots très courants tels que *la* seront également des co-occurents de *lettre*.

Le but

En étudiant les co-occurents, vous pouvez en apprendre plus sur les mots "accompagnant" le terme de recherche, ce qui permet de voir sa signification et son utilisation.

Options

Vous pouvez effectuer une concordance avec ou sans co-occurents : l'option sans co-occurents permet d'accélérer légèrement le processus et prend moins d'espace disque. Par défaut, les concordances s'effectuent avec les co-occurents.

Le nombre de co-occurents stockés dépend du [voisinage de co-occurrence](#).

Vous pouvez re-calculer les co-occurents après avoir édité votre concordance.

Si vous souhaitez filtrer votre liste de co-occurents, utilisez une [liste de correspondance](#) ou [une liste d'arrêt](#).

[Re-trier](#) une liste de co-occurents de diverses manières.

Vous pouvez voir le [degré de relation](#) entre un mot et le terme de recherche sur lequel la concordance est basée.

Les co-occurents peuvent être [affichés](#) une fois la concordance effectuée.

Note technique

Les [écrits](#) sur la collocation n'ont jamais fait de distinction très nette entre les co-occurents que nous considérons comme "associés" d'un mot (lettre – timbre) et les mots qui sont effectivement des co-occurents du mot (lettre - mon, ce, un, etc.).

On pourrait considérer que les premiers sont des "co-occurents logiques" et que les seconds sont des "co-occurents de voisinage". Il est très difficile de détecter les co-occurents logiques, car dès que l'on commence à regarder au-delà de 4 ou 5 mots de chaque côté du terme de recherche, on obtient tellement de mots qu'il devient difficile de s'y retrouver.

KeyWords vous permet d'examiner les [associés](#), qui sont des indices révélateurs de co-occurents logiques. **Concord** permet de rechercher des "co-occurents de voisinage".

WordList vous permet d'étudier les [informations mutuelles](#).

Rubriques connexes : [Affichage des collocations](#), [Paramètres de collocation](#), [Relation de collocation](#), [Affichage des informations mutuelles](#).

6.10 Concord : clusters

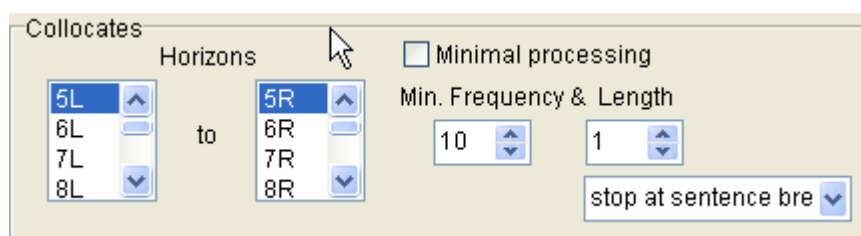
Le but

Ces clusters de mots vous permettent de voir les schémas de phraséologie qui se répètent dans votre concordance, surtout si votre concordance comporte plusieurs milliers de lignes. Ils comportent en général le terme de recherche, puisqu'ils sont basés sur des lignes de concordance.

Une autre fonctionnalité de **Concord**, [Schémas](#), vous permet de voir les schémas existants.

Fonctionnement

Les clusters sont automatiquement calculés si cette option n'est pas désactivée dans la fenêtre des paramètres du [contrôleur](#) de Concord (*Modifier les paramètres | Concord*) où s'affichera une fenêtre semblable à celle-ci :



où les paramètres par défaut peuvent être modifiés. Si l'option "Traitement minimal" est cochée, les co-occurents, les clusters, les schémas, etc. ne seront pas calculés lors de la concordance. (Il est toujours possible de les calculer ultérieurement si les fichiers texte source sont toujours présents).

La recherche de clusters s'effectue dans cette limite : la valeur par défaut est 5 mots à gauche et 5 mots à droite du terme de recherche, mais ces valeurs peuvent aller jusqu'à 25. Par défaut, les clusters comportent trois mots, et vous pouvez décider combien d'entre eux doivent être trouvés pour que les résultats soient affichés (on peut dire par exemple que 3 est la fréquence minimale).

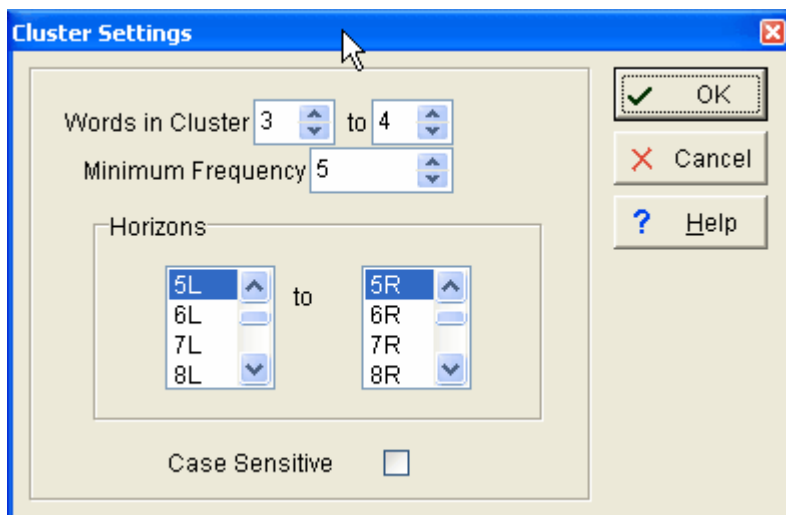
Les clusters sont calculés en se basant sur les lignes de concordance existantes. Cela signifie que toute ligne n'ayant pas été supprimée ou ignorée est utilisée pour le calcul des clusters.

Tout comme pour les [clusters d'index de WordList](#), le principe consistant à "s'arrêter en fin de

phrase" (d'autres alternatives sont possibles) est qu'un cluster qui s'étend sur deux phrases n'a pas de sens.

Re-calculer des clusters (⋮)

Les clusters calculés par défaut peuvent ne pas convenir, (vous pouvez décider de refaire le calcul après avoir supprimé certaines lignes), donc vous pouvez également sélectionner *Calculer / Clusters* (⋮) dans le menu Concord pour sélectionner le nombre de mots qu'un cluster doit contenir (2 à 4 est la taille recommandée) et de modifier les autres paramètres.



Les clusters sont calculés lorsque vous cliquez sur OK. Dans le cas présent, on a demandé des clusters comportant 3 ou 4 mots. Les résultats sont les suivants :



La plupart des clusters comportent 3 mots, mais celui mis en surbrillance en comporte 4. La ligne 10 montre qu'il y a 6 occurrences de *if this is*, et la ligne 12 montrent que la majorité de ces occurrences sont *if this is a*.

Une fenêtre dépendante

Chaque fenêtre Cluster est dépendante de la concordance à partir de laquelle elle a été générée. Si vous fermez la concordance d'origine, elles disparaissent.

Rubriques connexes : [informations générales sur les clusters](#), [clusters de WordList](#).

6.11 Concord : dispersion

Le but

Ceci montre où le terme de recherche apparaît dans le fichier auquel l'entrée actuelle appartient. Vous pouvez ainsi voir où apparaît le plus souvent votre terme de recherche dans chaque fichier.

Ce que vous voyez

Le schéma affiche ceci :

Fichier nom du fichier texte source

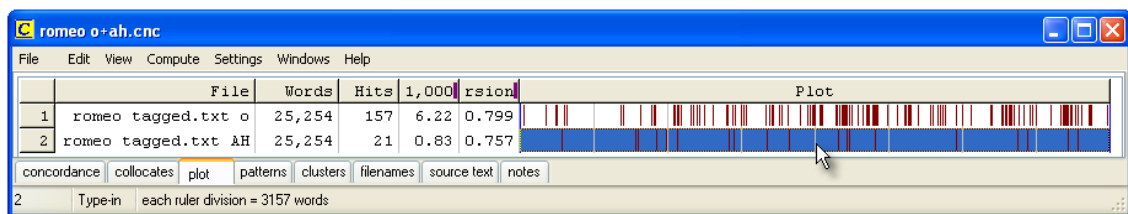
Mots nombre de mots dans le texte source

Occurrences nombre d'occurrences du terme de recherche

par 1 000 nombre d'occurrences par tranche de 1 000 mots

Dispersion la [valeur de dispersion](#) du schéma

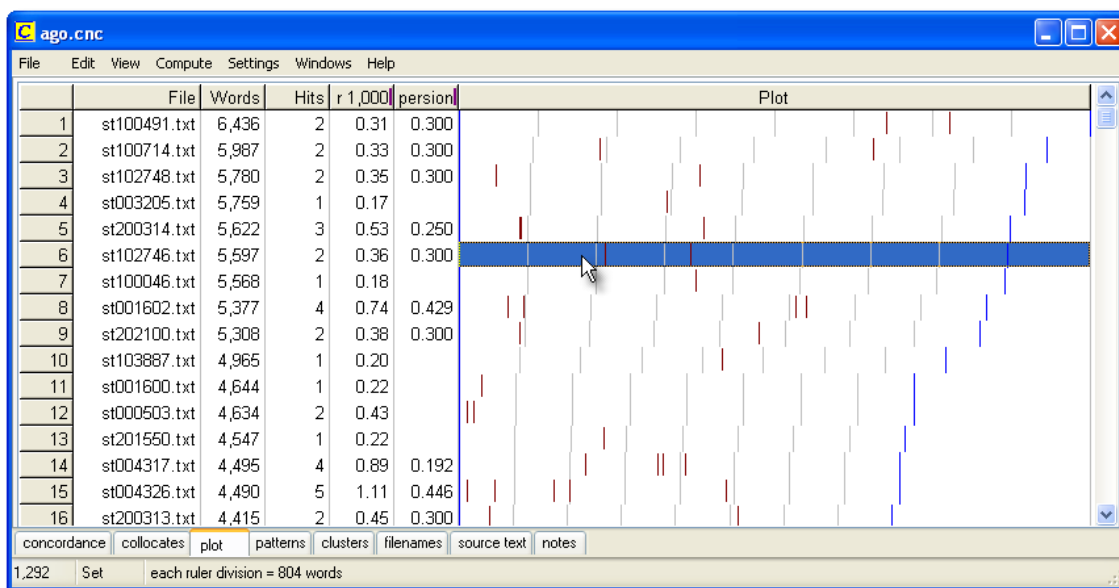
Schéma schéma montrant où les occurrences apparaissent. Le coin gauche du schéma représente "Once upon a time" et le coin droit "happily ever after".



Ici, on peut voir un schéma de "O" et un autre de "AH" de la pièce Roméo et Juliette. Les deux sont présentés sur deux lignes différentes car il y a deux termes de recherche. Il y a plus d'exclamations en "O" qu'en "AH"s. Une règle sépare le schéma affiché en 8 segments. La barre d'état indique que chaque segment représente environ 3 150 mots de la pièce.

Le schéma est initialement [trié](#) par nombre de mots par tranche de 1 000.

Il y a deux façons d'afficher le schéma : la première, par défaut, où tous les rectangles ont la même longueur, ou la façon *Schéma uniforme* où les rectangles reflètent la taille du fichier original (le rectangle le plus grand correspond au plus gros fichier). Il est possible de procéder à des modifications via le menu *Affichage* situé dans la partie supérieure.



La capture d'écran représente un schéma uniforme. Comme l'indique la barre d'état, chaque repère délimite des segments contenant 800 mots dans ce schéma de dispersion de "ago". En regardant la colonne Mots, vous voyez que le nombre de mots dans chaque fichier varie, ce qui explique pourquoi la position du repère bleu situé à l'extrême droite ainsi que des autres repères varie.

Si le nombre de marques est inférieur au nombre d'occurrences, c'est que les occurrences sont trop proches les unes des autres par rapport à la résolution de votre écran. Faites glisser le bord droit du schéma pour l'étendre. Vous pouvez exporter le schéma avec l'option [Enregistrer sous](#) créer des graphiques dans une feuille de calcul comme expliqué [ici](#).

Chaque fenêtre de schéma se rapporte à la concordance à partir de laquelle elle a été générée. Si vous fermez la concordance d'origine, elle disparaît. Vous pouvez *imprimer* le schéma. Il n'y a pas d'option *Enregistrer* car les données proviennent d'une concordance que vous devez [Enregistrer](#), ou *Imprimer dans un fichier*. Vous pouvez *Copier* les éléments dans le [Presse-papiers](#) (Ctrl+Insert) et les insérer en tant que graphiques dans un logiciel de traitement de texte en utilisant l'option Collage spécial.

Rubriques connexes : [couleur du schéma et des repères](#), [valeur de dispersion du schéma](#).

6.12 Concord : enregistrement et impression

Vous pouvez enregistrer la concordance (ainsi que les co-occurents et autres résultats qui s'y rapportent s'ils ont été gardés en mémoire lorsque la concordance a été générée) soit en tant que fichier texte (pour l'importer dans un logiciel de traitement de texte par exemple), soit en tant que fichier de résultats que vous pouvez *ouvrir* ultérieurement à partir du menu principal qui se trouve en haut. En quittant **Concord**, vous pouvez enregistrer votre travail si ce n'est pas déjà fait.

Enregistrer une concordance vous permet d'y revenir ultérieurement et de revoir les co-occurents, les schémas de dispersion et les clusters.

Vous pouvez [imprimer](#) en utilisant l'imprimante Windows de votre système. Vous pouvez spécifier le nombre de pages à imprimer. La police de caractères ressemblera à peu de choses près à ce qui s'affiche sur votre écran. Si vous utilisez une imprimante couleur ou à niveaux de

gris, les couleurs de l'écran seront transmises à votre imprimante. Si vous avez une imprimante noir et blanc, les éléments colorés sortiront en *italique* si votre imprimante le permet.

Concord imprime autant de concordances avec les détails associés que le permettent vos paramètres d'impression, les marges pouvant être visualisées en cliquant sur [Aperçu avant impression](#).

6.13 Concord : affichage des options

Ces options peuvent être activées ou désactivées à n'importe quel moment. Elles sont cochées lorsqu'elles sont activées. Elles comprennent :

Phrase uniquement

Cela n'affiche que la phrase dans laquelle le terme de recherche apparaît.

Balises et espaces masqués

Si vous avez spécifié des [balises à conserver](#), ces dernières seront normalement visibles dans votre concordance. Si vous souhaitez les masquer, activez cette option. Cette option permet également de supprimer tout espace superflu dans la ligne de concordance, qui peut venir de [balises](#) ayant été ignorées.

Rubriques connexes : [affichage des balises les plus proches](#), [Masquer](#) le terme de recherche.

6.14 Concord : gestion des fichiers son et vidéo

Le but

Supposons que vous effectuiez une concordance du mot "éléphant" et que vous souhaitez écouter la prononciation de ce mot en contexte. Est-ce que la dernière voyelle est un schwa ? La seconde voyelle se prononce-t-elle comme un "e" ou un "é" ?

Procédure...

Si vous avez défini des balises qui se réfèrent à des fichiers multimédia et si de telles balises se trouvent dans le contexte de balises d'une ligne de concordance donnée, vous entendez ou voyez la source multimédia. Pour lire le fichier multimédia, appuyez sur Ctrl+M ou cliquez sur *Fichier | Exécuter le fichier multimédia*. La balise sera [analysée](#) pour identifier le fichier nécessaire, ou le télécharger si besoin est, puis le jouer.

Rubriques connexes : [Gestion des balises](#), [Création d'un fichier de balises](#), [Affichage des balises les plus proches dans Concord](#), [Concordance des balises](#), [Types de balises](#), [Affichage des balises](#), [Utilisation des balises comme sélecteurs de texte](#), [Balises dans WordList](#)

6.15 Concord : ce que vous pouvez voir et faire

Une fenêtre contenant la liste de toutes les concordances s'affiche. Vous pouvez faire défiler la liste vers le haut, le bas, la droite ou la gauche avec la souris ou les touches directionnelles. Pour de plus amples informations, cliquez sur [menu](#).

Les colonnes :

Elles affichent les détails pour chaque entrée : numéro de l'entrée, ligne de concordance, classe, balise, position du mot (par exemple, le premier mot du texte est 1), [nom de fichier](#) du texte source, ainsi que le pourcentage de texte dans lequel il apparaît. Vous trouverez ci-

dessous une explication concernant les **marques violettes**.

Classe

C'est ici que vous pouvez classer vous-même les entrées, en utilisant n'importe quelle lettre, dans les [catégories définies par l'utilisateur](#). Supposons que vous souhaitez trier les utilisations des verbes de celles des noms. Il vous suffit d'appuyer sur V ou N. Pour saisir un élément plus long, (par exemple "Nom"), cliquez deux fois sur l'entrée de la colonne "classe" et saisissez ce que vous souhaitez. Si vous avez plusieurs [termes de recherche](#), vous constaterez que la colonne Classe affiche le terme de recherche pour chaque entrée. Pour effacer l'entrée actuelle, vous pouvez saisir le nombre 0. Pour effacer l'intégralité de la colonne Classe, cliquez sur *Edition | Vider la colonne Classe*.

Balise



Cette colonne affiche le [contexte de balises](#).

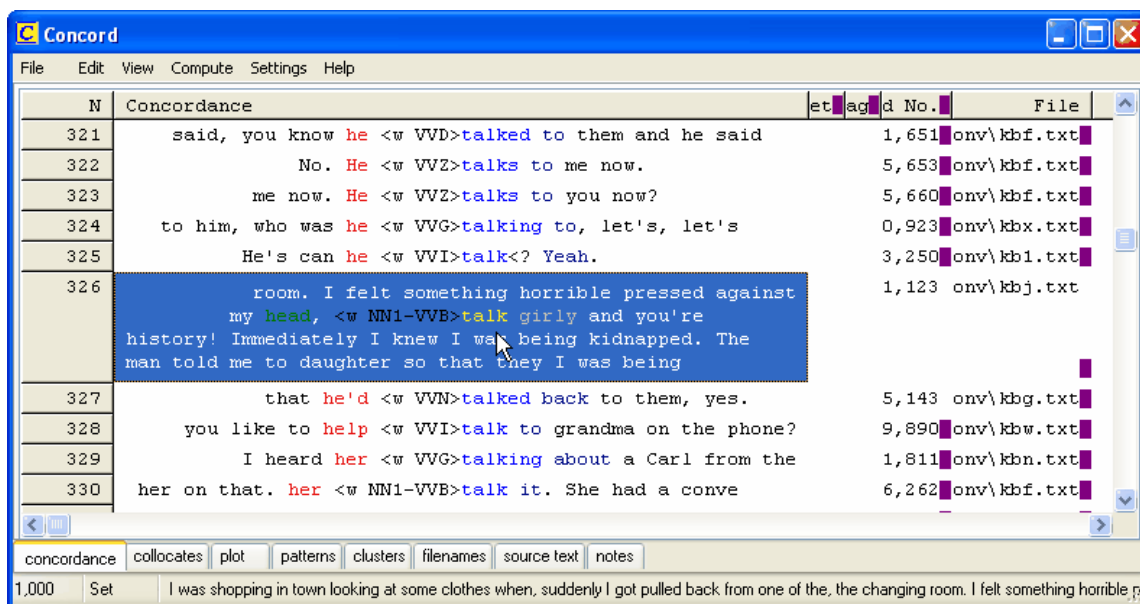
Étendre l'affichage

Étendre l'affichage de concordance pour élargir la colonne. Placez le curseur de la souris sur la barre située entre deux colonnes. Lorsque le curseur change de forme, vous pouvez étirer toute la colonne.

Étendre une ligne pour afficher plus de contexte

Faites de même pour chaque ligne : placez le curseur de la souris entre deux lignes dans la zone grise comportant un nombre, et faites glisser.

Vous pouvez également cliquer sur  ou appuyer sur F8 pour agrandir toutes les lignes. Pour les réduire, cliquez sur  ou appuyez sur Ctrl+F8. La touche 8 du pavé numérique permet d'agrandir la ligne sélectionnée, comme le montre la capture d'écran ci-dessous. Pour la réduire, appuyez sur la touche 2.



Marques en violet

Sur la capture d'écran, vous pouvez voir les marques violettes présentes lorsque la colonne n'est pas assez large pour afficher toutes les données. Très souvent, les nombres ne sont pas entièrement visibles, et le résultat que l'on voit est erroné. Par exemple, dans la concordance ci-dessous, la colonne *Nb de mots* affiche 4 569 mais le nombre réel pourrait être 14 569. Étirez la colonne pour faire disparaître les marques violettes.

	et	ag	ord	#	#	s.	#
brave new world of deregulated				4,569	0	11	0
brave soul who confronts a party				4,235	0	49	0
brave step, even though he is				4,295	0	10	0

Affichage du fichier original

Cette fonction n'est disponible que si le fichier se trouve encore au même emplacement que lorsque la concordance a été effectuée.

Barre d'état

Le panneau de [barre d'état](#) affiche ceci :

- le nombre d'entrées (1 000 sur la capture d'écran)
- si on est en mode Définition ou Édition
- la ligne de concordance active depuis son début.

Rubriques connexes :

[Re-trier](#) vos lignes de concordance

[Catégories définies par l'utilisateur](#)

[Modifier l'affichage](#)

[Masquer](#) le terme de recherche

[Collocation](#) (mots dans le voisinage du terme de recherche)

[Schéma](#) (montre, sous forme de graphique, à quel endroit le terme de recherche apparaît dans le texte)

[Clusters](#) (groupes de mots dans votre concordance)

[Segments de texte dans Concord](#)

[Modifications de la concordance](#)

[Ignorer des entrées](#)

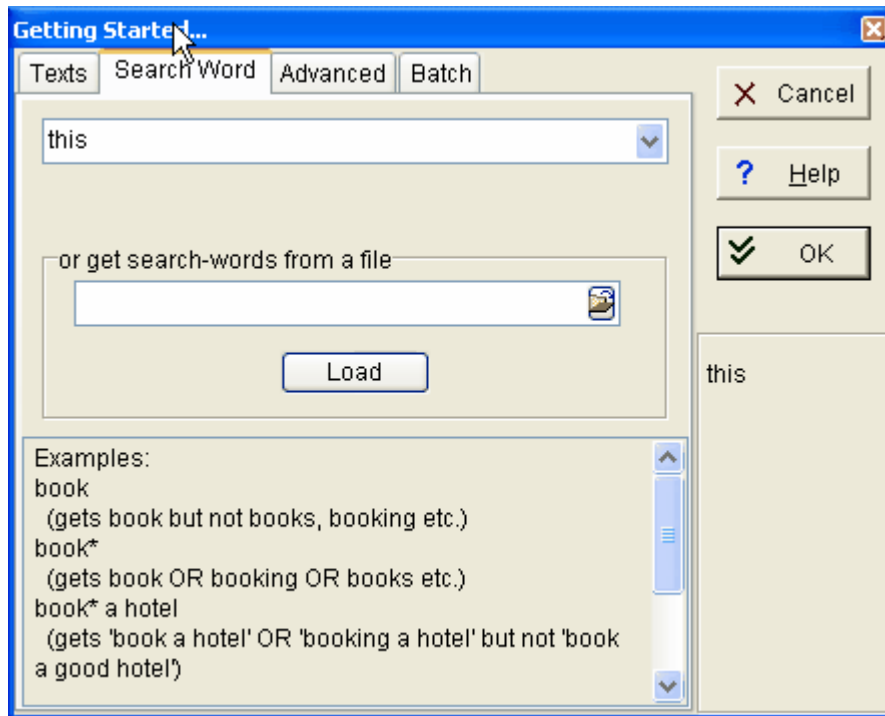
[Enregistrement et impression](#)

[Gestion de la fenêtre](#)

6.16 paramètres de concordance

Terme, syntagme ou balise de recherche

Saisissez le [mot ou le syntagme](#) que Concord doit rechercher lors de la concordance, ou saisissez le nom d'un [fichier de termes de recherche](#). Vous pouvez également effectuer une sélection à partir d'une [liste d'historique](#) de vos précédents termes de recherche. Pour de plus amples informations sur la syntaxe, reportez-vous à la rubrique [Syntaxe des termes de recherche](#) ou à l'ensemble d'exemples montrés dans la capture d'écran ci-dessous :

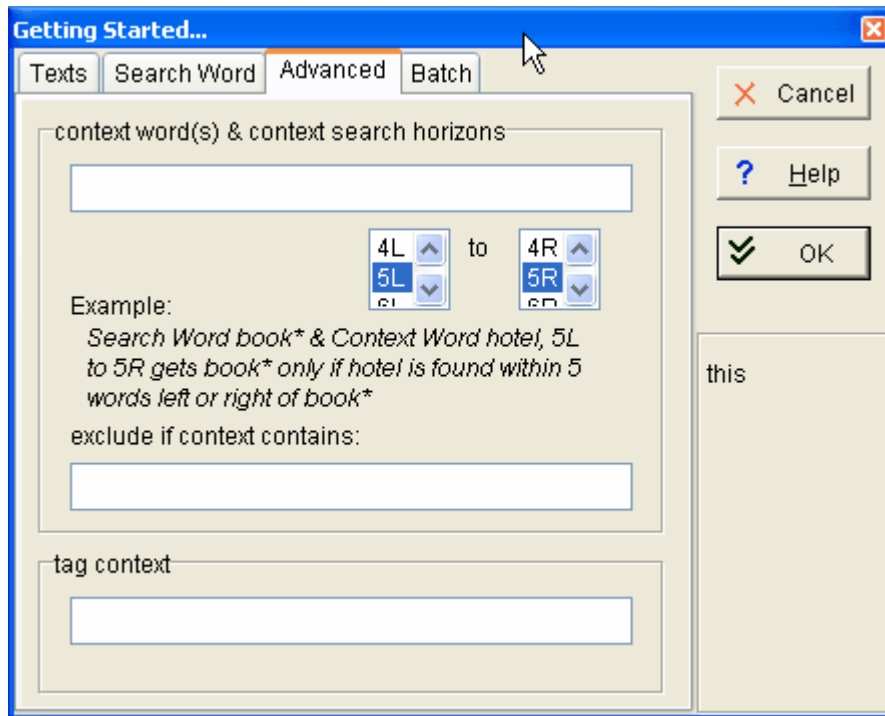


Si vous souhaitez effectuer de nombreuses concordances lors d'une [recherche basée fichiers](#), préparez tout d'abord un petit fichier texte comportant les termes de recherche, par exemple :

```
this
that
the other
==Major*==
```

Cliquez d'abord sur le bouton fichier pour localiser votre fichier texte, puis que le bouton *Charger*. Le nom changera et deviendra un nom du style *Clear 4*, où 4 signifie qu'il y a 4 termes de recherche différents à prendre en compte lors de la concordance. Reportez-vous à la rubrique Batch ci-dessous pour obtenir plus d'informations sur l'enregistrement de chaque recherche dans un fichier distinct. Dans le cas contraire, toutes les recherches seront combinées dans la même concordance.

Avancés



Mots de contexte et voisinage du contexte de recherche

Vous pouvez être amené à rechercher un mot ou un syntagme, en fonction du contexte. Dans ce cas, vous pouvez spécifier les termes de contexte que vous souhaitez ou que vous ne souhaitez pas (cela signifie que si le mot est trouvé, cette entrée ne sera pas utilisée). Si par exemple le terme de recherche est **book*** et le *contexte* word is **hotel**, vous obtiendrez **book**, **books**, **booked**, **booking**, **bookable**, mais seulement si **hotel** se trouve dans [le voisinage du contexte de recherche](#). Si le terme de recherche est **book*** et le mot d'*exclusion* est **hotel**, les résultats seront **book**, **books**, **booked**, **booking**, **bookable**, tant que **hotel** ne se situe pas dans le voisinage du contexte de recherche. Si le mot de recherche est **book*** et que le terme d'exclusion est **booked**, les résultats seront **book**, **books**, **booking**, **bookable**, mais pas **booked**.

Contexte de balises

Le contexte de balises est défini dans votre [fichier de balises](#), à gauche du terme de recherche. Dans cet exemple, le contexte de balises est `<u speaker=Simon>`:

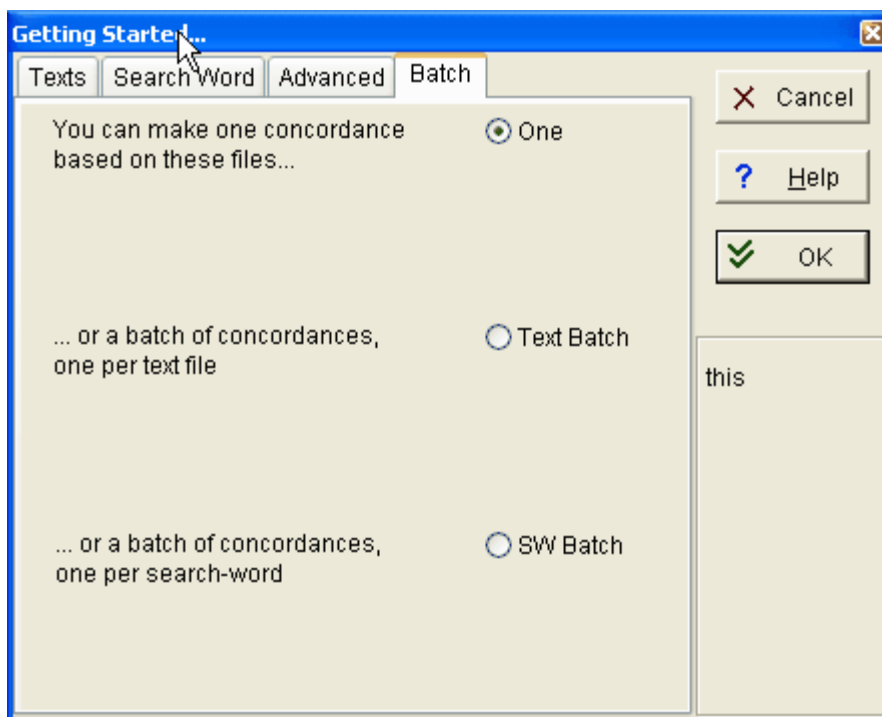
```
.....<u speaker=Simon> Between you and me, I wish I hadn't booked
that hotel ....
```

Le contexte de balises peut en fait inclure plus d'éléments – si cette partie se trouve dans `<section 1>`, et qu'une balise marque la fin de la section 1 (à priori `</section>`) apparaissant à n'importe quel endroit à droite du terme de recherche, alors le contexte de balise complet serait `<section 1><u speaker=Simon>` voire quelque chose de plus développé. Ainsi, le contexte de balises désigne en fait tout l'ensemble de balises utilisées à l'endroit où apparaît votre terme de recherche.

Vous pouvez être amené à rechercher un mot ou un syntagme en fonction du contexte de balises. Dans ce cas, spécifiez les attribut de balises que vous souhaitez. Par exemple, si le terme de recherche est **book*** et que le contexte de balise est **simon**, vous obtiendrez **book**, **books**, **booking**, **bookable**, **booked** seulement si **simon** apparaît dans le contexte de balises.

Lot

Supposons que vous effectuiez une concordance du mot **book*** au sein de 20 fichiers : vous pouvez effectuer *une* concordance basée sur les 20 fichiers (par défaut) ou 20 concordances séparées [dans un lot zippé](#) pouvant être affichées séparément (*Lot de textes*). Si vous avez des mots de recherche multiples dans une [recherche à partir de fichiers](#) comme expliqué ci-dessus, il peut être utile d'enregistrer chaque résultat séparément (*Lot par mot-clé*).



Il existe d'autres paramètres pour les concordances : voir [Paramètres de concordance du contrôleur WordSmith](#); [Caractères accentués](#); [Sélectionner une langue](#), [Termes de contexte et voisinage du contexte de recherche](#)

6.17 effectuer une concordance sur des balises

Le but

Supposons que vous cherchez à identifier certains types de structures (et non pas un mot ou syntagme donné), par exemple une séquence comme celle-ci : **Nom+Nom+Nom**. Vous pouvez saisir les balises sur lesquelles vous souhaitez effectuer la concordance (avec ou sans mots).

Procédure

Saisissez les balises souhaitées dans la zone de terme de recherche de Concord. Vous pouvez également définir ces dernières dans un [fichier de balises](#).

Exemples

`<w NN1>table` renvoie le nom anglais *table* au singulier (et non pas le verbe)

`<w NN1>* <w NN1>*` renvoie toute séquence de deux noms communs au singulier apparaissant dans l'échantillon du BNC.

Remarque : `<w NN1>table` renvoie *table* si votre texte est balisé avec les symboles `<` et `>`. Si

vous avez spécifié que [et] sont des symboles de balises, le résultat renvoyé sera [w NN1] **table**.

Vous trouverez plus d'exemples dans la section [Terme ou syntagme de recherche](#).

Que vous utilisiez un [fichier de balises](#) ou non n'a aucune importance car WordSmith identifie automatiquement vos balises. Bien évidemment, vous devez utiliser des textes précédemment balisés pour que la fonction puisse être appliquée.

Dans le deuxième exemple, les astérisques sont présents car dans le BNC, les balises se trouvent juste avant le mot auquel elles se réfèrent. Si vous oubliez l'astérisque, Concord traitera la commande comme si vous souhaitiez une balise avec un [séparateur](#) de chaque côté.

Effectuez-vous une concordance sur des balises ?

Si ce message apparaît et que vous effectuez une concordance sur des balises, répondez "Oui" à la question. Sinon, des guillemets " " seront insérés autour de chaque symbole < ou > que votre mot contient, comme l'explique la section [Syntaxe du terme de recherche](#).

Sensibilité à la casse

Les balises ne sont sensibles à la casse que si le terme de recherche ou le syntagme l'est. Par défaut, les termes de recherche ne le sont pas. Ainsi dans le premier exemple, les résultats renvoyés seraient *table*, *Table* et *TABLE* s'ils sont utilisés comme des noms. Aucun résultat n'aurait été renvoyé s'il n'y avait aucune balise dans vos textes source.

Rubriques connexes : [Présentation générale des balises](#), [Gestion des balises](#), [Affichage des balises les plus proches dans Concord](#), [Terme ou syntagme de recherche](#), [Types de balises](#), [Affichage des balises](#), [Utilisation des balises comme sélecteurs de texte](#)

6.18 terme de contexte

Vous pouvez limiter une recherche de concordance en spécifiant un terme de contexte qui doit ne doit pas être présent dans un certain "rayon" de mots à partir du terme de recherche.

Par exemple, **book** peut être le terme de recherche et **hotel*** le terme de contexte. Seul **book** sera trouvé si **hotel** ou **hotels** se trouvent dans le voisinage.

Par exemple, **book** peut être le terme de recherche et **paper*** le critère d'exclusion. Seul **book** sera trouvé si **paper** ou **papers** ne se trouvent pas dans le voisinage du contexte de recherche.

Voisinage du contexte de recherche

Le voisinage de contexte détermine la distance à gauche et à droite du terme de recherche que Concord doit prendre en compte lorsqu'il vérifie si les critères de recherche correspondent. La [valeur par défaut](#) est 5,5 (5 mots à gauche et 5 mots à droite du terme de recherche) mais elle peut aller jusqu'à 25 de chaque côté. En saisissant la valeur 0,2 la recherche ne s'effectuerait qu'à droite et dans un rayon de deux mots du terme de recherche.

Si vous avez spécifié un terme de contexte, vous pouvez re-trier les résultats en vous basant dessus. De plus, les termes de contexte seront affichés dans une [couleur](#) qui leur sera propre.

La syntaxe est semblable à celle du [terme de recherche ou syntagme](#),

* signifie que la fin du mot sera ignorée lors de la recherche. Ce signe peut être placé au début ou à la fin du terme de contexte.


== signifie que la casse sera respectée

/ sépare les alternatives. Vous pouvez spécifier jusqu'à 15 alternatives, dans une limite totale de 80 caractères.

Si vous souhaitez utiliser les caractères *, ? , == , ~ , :\ ou / comme caractères appartenant à votre terme de recherche, placez les entre guillemets comme ceci : "*"

6.19 modifier des concordances

Le but

Certaines entrées peuvent ne pas correspondre à ce que vous attendiez. Supposons que vous ayez effectué une recherche avec **SHRIMP*/PRAWN*** comme termes de recherche – le terme *Shrimpton* apparaît alors dans la liste de résultats. Pour épurer une liste, il suffit d'appuyer sur **Suppr** pour chaque ligne indésirable. (Effectuez d'abord un tri sur le terme de recherche de manière à grouper tous les *Shrimptons* les uns à côté des autres). La ligne devient alors grise. Il est possible de la restaurer en appuyant sur **Ins**. Pour supprimer ou restaurer TOUTES les lignes à partir de la ligne courante jusqu'en bas, cliquez sur la clé Fn configurée. Lorsque vous avez terminé, vous pouvez appuyer sur (**Alt+Z** ou ) pour [éliminer](#) les lignes supprimées. Pour un enseignant, [masquer](#) les termes de recherche peut s'avérer intéressant : il suffit d'appuyer sur la barre d'espace. En appuyant à nouveau sur la barre d'espace, l'apparence des termes est restaurée.

Rubriques connexes : [Gestion de la fenêtre](#)

6.20 termes de recherche issus du fichier

Le but

Cette fonction permet de ne pas perdre de temps à saisir des recherches complexes. Vous pouvez être amené à effectuer une recherche standard de manière répétée sur différents sous-corpus.

Or, comme Concord admet un nombre d'entrées quasi illimité, il est possible d'effectuer une concordance avec de nombreux [termes de recherche ou syntagmes](#).

Le nombre de caractères pour les termes de recherche multiples est limité à 80 (y compris / etc.). Si vos termes de recherche préférés dépassent cette limite ou si vous souhaitez utiliser une recherche standard, vous pouvez préparer un fichier contenant tous les termes de recherche.

Procédure

Un exemple (`\wsmith4\concordance_search_words.txt`) est inclus aux fichiers de distribution.

Utilisez un éditeur Windows (par exemple Bloc-notes) pour préparer le vôtre. Chacun des termes doit se trouver sur une ligne de votre fichier. Vous ne pouvez pas insérer de lignes de commentaire, mais vous pouvez insérer des lignes blanches pour la lisibilité.

Si vous cherchez un contexte pour un mot donné, saisissez `context:=` comme dans cet exemple :

```
réserver context:=hôtel
```

(qui recherche `réserver` et n'affiche un résultat que si `hôtel` figure dans le voisinage de contexte)..

Ensuite, au lieu de saisir chaque mot ou syntagme dans la boîte de dialogue des mots de recherche, recherchez simplement le fichier.

Lemmes et concordance à partir de fichiers

Remarque : un fichier temporaire est créé à l'endroit où WordList a [ouvert](#) Concord et où le mot surligné de la liste de mots est l'entrée principale avec des [lemmes](#). Il liste tous les groupes de lemmes et Concord utilise ces termes de recherche issus du fichier pour calculer la concordance. Votre fichier temporaire sera enregistré dans le répertoire `\wsmith4` sauf si vous travaillez en réseau auquel cas ce sera enregistré dans le fichier temporaire de Windows

\windows\temp.. C'est à vous d'effacer le fichier temporaire.

Concordances automatisées à partir de fichiers

Si vous voulez que Concord utilise de nombreux termes de recherche différents, en sauvegardant chaque résultat au fur et à mesure de sorte à le laisser travailler sans rester à côté, choisissez *Lot par mot-clé* dans [Paramètres de concordance](#).

6.21 balise la plus proche

Concord indique la balise la plus proche, à condition que vous ayez spécifié un [fichier de balises](#) qui indique à Oxford WordSmith Tools quels sont vos balises préférées. Lorsque la concordance s'affiche, la balise apparaît dans une des colonnes de la fenêtre de concordance.

Le but

L'avantage est que vous pouvez voir quels sont les liens entre votre terme de recherche pour la concordance et le texte balisé. Par exemple, si vous placez la balise [**Rob**] pour marquer le discours de Robert et la balise [**Marie**] pour marquer le discours de Marie, vous pourrez facilement repérer de qui émane une réplique dans une conversation entre Marie, Robert et des tiers en effectuant une concordance.

Vous pouvez également baliser votre texte avec `<Introduction>`, `<Body>` and `<Conclusion>`: La fonction balise la plus proche affichera chaque ligne comme suit :

```
1 ... ne put me donner l'heure ... <Introduction>
2 ... Rosemary, donne m'en un autre... <Body>
3 ... voulait l'aider ... <Body>
4 ... il n'avait pas beaucoup à donner... <Conclusion>
```

Pour baliser ainsi le texte, créez un [fichier de balises](#) à partir de vos sections et marquez-les comme étant des sections, comme dans les exemples ci-dessous :

```
<EXTRAIT> /description "section"
</ABSTRACT>
<INTRODUCTION> /description "section"
</INTRODUCTION>
<SECTION 1> /description "section"
</SECTION 1>
```

Vous pouvez [trier](#) en fonction des balises les plus proches. La capture d'écran ci-dessous montre une concordance du mot **such**, générée à partir d'un texte du BNC World Edition. **such** est tantôt marqué de la balise `< PRP>` (**such as**), tantôt de la balise `<w DT0>`. La colonne Balises affiche la balise la plus proche, et l'ensemble de la liste a été trié en fonction de cette colonne.

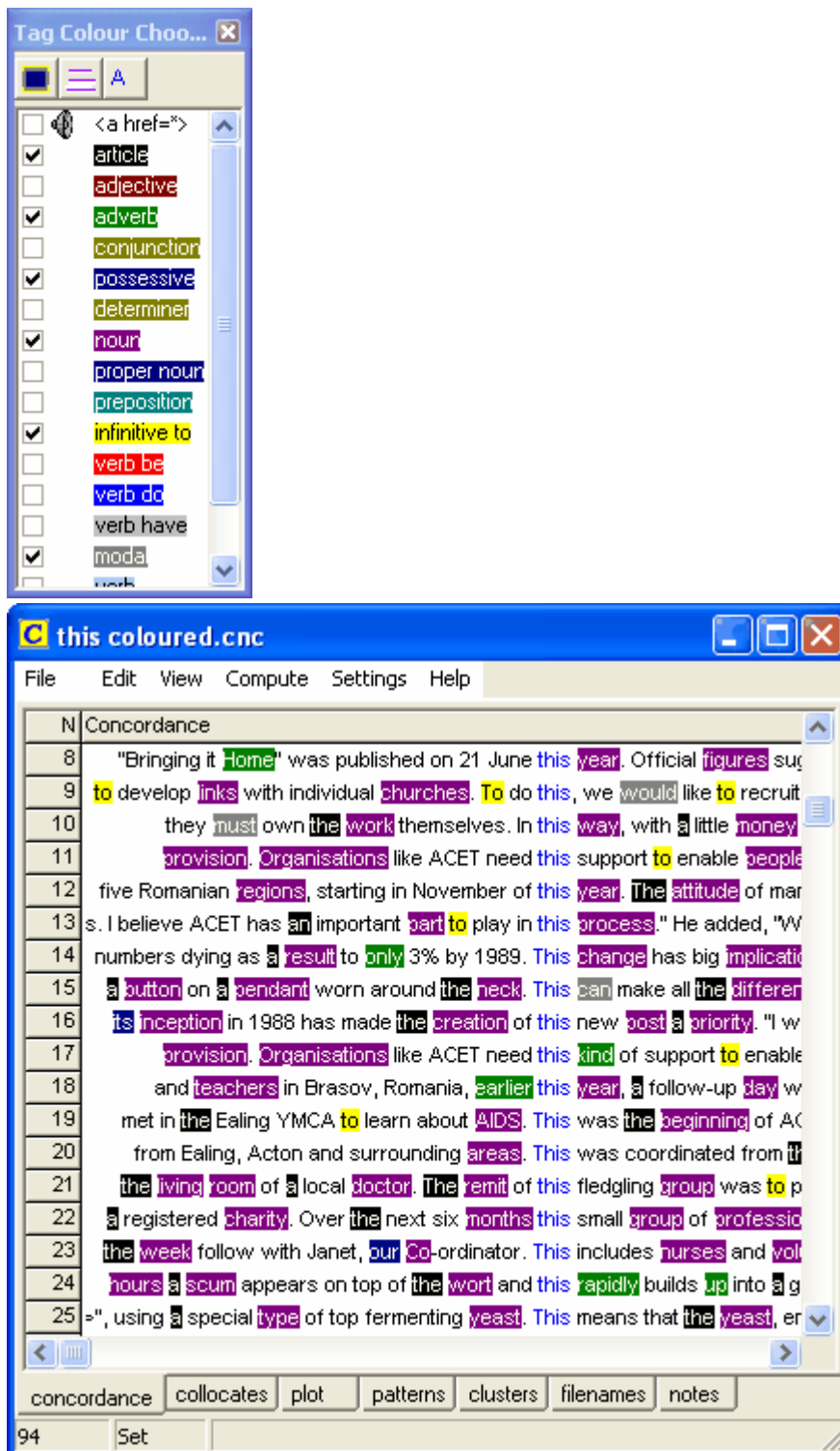
N	Concordance	et	Tag	ord #	#	s.
18	<w ATO>a <w NN1>scheme <w PRP>such as <w ATO>a <w NN1>mural	<w PRP>	<w PRP>	1,993	01	20
19	<w NNO>works <w PRP>such as <w NN2>altarpieces <w	<w PRP>	<w PRP>	2,310	12	9
20	not <w VVI>date, <w PRP>such as <w NN1>art <w	<w PRP>	<w PRP>	4,403	00	16
21	<w NN1>programme, <w PRP>such as <w ATO>the <w	<w PRP>	<w PRP>	1,417	07	9
22	<w NN1>title, <w PRP>such as <w NN1>futurism. <s	<w PRP>	<w PRP>	1,958	30	55
23	<w NN2>techniques <w PRP>such as <w AJO>infra-red <w	<w PRP>	<w PRP>	6,063	04	39
24	<w TOO>to <w VVI>make <w DTO>such <w NN2>identifications,	<w DT>	<w DT>	2,792	08	19
25	<w CJC>or <w DTO>such <w NN2>questions <w	<w DT>	<w DT>	3,133	20	16
26	<w CJC>and <w DTO>such <w NN2>treatises <w	<w DT>	<w DT>	3,170	21	35
27	<w VMO>may <w VVI>take <w DTO>such <w ATO>a <w	<w DT>	<w DT>	3,570	36	6
28	<w NN1>creation of <w DTO>such <w NN2>theories <w	<w DT>	<w DT>	3,722	41	25
29	<w PRP>by <w DTO>such <w ATO>an <w	<w DT>	<w DT>	4,857	95	17
30	<w PRP>in <w DTO>such <w NN2>papers <w CJS>as	<w DT>	<w DT>	4,951	99	5
31	<w NN1>criticism of <w DTO>such <w NN1>art <w PRP>within	<w DT>	<w DT>	7,534	01	5
32	<w DTO>This <w VVD>took <w DTO>such <w NN2>forms <w CJS>as	<w DT>	<w DT>	7,724	08	8

Si aucune balise ne s'affiche, cela est certainement dû au fait que les [balises à ignorer](#) sont réglées sur ce format. Par exemple, si vous spécifiez <*> dans Balises à ignorer, toutes les balises comme <title>, <quote>, etc. seront ignorées lors d'une concordance à moins que vous n'utilisiez un [fichier de balises](#). Si c'est le cas, spécifiez le fichier de balises et relancez le processus de concordance.

Vous pouvez afficher les balises en couleur ou les masquer tout en gardant la possibilité d'afficher le mot balisé en couleur. Voici une concordance pour **this** dans un texte du BNC World Edition, les balises sont en couleur :

N	Concordance
3	<w NN1>availability. <w DTO>This <w VBZ>is <w AVD>so <w PNP>we
4	<w NN1>spread <w PRF>of <w DTO>this <w AJO>terrible <w NN1>disease
5	<w NN1>need <w PRP>for <w DTO>this <w NN1>service. <w NN1>Home
6	<w NN1>opening <w PRF>of <w DTO>this <w NN1>office <w PRP>in
7	<w CJS>even though <w DTO>this <w VBZ>is <w AVD>increasingly
8	<w PRP>on 21 <w NPO>June <w DTO>this <w NN1>year. <w AJO>Official
9	<w TOO>To <w VDI>do <w DTO>this, <w PNP>we <w VMO>would
10	<w PRP>in <w DTO>this <w NN1>way, <w PRP>with <w ATO>
11	<w VVB-NN1>need <w DTO>this <w VVI>support <w TOO>to
12	<w NPO>November <w PRF>of <w DTO>this <w NN1>year. <w ATO>The
13	<w VVI>play <w PRP-AVP>in <w DTO>this <w NN1>process. &equo; <w PNP>He
14	3&percent; <w PRP>by 1989. <w DTO>This <w NN1>change <w VHZ>has
15	<w ATO>the <w NN1>neck. <w DTO>This <w VMO>can <w VVI>make
16	<w NN1>creation <w PRF>of <w DTO>this <w AJO>new <w NN1>sect <w ATO>sh

et voici ce que vous obtenez en cliquant sur *Affichage | Masquer les balises*.



Les balises n'apparaissent plus, et seuls 6 types de balises sont affichées en couleur.

Rubriques connexes : [Présentation générale des balises](#), [Gestion des balises](#), [Création d'un fichier](#)

[de balises](#), [Textes balisés](#), [Types de balises](#), [Affichage des balises](#), [Utilisation des balises comme sélecteurs de texte](#)

6.22 schémas

Lorsqu'une fenêtre de collocation est ouverte, l'un des onglets affiche "Schémas". Cela affiche les co-occurents (mots adjacents au terme de recherche), classés par fréquence au sein de chaque colonne. Le mot le plus haut de chaque colonne est celui qui apparaît le plus fréquemment dans cette position. Le deuxième mot est celui dont la fréquence est la deuxième plus élevée.

La fréquence minimum pour que les mots apparaissent est la [fréquence minimum des co-occurents](#).

Le but

Les éléments apparaissant le plus fréquemment au voisinage du terme de recherche apparaissent ainsi en haut. Tout comme une collocation, ceci vous permet de voir les schémas lexicaux de la concordance.

6.23 re-trier :

Procédure...

Pour trier, cliquez simplement sur la première ligne d'une liste. Vous pouvez également appuyer sur F6 ou Ctrl+F6, ou encore passer par l'option de menu.

Lorsqu'une concordance est générée, elle apparaît dans l'ordre de vos fichiers texte. Lorsque la concordance est terminée, elle est re-triée en fonction de vos [paramètres par défaut](#). Si vous souhaitez pas que la liste soit re-triée, indiquez "file,file" dans les paramètres.

Le but

Re-trier permet de repérer les schémas lexicaux caractéristiques. Il est parfois difficile de voir les tendances dans vos lignes de concordance, surtout lorsque ces dernières sont nombreuses. Le tri vous permet de différencier les mots de recherche multiples et d'examiner leur contexte immédiat, à gauche et à droite de l'élément. Admettons que la plupart des entrées comportent "dans le" ou "dans un" ou "dans mon" juste avant le mot de recherche, il est alors utile de faire un tri en fonction du deuxième mot à gauche du mot de recherche.

Le tri s'effectue par rapport à un certain nombre de mots à gauche ou à droite (L1 [=1 mot à gauche du mot de recherche], L2, L3, L4, L5, R1 [=1 un mot à droite], R2, R3, R4, R5) du mot de recherche, au terme de contexte si vous en avez spécifié un, à la [balise la plus proche](#), à la distance jusqu'à la balise la plus proche, à une [catégorie définie par l'utilisateur](#) ou par rapport au fichier original.

Tri principal

Vous pouvez appliquer un maximum de trois critères de tri à une liste. Si vous appliquez L1 comme tri principal, les entrées sont triées par ordre alphabétique des mots se trouvant directement à gauche du terme de recherche. Si vous effectuez un second tri par rapport à R2, la liste est triée tout en départageant : seules les entrées dont les mots placés en L1 correspondent exactement sont classées par ordre alphabétique, suivant les mots en R2 par rapport au mot de recherche. Dans le cas de grosses concordances, il peut être utile de procéder à un troisième tri : cela permet de différencier les entrées auxquelles vous avez appliqué un tri secondaire.

Il n'est pas toujours utile de départager si les deux premiers tris sont identiques (par exemple L1 et L1) ou si vous ne cochez pas la case.

trier par ensemble ([catégories définies par l'utilisateur](#))

Vous pouvez également trier en fonction de catégories que vous aurez prédéfinies, en utilisant des lettres de **A** à **Z** ou **a** à **z**. Le tri place en premier les lignes définies par l'utilisateur, classées par catégories, suivies de toutes les autres lignes qui ne correspondent pas à vos catégories (ces dernières apparaissent en grisé). Reportez-vous à la section [Balise la plus proche](#) pour de plus amples informations sur le tri par balises.

La couleur du terme de recherche change en fonction du système de tri utilisé.

Tout

Par défaut, toutes les lignes sont triées, mais vous pouvez saisir par exemple 5-49 pour trier uniquement ces lignes.

Croissant

Si cette option est activée, le tri se fait de **A** à **Z** ou de **Z** à **A**.

Rubriques connexes : [Tri dans WordList](#), [Tri dans KeyWords](#), [Sélectionner une langue](#)

6.24 re-trier : co-occurents

Les collocations affichées par ordre de fréquence peuvent re-triées pour afficher les fréquences triées par fréquences *totales* dans l'ensemble des fichiers (paramètres par défaut), par le total des fréquences à gauche ou à droite ou par position de fréquence individuelle, allant de 25 mots à gauche ou 25 mots à droite du terme de recherche. Cliquez simplement sur l'en-tête d'une colonne pour la trier. Cliquez à nouveau pour basculer entre tri croissant et tri décroissant.

Le but

Ceci permet de repérer des schémas de collocation, afin de comprendre totalement quels sont les mots entourant votre terme de recherche.

Les options dépendent du [voisinage de co-occurrence](#).

Rubriques connexes : [Collocation](#), [Affichage des collocations](#)

6.25 re-trier : schéma de dispersion

Ceci permet de re-trier automatiquement un schéma de dispersion et de sélectionner l'une des options suivantes :

ordre alphabétique (par nom de fichier)

par ordre de *fréquence* (termes trouvés par tranche de 1 000 mots de texte)

par la première occurrence dans les textes sources : *ordre du texte*

par *rang* : l'intervalle entre la première et la dernière occurrence dans le texte source.

Rubriques connexes : [Schéma de dispersion](#)

6.26 segments de texte dans Concord

Une ligne de concordance donne des informations sur le segment de texte dans lequel elle a été trouvée.

Sur la capture d'écran ci-dessous, une concordance du mot **year** a été effectuée. La liste a été triée par position dans l'en-tête. Au niveau des deux premières lignes, le mot **year** apparaît en tant que troisième mot d'en-tête. L'avantage est que cela permet d'identifier des termes de recherches apparaissant près du début d'une phrase, d'une section, d'un en-tête ou d'un paragraphe.

N	Word #	Sent. #	Sent. Pos.	Para. #	Para. Pos.	Head. #	Head. Pos.	Sect. #	Sect. Pos.
1	6	1	3	0	6	0	3	1	4
2	18,829	1,028	3	203	8	15	3	16	4
3	24,527	1,364	6	344	9			53	7
4	35,743	1,983	4	519	31			89	15
5	16,762	929	7	226	7			45	16
6	19,025	1,051	3	260	26			47	35
7	20,297	1,132	30	228	30			18	43
8	34,829	1,933	21	505	40			87	49
9	20,305	1,133	9	228	38			18	51
10	24,792	1,376	16	347	60			54	64

Rubriques connexes : [Début et fin des segments de texte.](#)

6.27 syntaxe du terme de recherche

Par défaut, Concord effectue une recherche par mot entier, non sensible à la casse.

Exemples

terme de recherche	trouve
book	<i>Book</i> ou <i>book</i> ou <i>BoOk</i>
book	<i>book, books, booking, booked</i>
*book	<i>textbook</i> (mais pas <i>textbooks</i>)
bo* in	<i>book in, books in, booking in</i> (mais pas <i>book into</i>)
book * hotel	<i>book a hotel, book the hotel, book my hotel</i>
bo* in*	<i>book in, books in, booking in, book into</i>
book?	<i>book, books, book; book.</i>
book^	<i>book, books</i>
b^^k	<i>book, back, bank, etc.</i>
==book==	<i>book</i> (mais pas <i>BOOK</i> ou <i>Book</i>)
book/paperback	<i>book</i> ou <i>paperback</i>

symbol	signification	exemples
e		
*	ignorer la fin du mot, ignorer le mot entier	<i>tele*</i> <i>*ness</i> <i>*happi*</i>
?	tout caractère simple (y compris la ponctuation)	<i>book * hotel</i> <i>Engl???</i> <i>?50.00</i>

#	n'importe quel chiffre entre 0 et 9	\$### £##,00
^	n'importe quelle lettre de l'alphabet	Fr^nc^
==	sensible à la casse	==French== ==Fr*==
:\	utiliser un fichier pour de nombreux mots de recherche (reportez-vous à la section termes de recherche issus du fichier)	c:\text\frd.txt
/	séparer les mots de recherche. Vous pouvez spécifier des mots de recherche différents, dans une limite de 80 caractères.	may/can/will
<>	début et fin de balises	<w NN1>

Pour utiliser *, ? , == , # , ^ , :\ , > , < ou / comme caractère dans votre mot de recherche, placez-le entre guillemets. Exemples :

```
"*"
"why"?"
"and"/"or"
":\"
"<"
```

N'oubliez pas que les points d'interrogation se trouvent en fin de mot, vous aurez peut-être besoin de "*"?"

Balises

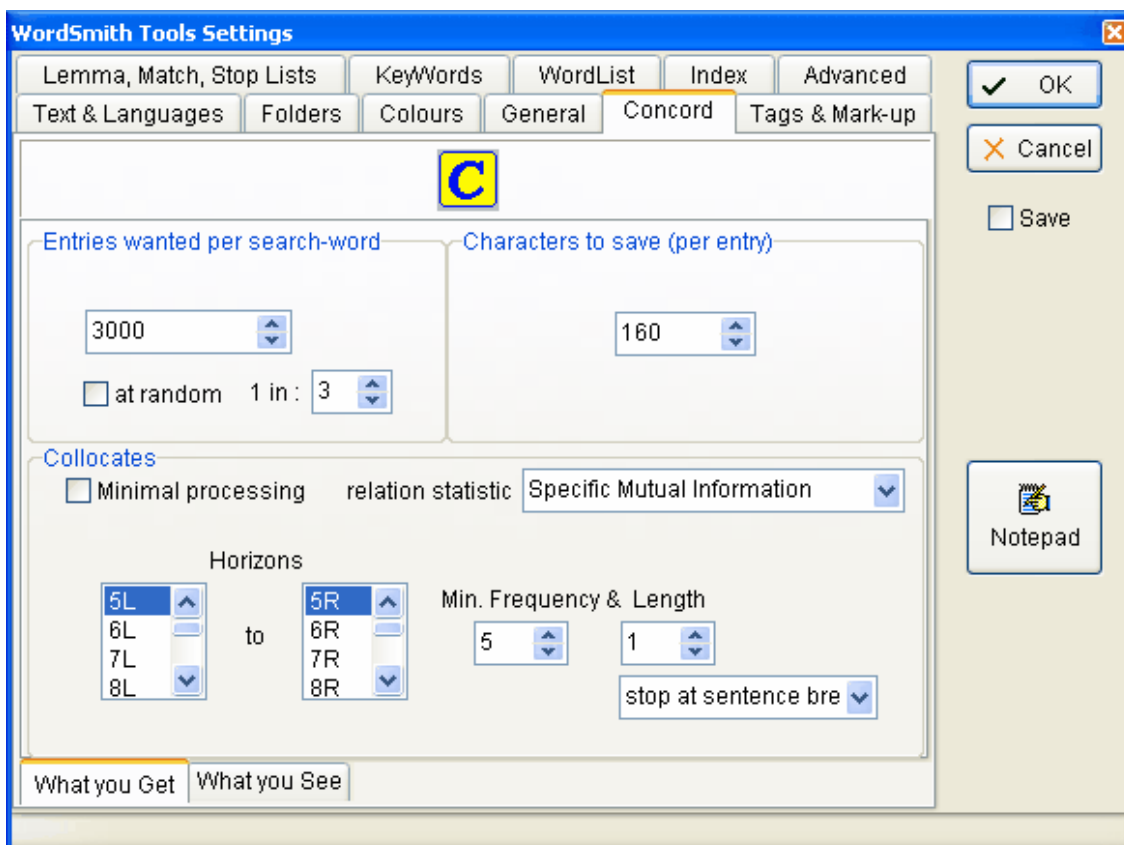
Si votre texte est balisé, vous pouvez inclure des balises dans vos mots de recherche. Exemples :

symbole	signification	exemples
<w NN1>*	nom commun (BNC)	<i>book, chair, elephant</i>
<w NN?>*	nom commun singulier ou pluriel	<i>book, chairs</i>
<w NN1>t*	nom commençant par T ou t	<i>table, teacher</i>
<w NN1>* <w NN1>*	deux noms communs qui se suivent	<i>campaign manager</i>

Rubriques connexes : [Concordance des balises](#), [Mot de contexte](#)

6.28 contrôleur WordSmith : Concord : paramètres

Ces derniers sont accessibles via le [contrôleur](#) dans *Modifier les paramètres | Concord*. Certains des choix effectués, par exemple le [voisinage de co-occurrence](#), peuvent affecter d'autres outils.



CE QUE VOUS OBTENEZ et CE QUI S'AFFICHE

Deux onglets de paramètres affectent *ce que vous obtenez* dans la concordance et *ce qui s'affiche* dans la fenêtre. La capture d'écran ci-dessous affiche les options de *ce qui s'affiche*.

CE QUE VOUS OBTENEZ

Entrées demandées

La valeur maximum dépasse les 2 milliards de lignes. Cette fonctionnalité est utile si vous effectuez plusieurs recherches et que vous souhaitez une centaine d'exemples pour chacune d'entre elles. Les 100 entrées sont les 100 premières trouvées dans les textes que vous avez sélectionnés. Si vous recherchez plus d'un terme de recherche, par exemple **livre/poche**, vous obtiendrez 100 occurrences de **livre** et 100 de **poche**.

"*aléatoire*" est une fonctionnalité permettant de rendre la recherche aléatoire. Ici, **Concord** parcourt les fichiers texte et récupère les 100 entrées en donnant à chaque entrée une chance sur trois d'être sélectionnée. Pour obtenir 100 entrées, **Concord** doit trouver 250 à 350 entrées. Vous pouvez définir la valeur aléatoire de 1 sur 2 à 1 sur 1 000.

Caractères à enregistrer

C'est ici que vous définissez le nombre de caractères d'une ligne de concordance devant être enregistrés au format texte (par opposition à l'envoi de caractères vers le [presse-papiers](#)). La valeur par défaut est de 80 (20 au minimum, 8 000 au maximum). Ainsi, vous pouvez spécifier un nombre fixe de caractères pour que vos termes de recherche s'alignent correctement lorsque vous utilisez une police de caractères non proportionnelle telle que Courier ou Lucinda Console.

Co-occurents

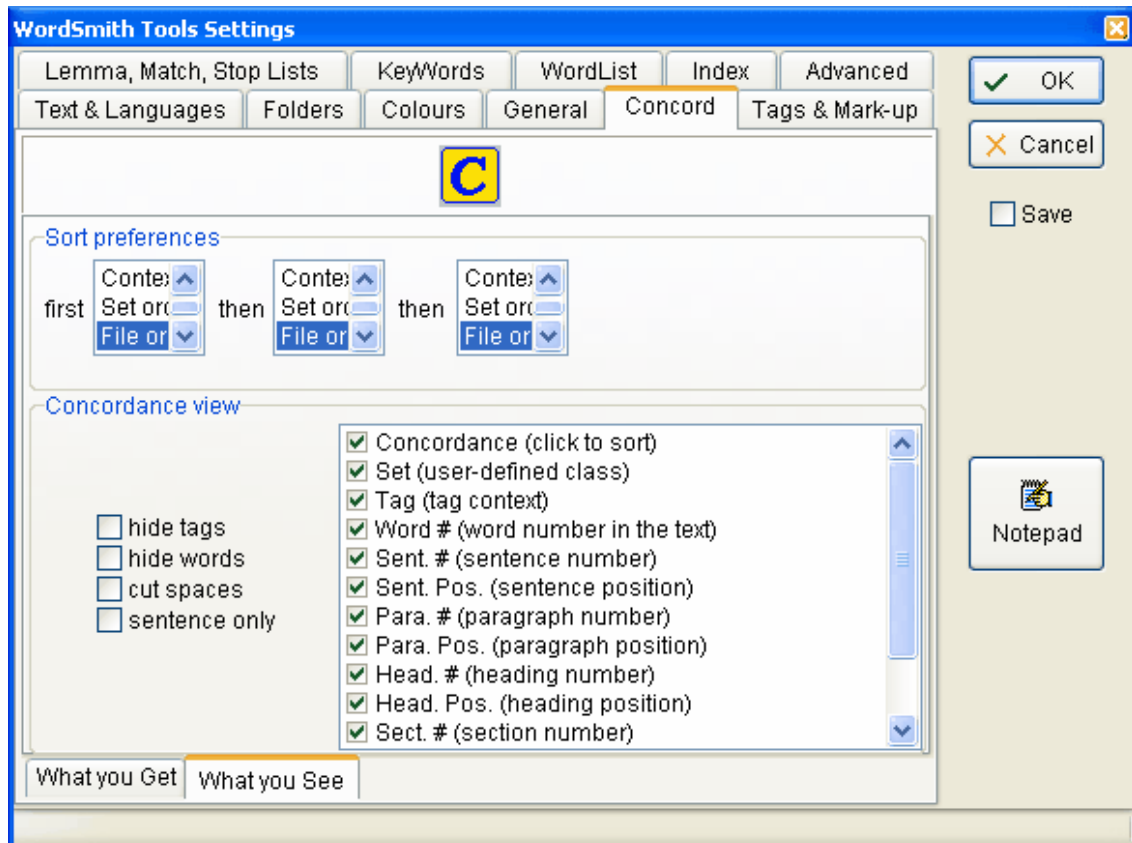
Par défaut, **Concord** calcule les co-occurents ainsi que la concordance, mais vous pouvez

désactiver cela (*Traitement minimal*). Pour de plus amples informations, reportez-vous à la section [Voisinage des co-occurents](#) ou [Collocation](#)

Statistiques de relation des co-occurents

Vous devez choisir entre : informations mutuelles spécifiques, statistiques MI3, note Z et log de vraisemblance. Reportez-vous à la section [Affichage des informations mutuelles](#) pour des exemples montrant les différences entre ces différentes entités statistiques.

CE QUI S'AFFICHE



Préférences de tri

Par défaut **Concord** trie les concordances dans l'ordre des fichiers d'origine, mais vous pouvez modifier ceci si vous le souhaitez. Pour de plus amples informations, reportez-vous à la section [Re-trier une concordance](#).

Affichage de la concordance

Plusieurs possibilités s'offrent à vous pour afficher les données et vous pouvez choisir quelles colonnes afficher pour chaque nouvelle concordance. Vous pouvez rétablir ceci ultérieurement en modifiant la [mise en page](#).

Rubriques connexes : Sommaire de l'aide de Concord, [Paramètres de collocation](#).

WordSmith Tools

Key Words

Section

VIII

7 KeyWords

7.1 but



Ce programme permet d'identifier les "mots-clés" dans un ou plusieurs textes. On appelle mots-clés ceux dont la fréquence est élevée par rapport à une certaine norme. Cliquez pour afficher un [exemple](#).

Le but

KeyWords est utile pour caractériser un texte et son genre. Applications possible du programme : enseignement des langues, linguistique légale, stylistique, analyse de contenu, extraction de texte.

Le programme compare deux listes de mots préexistantes (qui doivent avoir été créées avec l'outil WordList). L'une d'entre elles doit être une liste conséquente qui servira de fichier de référence. L'autre liste est celle basée sur le texte que vous souhaitez étudier.

Le but est de trouver quels sont les mots qui caractérisent le texte qui vous intéresse : on considérera par défaut qu'il s'agit du plus petit des deux textes choisis. Le plus gros des deux fournira des données de référence pour la comparaison.

Les mots-clés et les [liens](#) qui les unissent peuvent être représentés sous forme de [schéma](#), placés dans une [base de données](#), ou regroupés en fonction de leurs [associés](#).

7.2 Analyse de deux listes de mots

Ce n'est ni plus ni moins qu'une analyse de **KeyWords**. Le programme compare le texte qui vous intéresse avec un corpus de référence basé sur une grande quantité de textes.

Sélectionner des listes de mots

La boîte de dialogue vous permet de sélectionner deux fichiers : Le fichier, dans la partie supérieure, et le corpus de référence dans la partie inférieure.

Rubriques connexes : [Calcul des mots-clés](#), [Paramètres de KeyWords](#)

7.3 définition des associés

Un "associé" du mot-clé X est un autre mot-clé (Y) qui apparaît avec X dans un certain nombre de textes. Il peut apparaître ou ne pas apparaître à proximité du mot-clé X. Un *co-occurent* doit apparaître à une distance donnée du mot-clé alors qu'un associé est "associé" en étant clé dans le même texte.

Par exemple, dans une base de données de mots-clés du journal *Guardian*, *wine* s'est avéré être un mot-clé dans 25 des 299 articles de la page "tabloïd" de l'édition du dimanche, devenant ainsi [mot-clé clé](#) de cette section. Les associés les plus notoires de *wine* étaient : *wines*, *Tim*, *Atkin*, *dry*, *le*, *bottle*, *de*, *fruit*, *region*, *chardonnay*, *red*, *producers*, *beaujolais*.

On est ici très près de la notion première de "co-occurent".

Les associations fonctionnent de différentes manières. Une association peut être forte ou faible, réciproque ou à sens unique. En anglais par exemple, l'association entre *to* et *fro* est à sens unique (*to* se trouve presque toujours à côté de *fro* mais il est rare de trouver *fro* à côté de *to*).

Rubriques connexes : [Définition d'un mot-clé](#), [Associés](#), [Définitions](#), [Informations mutuelles](#)

7.4 associés

Le mot "associés" désigne des mots-clés associés à un [mot-clé clé](#).

Le but

Le but est d'identifier les mots qui sont associés à un même mot-clé clé parce qu'ils ont le statut de mots-clés dans les mêmes textes que le mot-clé clé. Un exemple permettra d'y voir plus clair. Supposons que le mot *vin* soit un mot-clé clé d'un ensemble de textes tels que des articles de journaux. Certains de ces articles traitent des différents vins et de leurs saveurs, d'autres traitent de la cuisine et parlent de l'utilisation du vin dans des ragoûts ou en sauce, enfin, d'autres traitent des prix du vin dans l'industrie agricole ou mentionnent les maladies qui affectent les vignes. Dans ce cas, les associés de *vin* seraient des éléments tels que *Chardonnay, Chile, sauce, fruit, infecté, terre*, etc.

La liste affiche les associés par ordre de fréquence. Une option de menu vous permet de les trier différemment.

Vous pouvez définir un nombre minimum de fichiers pour la procédure d'association en cliquant sur *Paramètres / Base de données*.

Rubriques connexes : [définition d'un associé](#).

7.5 sélection des fichiers

Liste de mots du fichier actif

Sélectionnez un fichier de liste de mots dans la boîte supérieure.

Pour sélectionner plusieurs fichiers de listes de mots, appuyez sur Ctrl tout en cliquant pour sélectionner des listes non adjacentes, ou appuyez sur Maj pour sélectionner un bloc.

Cette boîte détermine les listes de mots dont vous allez rechercher les mots-clés.

Liste de mots d'un corpus de référence

Dans la boîte ci-dessous, vous pouvez sélectionner votre liste de [Corpus de référence](#). Ceci peut être défini de façon permanente dans la section Paramètres du contrôleur principal.

Aucune liste de mots visible

Si vous ne voyez aucune liste de mots s'afficher, vous pouvez soit modifier les dossiers ou retourner dans WordList et créer au moins deux listes de mots. Cette procédure nécessite deux listes de mots pour pouvoir effectuer une comparaison.

7.6 groupes

Le mot "groupe" désigne un ensemble de mots-clés [associés](#) à un [mot-clé clé](#).

Le but (1)


L'idée ici est de redéfinir les associés en groupant des mots ayant le statut de mots-clés au sein d'un même sous-ensemble de fichiers texte. Un exemple permettra d'éclaircir cette notion.

Supposons que le mot *vin* soit un mot-clé clé d'un ensemble de textes tels que des articles de journaux. Certains de ces articles traitent des différents vins et de leurs saveurs, d'autres traitent de la cuisine et parlent de l'utilisation du vin dans des ragoûts ou en sauce, enfin, d'autres traitent des prix du vin dans l'industrie agricole ou mentionnent les maladies qui affectent les vignes.

Dans ce cas, les associés de *vin* seraient des éléments tels que *Chardonnay, Chile, sauce, fruit, infecté, terre*, etc. La procédure traitant les associés affiche des éléments non-triés.

La procédure de groupement, en revanche, effectue un tri en se basant sur les diverses utilisations. Le fonctionnement est le suivant : les mots-clés de chaque texte donnent un aperçu général du "sujet", et les "sujets" des différents textes peuvent être regroupés en faisant correspondre les listes de mots-clés. Ainsi, les ensembles de mots-clés peuvent être groupés en fonction du degré de chevauchement sémantique dans chaque fichier texte.

Deux étapes

Le **processus initial de groupement n'effectue pas de groupement réel** : vous verrez chaque ensemble de mots-clés correspondant à chaque fichier texte s'afficher séparément. Pour effectuer un [groupement](#), il suffit d'associer les groupes allant ensemble (en les faisant glisser), ou de les regrouper en cliquant sur .

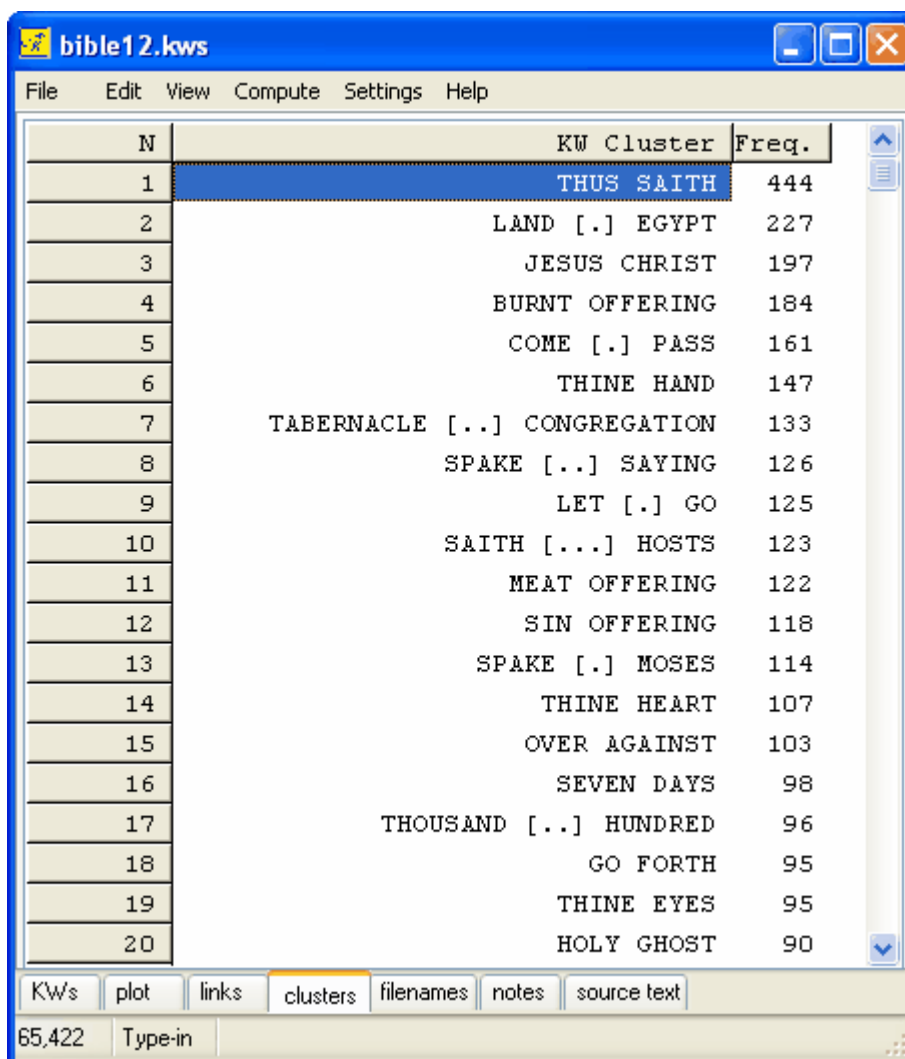
La liste affiche les groupes classés par ordre alphabétique. Vous pouvez les re-trier par fréquence (le nombre d'apparitions de chaque mot-clé du groupe dans tous les fichiers contenant ce groupe).

Rubriques connexes : [définition d'un associé](#), [effectuer un regroupement de synonymes](#)

7.7 Clusters KeyWords

Tout comme le cluster WordList, le cluster KeyWords représente deux mots ou plus qui apparaissent fréquemment à proximité les uns des autres. Veuillez noter qu'un cluster KeyWords ne considère que les mots-clés.

Vous comprendrez mieux avec cette capture d'écran.



N	KW Cluster	Freq.
1	THUS SAITH	444
2	LAND [.] EGYPT	227
3	JESUS CHRIST	197
4	BURNT OFFERING	184
5	COME [.] PASS	161
6	THINE HAND	147
7	TABERNACLE [...] CONGREGATION	133
8	SPAKE [...] SAYING	126
9	LET [.] GO	125
10	SAITH [...] HOSTS	123
11	MEAT OFFERING	122
12	SIN OFFERING	118
13	SPAKE [.] MOSES	114
14	THINE HEART	107
15	OVER AGAINST	103
16	SEVEN DAYS	98
17	THOUSAND [...] HUNDRED	96
18	GO FORTH	95
19	THINE EYES	95
20	HOLY GHOST	90

Ces clusters ont été calculés en prenant comme texte source la Bible. Chacun de ces mots est "clé" par rapport à un corpus de référence. Les clusters montrent les cas où ces mots-clés

apparaissent dans le voisinage de co-occurrence actuel. Les crochets ([. . .]) indiquent les cas où les mots-clés n'apparaissent pas ensemble. Dans **come** [.] **pass** par exemple, il y a un point car les occurrences qui se répètent sont **come to pass**.

Rubriques connexes : [Calcul des schémas](#).

7.8 concordance

Lorsqu'une liste de mots-clés s'affiche à l'écran, vous pouvez appuyer sur **C** pour appeler une concordance du mot ou des mots actuellement sélectionné(s). La concordance recherchera ce même mot dans le fichier texte original dont votre liste de mots-clés provient.

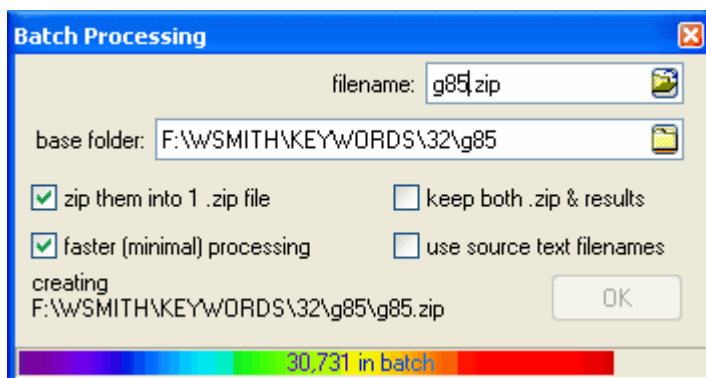
Le but

Cela vous permet d'afficher les mêmes mots-clés dans leurs contextes originaux.

7.9 Création d'une base de données

Pour créer une base de données de mots-clés, vous avez besoin d'un ensemble de listes de mots-clés. Pour obtenir une base de données de taille correcte, il est préférable de la construire comme suit :

1. Créez un [lot](#) de listes de mots.
2. Utilisez-le pour créer un [lot](#) de listes de mots-clés. Activez "traitement minimal rapide" comme le montre la capture d'écran, afin de ne pas perdre de temps à calculer les schémas, etc.



3. Ensuite, dans **KeyWords**, cliquez sur *Nouveau | Base de données*.

Vous pouvez alors sélectionner l'ensemble des fichiers de mots-clés.

Remarque : la création d'une base de données signifie que seuls les mots-clés dont la valeur est [positive](#) seront retenus.

Dans les [Paramètres du contrôleur KeyWords](#), vous pouvez spécifier le nombre de mots-clés par texte devant être intégrés à votre base de données.

Spécifiez ensuite l'emplacement où la base de données doit être stockée.

Une fois la base de données créée, vous pouvez l'*ouvrir* pour afficher les [mots-clés clés](#) et les [associés](#).

7.10 exemple de mots-clés

Vous disposez de quelques articles de journaux. Vous créez une liste de mots à partir de ces articles et voyez que le mot le plus fréquent est *le*. Parmi les mots les moins fréquents de la liste vous trouvez par exemple *terre*, *menoté*, *protestations*, etc.

Créez ensuite une liste de mots à partir des 1 000 mots de l'un de ces articles. Le mot le plus fréquent est à nouveau *le*. Jusqu'ici, il n'y a aucune différence.

Vous pouvez ensuite utiliser **KeyWords** pour analyser les deux listes de mots. **KeyWords** indique que les mots-clés les plus pertinents sont : *squat*, *police*, *juge*, *prison*, *Dupond*, *résisté*, *communauté*.

Ces mots-clés ne sont pas les mots qui reviennent le plus fréquemment (comme c'est par exemple le cas pour *le*) mais ceux qui sont fréquents de manière significative dans l'article de 1 000 mots. On peut généralement se faire une assez bonne idée du sujet d'un texte à partir de ses mots-clés.

7.11 définition des mots-clés clés

Un "mot-clé clé" est un mot qui est mot-clé dans au moins deux textes d'un corpus. Plus le nombre de textes dans lequel il est mot-clé est élevé, plus le mot est un mot-clé "pertinent". Cela dépend beaucoup de l'homogénéité du contenu textuel de votre corpus. Dans un corpus de textes ayant trait au monde des affaires, des éléments tels que *banque*, *bénéfice*, et *entreprises* sont des mots-clés clés, par contre, *ordinateur* n'en est pas un même s'il peut être mot-clé dans certains articles traitant de transactions effectuées par IBM ou Microsoft.

Rubriques connexes : [Calcul des mots-clés](#), [Définition d'un mot-clé](#), [Création d'une base de données](#), [Définitions](#)

7.12 définition de pertinence

Le terme "mot-clé", bien que couramment utilisé, n'a pas de définition linguistique. Le programme identifie les mots-clés mécaniquement en comparant des schémas de fréquence. L'être humain, lui, considérera plutôt un syntagme ou un hyperonyme comme étant un mot-clé. Un mot est dit "clé" si :

- la fréquence du mot dans le texte dépasse la fréquence minimum spécifiée par l'utilisateur
- sa fréquence dans le texte, comparée à sa fréquence dans un corpus de référence, est telle que la probabilité statistique, calculée à l'aide d'une [procédure appropriée](#), est plus petite que ou égale à une [p valeur](#) spécifiée par l'utilisateur.

pertinence positive et négative

Un mot-clé *positivement* pertinent est *plus* fréquent que ce à quoi l'on peut s'attendre par rapport au corpus de référence.


Un mot-clé *négativement* pertinent est *moins* fréquent que ce à quoi l'on peut s'attendre par rapport au corpus de référence.

mots-clés typiques

Les mots-clés trouvés par KeyWords sont généralement de 3 sortes :

Premièrement, il y a les noms propres. Les noms propres sont souvent "clés" dans un texte, bien qu'un texte sur les courses hippiques puisse indiquer des noms de chevaux, plutôt secondaires à l'histoire, comme étant des mots-clés. Ceci peut être évité en spécifiant une fréquence minimum plus élevée.

Deuxièmement, on trouve les mots-clés que l'être humain aurait relevé. Le programme les trouve aisément, ils donnent un bon aperçu du sujet du texte. Cependant, le programme ne regroupe pas les synonymes, et un mot qui n'apparaît qu'une fois dans un texte peut parfois être un mot-clé pour l'être humain. Et **KeyWords** n'identifiera les "syntagmes-clés" que si vous comparez des listes de mots basées sur des [clusters de mots](#).

Pour finir, on trouve des mots ayant une haute fréquence d'apparition comme "parce que", "pourrait" ou "déjà". Le lecteur ne les considère généralement pas comme des mots-clés. Ces mots sont davantage des indicateurs de style que de contenu. Le fait que KeyWords identifie de tels mots devrait vous inciter à reprendre l'analyse du texte, avec Concord par exemple (cliquez simplement sur ) , afin de déterminer pourquoi ces mots sont inhabituellement fréquents.

Rubriques connexes : [Calcul des mots-clés](#), [Définition d'un mot-clé clé](#), [Définitions](#), [Paramètres de KeyWords](#)

7.13 base de données KeyWords

(extension de fichier par défaut *.KDB)

Le but

Cette base de données vous permet de voir les "mots-clés clés" de votre ensemble de fichiers, c'est-à-dire les mots-clés clés les plus fréquents dans ces fichiers.

Si vous avez par exemple 500 rapports d'activité, chacun aura ses propres mots-clés. Ils se répartiront probablement en deux catégories principales. Il y aura des mots-clés clés, qui sont clés pour un texte mais pas pour l'ensemble des textes (noms d'entreprises et mots se rapportant à ce qu'elles produisent), et il y aura les autres mots, plus généraux, tels que consultant, profit ou employés qui sont typiques de la documentation du monde des affaires. Constituer une base de données vous permet de les trier. Lorsque vous les affichez, ceux situés en haut de la liste sont les plus typiques. La liste est ordonnée selon la "pertinence" mais il est possible de basculer en tri alphabétique.

Vous pouvez définir un nombre minimum de fichiers dans lesquels chaque mot doit être mot-clé en cliquant sur *Paramètres | KeyWords | Base de données*.

Lorsque vous affichez une base de données, vous pourrez examiner les [associés](#) des mots-clés clés.

Dans Statistiques, vous pouvez afficher les détails des fichiers de mots-clés qui comprennent la base de données (nom du fichier et mot-clés par fichier) ainsi que les statistiques générales sur le nombre de types et tokens différents (total de tous les mots-clés dans toute la base de données, répétitions incluses).

Rubriques connexes : [Création d'une base de données](#), [Définition d'un mot-clé clé](#)

7.14 KeyWords : conseils

1. N'appellez pas de schémas de mots-clés issus de textes différents. Cela n'aurait aucun sens. De toute manière, le schéma ne montrera que les mots du premier texte. Si vous souhaitez voir le schéma d'un mot ou syntagme dans différents textes, utilisez la fonction [dispersion](#) dans Concord.

2. Il n'est pas certain que les mots-clés soient "clés" au sens où vous l'entendez. Un mot "important" peut n'apparaître qu'une fois dans un texte. Les mots indiqués sont simplement ceux qui sont inhabituellement fréquents ou peu fréquents en comparaison avec un corpus de référence.

3. Comparez ce qui est comparable. Choisissez avec soin votre corpus de référence. L'ordinateur n'est pas intelligent et effectuera toutes les comparaisons que vous lui demanderez d'effectuer. Il vous appartient de choisir judicieusement les comparaisons à effectuer.

7.15 KeyWords : calcul

Les mots-clés sont calculés en comparant la fréquence de chaque mot de la liste de mots du texte qui vous intéresse avec la fréquence du même mot dans la liste de mots de référence. Tous les mots qui apparaissent dans la plus petite liste sont pris en compte, sauf s'ils se trouvent dans une [liste d'arrêt](#).

Si **le** apparaît dans 5% des cas dans la petite liste et dans 6% des cas dans le corpus de référence, il ne sera pas un mot-clé, même si c'est le mot le plus fréquent. Si le texte traite de l'anatomie des araignées, il se peut que les noms des chercheurs et les éléments **araignée**, **patte**, **huit** etc. soient plus fréquents qu'ils ne le sont dans le corpus de référence, à moins que votre corpus ne parle que d'araignées.

Pour calculer la pertinence d'un élément, le programme calcule donc

- sa fréquence dans la petite liste de mots
- le nombre de mots dans la petite liste de mots
- sa fréquence dans le corpus de référence
- le nombre de mots dans le corpus de référence

Il met ensuite tous ces éléments en regard dans un tableau.

Les tests statistiques incluent :

- le classique test du chi-deux avec la correction de Yates pour un tableau 2 X 2
- le test du log de vraisemblance de [Ted Dunning](#) donne une meilleure estimation de la pertinence, surtout lorsque vous comparez de longs textes ou l'intégralité d'un genre avec votre corpus de référence.

Un mot apparaîtra dans la liste s'il est significativement fréquent (ou peu fréquent) par rapport à ce à quoi vous vous attendez au vu de la grande liste de mots.

Les mots-clés significativement *non* fréquents sont appelés "mots-clés négatifs" et apparaissent en fin de liste dans une couleur différente. À noter que les mots-clés négatifs seront automatiquement ignorés par une [base de données](#) et un schéma.

7.16 KeyWords : liens

Le but

Il s'agit de trouver quels sont les mots-clés directement liés à un mot-clé donné. Un [schéma](#) indique à quel endroit chaque mot-clé apparaît dans le fichier original. Il indique également le nombre de liens entre les mots-clés.

Qu'est-ce qu'un lien ?

Les liens sont des "co-occurrences de mots-clés au sein d'une collocation". Vous comprendrez mieux avec cet exemple :
Supposons que le mot *éléphant* soit un mot-clé d'un texte sur l'Afrique et que *eau* soit un mot-clé de ce même texte. Si le rayon qui sépare *éléphant* et *eau* ne dépasse pas 5 mots, on dit que les deux mots sont "liés". La fenêtre Liens indique combien de fois les deux termes sont ainsi liés.

Ce que vous voyez

La fenêtre Liens affiche le nombre de liens, suivi d'une colonne "dans" et d'un pourcentage. Ce

pourcentage correspond au nombre de liens divisé par le nombre total d'occurrences du mot en question, c'est-à-dire le nombre qui figure dans la colonne "dans". Ainsi, si vous recherchez les liens du mot *éléphant*, qui apparaît 10 fois dans votre texte original, et qu'à chaque fois *éléphant* se trouve près du mot *eau*, vous obtiendrez 100% même si *eau* apparaît 40 fois au total. Si vous recherchez les liens de *eau*, le pourcentage pour *éléphant* sera de 25%.

Le [voisinage](#) de co-occurrence est défini par **Concord**, et peut aller jusqu'à 25 mots de part et d'autre du terme choisi. La [valeur par défaut](#) est 5,5.

Vous pouvez double-cliquer sur n'importe quel mot de [la liste des schémas](#) pour ouvrir une fenêtre (20 fenêtres maximum) contenant les mots-clés liés.

Rubriques connexes : [Calcul des schémas](#), [Clusters KeyWords](#)

7.17 créer une liste de mots à partir de données de mots-clés

Lorsque vous avez une liste de mots-clés sur votre écran, vous pouvez cliquer sur **W** pour enregistrer vos données sous forme de liste de mots (dans le but d'effectuer ultérieurement une comparaison etc, en utilisant les fonctions de **WordList**).

7.18 p valeur

La **p** valeur est utilisée lors de tests du chi-deux ou autres tests statistiques. Cette valeur est comprise entre 0 et 1. Une valeur égale à 0,01 signifie qu'il y a une marge d'erreur de 1 % lors du calcul de relation, et 0,05 signifierait que la marge d'erreur est de 5 %. En sciences sociales, une marge de 5 % est considérée comme acceptable.

Dans le cas d'analyses de mots-clés, où la notion de risque est moins importante que la notion de sélectivité, vous serez souvent amenés à définir un seuil de valeur **p** relativement faible (par exemple 0,000001) afin d'obtenir moins de mots-clés. Vous pouvez également définir une valeur faible pour le "maximum voulu" dans le [contrôleur](#) principal, sous *Modifier les paramètres / KeyWords*.

Si la [procédure du chi-deux](#) est utilisée, la p valeur calculée n'apparaîtra que si les critères statistiques sont respectés (toutes les valeurs ≥ 5).

Rubriques connexes : [Définitions](#)

7.19 calcul des schémas

Le but

Cela vous permet de voir comment les mots-clés sont répartis dans le texte. Sont-ils regroupés au milieu du texte ou en début de texte?

Procédure

Les relations entre tous les mots-clés identifiés sont calculées, à l'exception de ceux que vous avez supprimés ou [ignorés](#).

1. effectue une concordance du texte qui recherche toutes les occurrences de chaque mot-clé ;
2. détermine ensuite, parmi tous les autres mots-clés, quels sont ceux qui apparaissent dans le voisinage de co-occurrence (défini dans *Paramètres*). Le plus grand des deux voisinages de co-occurrence est utilisé.
3. schématise tous les mots et montre l'emplacement de chaque occurrence dans le fichier original (avec une "règle" indiquant le nombre de mots dans chaque partie du fichier).
4. calcule combien d'autres mots-clés apparaissent avec chacun d'entre eux au sein du voisinage de co-occurrence.

5. calcule la [valeur de dispersion du schéma](#).

Remarque : ce processus n'est possible que si KeyWords est en mesure de trouver les [textes source](#) sur lesquels votre liste de mots originale est basée.

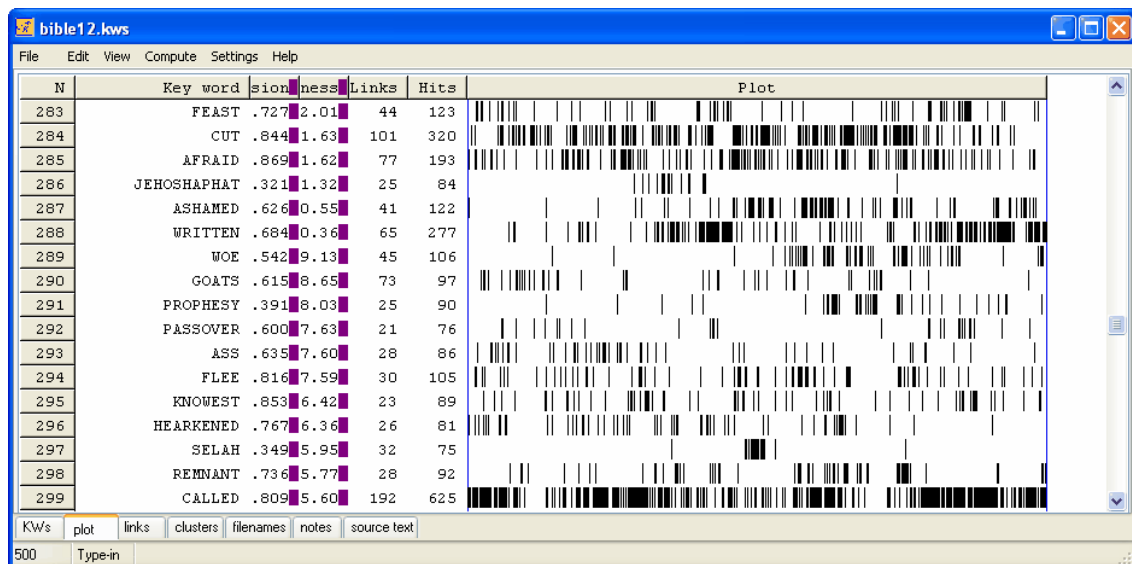
Il peut être utile d'[exporter votre schéma](#) et de créer d'autres graphiques, comme l'explique la section [Enregistrer sous](#).

Rubriques connexes : [Liens des schémas](#), [Affichage du schéma des mots-clés](#)


7.20 affichage des schémas

Le schéma vous permet de voir la fréquence et l'emplacement où apparaissent les mots-clés dans le texte. Le schéma est initialement [trié](#) pour montrer quels mots apparaissent plus en début de texte (par exemple dans l'introduction), puis ceux qui apparaissent plus loin dans le texte.

La capture d'écran ci-dessous montre les mots-clés de la Bible ainsi que leur emplacement. Le nom **Jehoshaphat** par exemple apparait principalement dans le premier tiers du texte.



re-trier :

Vous pouvez [re-trier](#) la liste en cliquant sur . Re-trier permet de basculer entre les modes suivants :

- première mention de chaque mot-clé dans le texte
- dispersion dans le texte
- ordre original du schéma (basé sur la pertinence)
- ordre alphabétique
- nombre total de liens avec d'autres mots-clés

liens


Affiche le nombre total de [liens](#) entre le mot-clé et les autres mots-clés dans le même texte, au sein de la collocation actuelle ([valeur par défaut](#) = 5,5). En d'autres termes, il s'agit du nombre de fois que chaque mot-clé est trouvé dans un rayon de 5 mots à gauche ou à droite des autres mots-clés de votre schéma.

occurrences

Cette colonne vous indique le nombre d'occurrences de chaque mot-clé.

Les schémas vous permettent de voir la relation qui existe entre les mots. Pour ce faire, reportez-vous à la fenêtre de liens dans les onglets du bas, qui montre quels sont les autres mots-clés sont les plus [liés](#) au mot que vous avez sélectionné. Il s'agit des autres mots qui apparaissent le plus souvent dans le voisinage de co-occurrence que vous avez défini. La fenêtre de liens devrait vous permettre de mieux cerner les relations lexicales.

Chaque fenêtre de schéma dépend de la liste de mots-clés à partir de laquelle elle a été générée. Si vous fermez la liste, le schéma disparaît. Vous pouvez *imprimer* le schéma. Il n'y a pas d'option *Enregistrer* car le schéma provient d'une liste de mots-clés que vous devez *Enregistrer*, ou *Enregistrer Sous*. Il n'y a pas d'option [enregistrer au format texte](#) car le schéma comporte des graphiques qui ne peuvent pas être correctement représentés sous forme de symboles texte, mais vous pouvez le *Copier* dans le [presse-papiers](#) (Ctrl+Ins) et le coller ensuite dans un logiciel de traitement de texte en tant que graphique. Sinon, utilisez l'option *Affichage | Données sous forme de fichier texte* qui enregistre les données de votre schéma (chaque mot est suivi du nombre total de mots dans le fichier et de la position de chaque occurrence).


La [règle](#) du menu  permet de voir le schéma divisé en 8 segments de même taille s'il est basé sur un seul fichier, ou les divisions du fichier texte s'il y'en a plusieurs.

Rubriques connexes : [Schéma des mots-clés](#), [valeur de dispersion du schéma](#)

7.21 effectuer un regroupement de synonymes

Procédure

Lorsque vous estimez que deux groupes peuvent être associés en raison d'une similarité sémantique, il vous suffit de les faire glisser pour les regrouper.

Autrement, si vous cliquez sur , **KeyWords** indiquera les deux groupes de synonymes qui correspondent le mieux. S'affiche une liste des mots trouvés dans l'un des deux groupes, une liste de ceux trouvés dans l'autre, et, au centre, une liste des mots qui correspondent. C'est à vous de décider si la correspondance est suffisamment bonne pour constituer un groupe fusionné.

En cas de doute, cliquez sur **Annuler**.

Pour les regrouper, cliquez sur **Regrouper**.

Si vous êtes certain de ne **pas** souhaiter les regrouper et ne voulez pas que **KeyWords** propose à nouveau cette paire, cliquez sur **Ignorer**. **KeyWords** peut ignorer jusqu'à 50 paires. Pour ôter de la mémoire les éléments à ignorer, cliquez sur **Effacer les ignorés**.

Le but (2)

[Scott](#) (1997) indique comment le regroupement révèle les différents rôles des femmes dans un ensemble d'articles issus du journal *Guardian*.

Rubriques connexes : [groupes](#)

7.22 re-trier : KeyWords

Procédure...

Pour trier, cliquez simplement sur la première ligne d'une liste. Vous pouvez également appuyer sur F6 ou Ctrl+F6, ou encore passer par l'option de menu. Cliquez à nouveau pour basculer entre tri croissant et décroissant.

Une **liste de mots-clés** peut être triée par :

pertinence (le mot le plus pertinent se trouve en haut)
ordre alphabétique (de A à Z)

fréquence dans la petite liste (les mots les plus fréquents sont les premiers)
fréquence dans la liste de référence (les mots les plus fréquents sont les premiers)
Un **schéma de mots-clés** permet le tri par
pertinence (le mot le plus pertinent se trouve en haut)
ordre alphabétique (de A à Z)
fréquence (les mots les plus fréquents sont les premiers)
nombre de liens (les mots les plus liés apparaissent en premier)
première mention de chaque mot-clé dans le texte
rang (les mots des plus petites sections apparaissent en premier)
Une **une base de données de mots-clés clés** permet de trier par :
fréquence (les mots- *clés clés* les plus pertinents apparaissent en haut)
ordre alphabétique (de A à Z)
Une liste d'[associés](#) permet de basculer entre le tri par :
fréquence (association du mot titre et de l'élément)
ordre alphabétique (de A à Z)
fréquence (association de l'élément et du mot titre)

7.23 l'écran des mots-clés

Les éléments suivants s'affichent :

1. chaque mot-clé
2. sa fréquence dans les textes sources dans lesquels ces mots-clés sont clés, en *italique*.
3. le nom du fichier texte source (ou le nom du fichier de liste de mots s'il y en a plusieurs) et le %, également en *italique*.
4. sa fréquence dans le corpus de référence
5. le nom du corpus de référence (ou le nom du fichier de liste de mots du corpus s'il est basé sur plusieurs textes) et le %
6. pertinence ([statistiques](#) du chi-deux ou du log de vraisemblance)
7. [p valeur](#).

Pour calculer dans quelle mesure la fréquence est inhabituelle, on se base sur le [calcul statistique](#) effectué. Les statistiques s'affichent sur la droite. Si la procédure est le log de vraisemblance, ou si la technique du chi-deux est utilisée et que les conditions habituelles du chi-deux sont remplies (la valeur doit être supérieure ou égale à 5 dans les quatre cellules), la probabilité (p) sera affichée à droite de la valeur du chi-deux.


Le critère de fréquence est basé sur la valeur de probabilité sélectionnée avant que les mots-clés ne soient calculés. Plus cette valeur est faible, plus le nombre de mots-clés affichés sera faible.

On ne traite généralement pas plus de 40 mots-clés.

Les mots sont [triés](#) en fonction de la pertinence de leur fréquence d'occurrence. Ceux qui apparaissent en haut sont les plus fréquents. À la fin de la liste, dans une couleur différente, se trouvent les mots très [peu fréquents](#) (mots-clés négatifs).

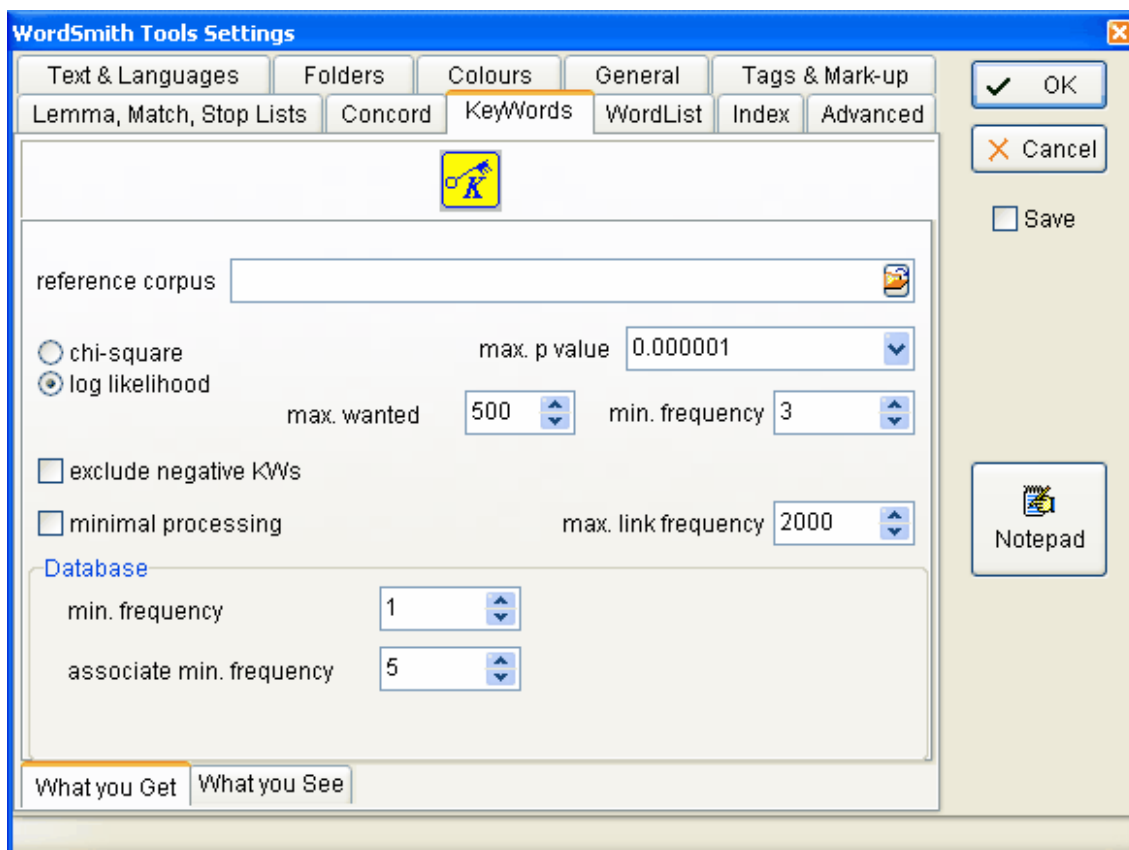
bouton de visualisation

Cela vous permet de voir le texte source à l'aide de [Viewer](#), qui mettra les mots-clés en surbrillance.

Accédez à la section [mise en page](#)  pour changer la couleur ou la police de chaque colonne de données, par exemple si vous n'aimez pas le texte en italique.

7.24 contrôleur WordSmith : Paramètres de KeyWords

Ces derniers sont accessibles via le [contrôleur](#) dans *Modifier les paramètres | KeyWords*.



La raison derrière tout ça est que certains choix effectués peuvent affecter d'autres outils. KeyWords et WordList utilisent tous deux des routines similaires : KeyWords utilise ces routines pour calculer les mots-clés d'un fichier texte, et WordList pour [comparer des listes de mots](#).

Procédure

Chi-deux ou log de vraisemblance. Le log de vraisemblance est la procédure par défaut. Reportez-vous à la rubrique [procédure](#) pour obtenir de plus amples informations.

p valeur max.

Il s'agit du niveau de signification. Reportez-vous à la rubrique [p valeur](#) pour de plus amples informations.

Valeur max. souhaitée (500) et fréquence min. (3)

Vous pouvez limiter le nombre de mots-clés identifiés afin de trouver par exemple les 10 mots-clés les plus importants pour chaque texte. Le programme identifie tous les mots-clés, les trie par ordre de pertinence, et ensuite, il ignore les autres. Ainsi, les [mots-clés positifs](#) sont favorisés par rapport aux négatifs.

La fréquence minimum est un paramètre permettant d'aider à éliminer tous les mots ou clusters qui sont inhabituels et peu fréquents. Par exemple, un nom propre tel que le nom d'un village sera extrêmement peu fréquent dans votre corpus de référence, et s'il n'est mentionné qu'une fois dans le texte que vous analysez, il y a toutes les chances qu'il n'apparaisse pas comme mot-clé. Le paramètre par défaut de 3 mentions minimum permet de réduire le nombre de

résultats non pertinents. Dans le cas de textes courts, comportant moins de 600 mots, on applique automatiquement un minimum de 2.

Exclure les mots-clés à valeur négative

Si cette option est cochée, KeyWords ne calcule pas les mots-clés négatifs (ceux qui se distinguent par leur *manque* de fréquence).

Traitement minimal

Si cette option est cochée, KeyWords ne calcule pas les [schémas](#), [liens](#) ou [clusters de mots-clés](#) lorsqu'il calcule les mot-clés (ils peuvent toujours être calculés ultérieurement, pourvu que vous ne supprimiez ni ne déplaciez les fichiers texte originaux). Cette option s'avère utile pour calculer un ensemble de fichiers KeyWords dans un lot, par exemple pour créer une base de données.

Fréquence max. de lien

Calculer un schéma est un processus lourd car il faut faire une concordance de tous les mots-clés pour déterminer où ils apparaissent. Calculer les liens entre les mots-clés est un processus plus compliqué encore et peut prendre beaucoup de temps, surtout s'il y a parmi vos mots-clés des mots qui apparaissent des centaines ou des milliers de fois dans le texte. Pour que ce processus soit gérable, vous pouvez définir un paramètre par défaut. Ici, 2 000 signifie que tout mot-clé apparaissant plus de 2 000 fois dans le texte ne sera pas utilisé pour le calcul de [liens](#). Cependant, ils apparaissent tout de même dans les schémas et les listes de mots-clés.

Base de données : fréquence minimum

La valeur par défaut est 1. Reportez-vous à la section [base de données](#).

Base de données : fréquence minimum des associés

La valeur par défaut est 5. Reportez-vous à la section [associés](#).

Rubriques connexes : Sommaire de l'aide de KeyWords, [Calcul KeyWords](#).

WordSmith Tools

WordList

Section

VIII

8 WordList

8.1 but



Ce programme permet de générer des listes de mots à partir d'un ou plusieurs fichiers texte [ASCII](#) ou [ANSI](#). Les listes de mots sont générées automatiquement par ordre alphabétique et par ordre de fréquence. Vous pouvez également générer une liste d'[index de mots](#).

Le but

On utilise les listes de mots pour :

- 1 étudier le type de vocabulaire employé ,
- 2 identifier les [clusters](#) de mots les plus courants,
- 3 comparer la fréquence d'un mot au sein de différents fichiers texte ou genres,
- 4 comparer les fréquences de mots apparentés ou d'équivalents de traduction entre [différentes langues](#).

Dans **WordList**, vous pouvez comparer deux [listes](#) ou effectuer une analyse de cohérence ([simple](#)) ou ([détaillée](#)) pour effectuer des comparaisons de style.

Vous pouvez également utiliser ces listes de mots en tant qu'entrées dans le programme [KeyWords](#) qui analyse les mots dans un texte donné et compare les fréquences avec un corpus de référence, afin de générer des listes de mots-clés et de mots-clés clés.

Rubriques connexes : [Affichage WordList](#)

8.2 regroupement automatique de lemmes

L'option de menu *Regroupement automatique* peut être utilisée pour spécifier une chaîne telle que `s;ED;ING`. Le programme parcourt ensuite toute la liste de mots et lemmatise toutes les entrées où un mot diffère du suivant par le fait qu'il comporte `s` ou `ED` ou `ING` à la fin. Le caractère `;` permet de séparer plusieurs suffixes.

Préfixe / Suffixe / Infixe

Par défaut, toutes les chaînes saisies sont supposées être des suffixes ; pour regrouper des préfixes, tapez un astérisque (*) à droite de la fin du préfixe. Si vous souhaitez rechercher des infixes (par ex. `bloody` dans `absobloodylutely` en anglais), tapez un astérisque à chaque extrémité du mot.

Exemples

```
s;ED;ING regroupera books avec book, booked avec book et booking avec book
s;*ED;*ING regroupera books avec book, booked avec book et booking avec book
UN*;ED;ING regroupera undo avec do, booked avec book et booking avec book
*BLOODY* regroupera absobloodylutely avec absolutely
```

Il vous est possible de laisser le processus s'exécuter rapidement et automatiquement, mais il vous est également possible de paramétrer le programme pour que confirmation vous soit demandée avant chaque regroupement. La lemmatisation automatique, tout comme la fonction rechercher et remplacer, peut parfois poser des problèmes et donner des résultats incongrus. Pour stopper le processus en plein milieu d'un regroupement automatique, appuyez sur Échap.

Astuce

Essayez, avec une liste précédemment enregistrée, de procéder à un regroupement

automatique *sans* confirmation des modifications (ou cliquez sur *Oui à tous*). Sélectionnez ensuite la version alphabétique de la liste (par opposition à un classement par fréquence) et triez les lemmes en cliquant sur l'en-tête *Lemmes*). Toutes les entrées jointes s'afficheront en haut de la liste. Il peut s'avérer plus simple et plus pratique de [dissocier](#) (Ctrl+F4) les entrées regroupées par erreur plutôt que de confirmer chaque regroupement. Pour finir, triez par *mot*, puis enregistrez.

Rubriques connexes : [Lemmatisation](#)

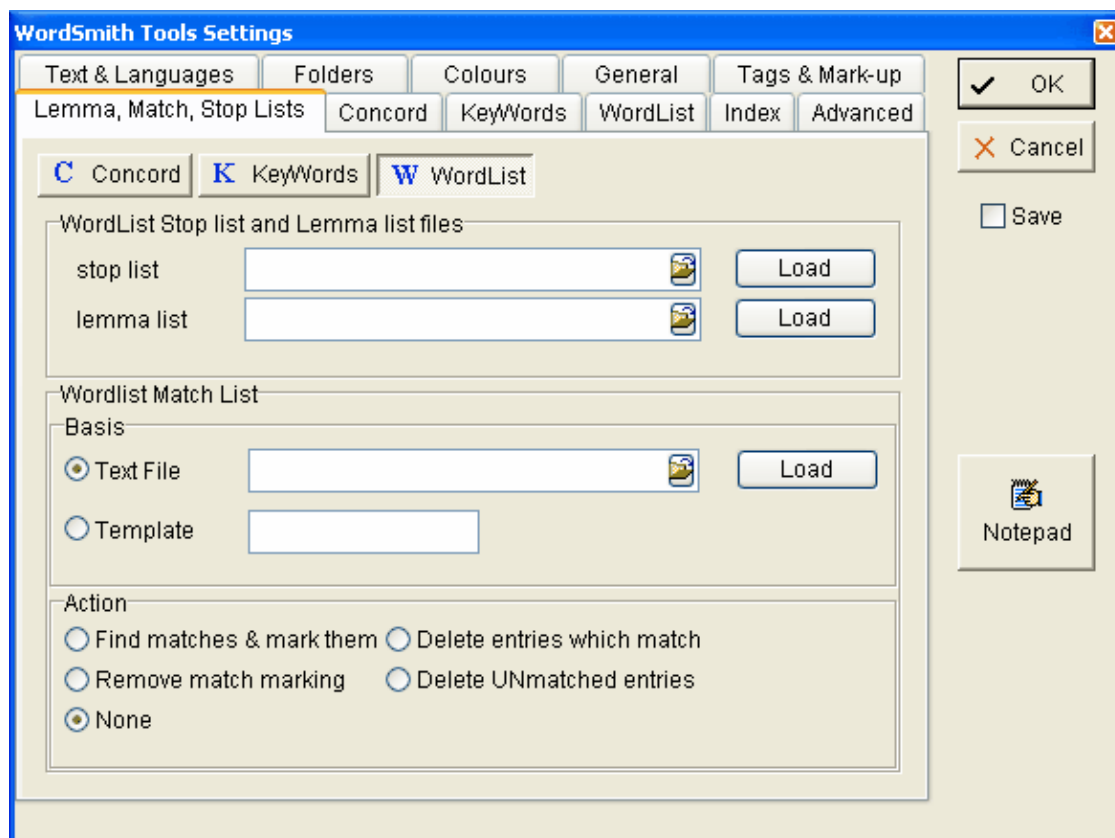
8.3 sélection d'un fichier de lemmes

Le but

Vous pouvez lemmatiser tous les éléments de la liste de mots courante en utilisant un fichier texte standard regroupant tous les mots qui vont ensemble (être -> était, est, étaient etc). Créer le fichier texte prend du temps, mais il est très utile pour lemmatiser de nombreuses liste de mots et est moins aléatoire que le [regroupement automatique](#).

Procédure

Dans le contrôleur principal, *Paramètres | Modifier les paramètres | Listes*, l'écran suivant s'affiche :



Cliquez sur le bouton approprié (pour Concord, KeyWords ou WordList) et saisissez le nom du fichier ou utilisez parcourir pour le sélectionner.

Ce fichier contient une liste de lemmes au format texte brut, avec des éléments tels que :

ÊTRE -> SUIS, SONT, ÉTAIT, ÉTAIENT, EST
ALLER -> VA, VONT, ALLA, ALLÈRENT

Le contrôleur lit les fichiers et affiche les lemmes (ou un échantillon si la liste est trop longue).

Le calcul des listes a lieu lorsque vous cliquez sur l'option de menu *Faire correspondre les lemmes* (≡) dans WordList, Concord ou KeyWords. Pour des explications plus complètes avec des captures d'écran, reportez-vous à la section [Liste de correspondance](#).

Que se passe-t-il si mes fichiers texte ne contiennent pas ÊTRE ?

Supposons que vous fassiez correspondre SUIS, SONT etc. avec ÊTRE comme dans la liste ci-dessus, et que vos texte ne contiennent pas le mot ÊTRE. WordList ne le trouvera donc pas. Pour y remédier, il faut créer une [nouvelle liste de mots à partir d'un fichier de texte brut](#) (dans lequel vous incluez ÊTRE ainsi que les autres formes infinitives désirées), l'enregistrer, et le [fusionner](#) avec votre liste de mots. WordList trouve alors la forme ÊTRE et l'ajoute à SUIS, SONT, ÉTAIENT, etc.

Rubriques connexes : [Lemmatisation](#), [Liste de correspondance](#), [Liste d'arrêt](#)

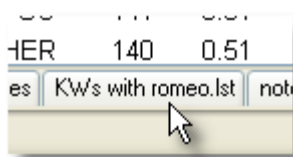
8.4 comparaison de listes de mots

Cette fonctionnalité permet de faciliter les comparaisons stylistiques. Supposons par exemple que vous étudiez plusieurs versions d'une histoire, ou ses différentes traductions. Si une version emploie *tuer* et qu'une autre emploie *assassiner*, vous pouvez utiliser cette fonction.

Cette procédure compare tous les mots dans les *deux* listes. Elle indique quels sont ceux qui apparaissent radicalement plus souvent dans l'un que dans l'autre, y compris ceux qui apparaissent plus que le nombre minimum de fois spécifiées dans un texte, même s'ils n'apparaissent pas dans l'autre.

Procédure

1. Ouvrez une liste de mots.
2. Dans le menu, cliquez sur *Fichier | Comparer 2 listes de mots*.
3. Sélectionnez une liste de mots servant à la comparaison. Les résultats s'affichent dans un des onglets de la partie inférieure de l'écran.



La fréquence minimum (que vous pouvez modifier dans le [contrôleur](#), via l'onglet *Modifier les paramètres, KeyWords*) peut être définie sur 1. Si la valeur est augmentée à 3, la comparaison ignore les mots qui n'apparaissent pas au moins trois fois dans au moins une des deux listes. Choisissez la valeur significative (*tout*, ou une [p valeur](#) allant de 0,1 à 0,000001 ou toute autre valeur de votre choix). Plus la [p valeur](#) est petite, plus la comparaison est sélective. En d'autres termes, une valeur de 0,1 permettra d'afficher plus de résultats qu'une valeur de 0,0001.

Le format d'[affichage](#) est similaire à celui utilisé dans [KeyWords](#).

Rubriques connexes : [Analyse de cohérence](#), [Liste de correspondance](#)

8.5 fusion des listes de mots

Le but

Vous pouvez fusionner 2 listes de mots (ou de concordances, des listes d'informations mutuelles, etc) si le traitement de chacune d'entre elles est trop long ou si vous établissez au fur et à mesure une liste de mots ou concordance principale basée sur un certain nombre de genres ou de types de textes.

Procédure

Ouvrez une liste de mots (ou concordance), cliquez sur *Fichier | Fusionner avec* et sélectionnez-en une deuxième.

Attention

Fusionner des listes de mots nécessite que les ensembles de textes source soient différents. Si vous décidez de fusionner deux listes de mots contenant des informations provenant du même fichier texte, WordSmith le fera même si le nombre d'occurrences et de textes dans lesquels chaque type de mots a été trouvé est (vraisemblablement) inexact.

Fusionner une liste en anglais avec une liste en espagnol : si vous commencez par la liste en espagnol, la liste en anglais sera fusionnée dans la première et sera, par conséquent, traitée comme si c'était de l'espagnol, par exemple au niveau du tri. Si vous tentez de fusionner une liste en anglais dans une liste en arabe (je n'ai jamais essayé), vous pourrez vraisemblablement afficher toutes les formes, mais vous obtiendrez un résultat différent en fusionnant la liste en arabe dans celle en anglais (tous les mots en arabe seraient considérés comme des mots anglais).

8.6 affichage de comparaison

Voici une fenêtre de comparaison. Nous avons comparé deux oeuvres de Shakespeare : *Le Roi Lear* et *Roméo et Juliette*.

Les éléments suivants s'affichent :

la fréquence d'apparition dans le texte de référence, ici *Le Roi Lear*, (le pourcentage ne s'affiche que s'il est supérieur à 0,01 %)

la fréquence d'apparition dans l'autre texte, c'est-à-dire *Roméo et Juliette*, (le pourcentage ne s'affiche que s'il est supérieur à 0,01 %)

[chi-deux ou log de vraisemblance](#), et la [valeur](#) p.

Le critère de fréquence est basé sur la valeur de probabilité saisie avant que les listes ne soient comparées. Plus cette valeur de probabilité est faible, plus le nombre de mots affichés sera faible.

Les mots apparaissent triés en fonction de leur fréquence d'occurrence. Ceux situés en haut apparaissent très fréquemment dans votre liste de mots principale. En bas de la liste, vous trouvez ceux dont la fréquence d'apparition est très faible dans le premier texte choisi : ce sont des mots-clés dans le second texte.

Cette comparaison est similaire à l'analyse de "mots-clés" avec le programme [KeyWords](#). Cependant, l'analyse avec KeyWords est légèrement plus rapide et permet un traitement par lots.

Le mot **Lear** est le plus pertinent. Sa valeur est de 304 dans la colonne de pertinence. (04,56 s'affiche sur la capture d'écran ci-dessous car la colonne n'a pas été élargie).

N	Key word	Freq.	%	Freq.	RC. %	yness	P	emmas
1	LEAR	229	0.83	0		04.56	00000	
2	KENT	173	0.63	0		29.91	00000	
3	FOOL	119	0.43	6	0.02	32.44	00000	
4	KING	70	0.25	2		77.51	00000	
5	EDMUND	52	0.19	0		69.00	00000	
6	GLOUCESTER	50	0.18	0		66.34	00000	
7	YOUR	223	0.81	103	0.40	53.21	00000	
8	EDGAR	37	0.13	0		49.08	00000	

frequency alphabetical statistics filenames KWs with romeo.lst notes
4,213 Type-in LEAR

Les mots en noir, ci-dessus, sont les mots-clés de Lear. Ci-dessous se trouve le milieu de la liste. Les mots en rouge sont les mots-clés de Romeo. Le mot "most" est le mot-clé le moins pertinent pour "Lear", et "death" le mot le moins pertinent pour Romeo ; tous deux ont une valeur de pertinence avoisinant 25 (positive ou négative).

N	Key word	Freq.	%	Freq.	RC. %	yness	P	emmas
28	DAUGHTERS	28	0.10	2		25.34	04780	
29	TOM	19	0.07	0		25.20	05142	
30	HAVE	207	0.75	127	0.49	24.72	06596	
31	MOST	54	0.20	16	0.06	24.21	08608	
32	DEATH	22	0.08	71	0.27	24.29	08253	
33	BALTHASAR	0		19	0.07	25.20	05142	
34	GREGORY	0		19	0.07	25.20	05142	
35	MARRIED	0		19	0.07	25.20	05142	

frequency alphabetical statistics filenames KWs with romeo.lst notes
4,213 Type-in LEAR

Au bas de la liste s'affichent les mots-clés les plus pertinents pour la pièce *Roméo et Juliette*.

N	Key word	Freq.	%	Freq.	RC. %	yness	P	emmas
53	TYBALT	0		71	0.27	94.23	00000	
54	BENVOLIO	0		80	0.31	06.19	00000	
55	MERCUTIO	0		83	0.32	10.17	00000	
56	FRIAR	0		100	0.38	32.77	00000	
57	NURSE	1		149	0.57	87.38	00000	
58	CAPULET	0		145	0.56	92.63	00000	
59	JULIET	0		178	0.68	36.57	00000	
60	ROMEO	0		296	1.14	94.02	00000	

Le mot-clé le plus pertinent est Romeo, dont la valeur de pertinence est 394 (la colonne doit être élargie pour que les valeurs soient lisibles).

8.7 analyse de cohérence (détaillée)

Cette fonction est identique à la [cohérence simple](#), mais elle donne plus de détails.

Le but

Cette fonctionnalité permet de faciliter les comparaisons stylistiques. Supposons par exemple que vous étudiez plusieurs versions d'une histoire, ou ses différentes traductions. Cette fonction vous permet d'afficher tous les mots utilisés dans la liste de mots que vous avez appelée. L'affichage classe les mots de façon à ce que le premier groupe contienne tous ceux qui apparaissent dans toutes les versions, puis ceux qui apparaissent dans toutes les versions sauf une, et ainsi de suite jusqu'aux mots n'apparaissant que dans une seule version.

N	Word	Freq.	Texts	s	t	00.txt	0a.txt	0w.txt	01.txt	1d.txt	1e.txt	1f.txt
1	A	1,151	7			135	174	141	206	116	214	165
2	ABOUT	77	7			13	5	5	16	7	20	11
3	ACT	16	7			1	2	1	1	7	1	3
4	AFTER	53	7			5	8	13	5	3	13	6
5	AGAIN	36	7			4	1	12	2	2	9	6
6	ALL	120	7			14	9	34	23	7	11	22
7	ALSO	89	7			16	9	21	11	5	17	10


Au sein de chaque groupe, les mots sont classés par ordre alphabétique. La colonne Fréq. affiche le nombre de fois où chaque mot apparaît dans l'ensemble des textes, et la colonne

Textes affiche le nombre de textes dans lesquels chaque mot apparaît. Il y a ensuite deux colonnes, Nbre de lemmes et Définir (de la même manière que dans une liste de mots), puis une colonne pour chaque texte. Dans notre exemple, le mot **about** apparaît dans les 7 textes, 77 fois au total. Sa plus grande fréquence d'apparition est dans **1e.txt** où il apparaît **20** fois. En cliquant sur les onglets du bas, vous pouvez afficher les statistiques ainsi que les noms des 7 fichiers texte utilisés ici. Les notes peuvent être éditées et enregistrées en même temps que la liste de cohérence détaillée.

Remarque : le nom du fichier est **test.dcl** (**detailed consistency list** = liste de cohérence détaillée).

Il n'y a aucune limite, sauf bien sûr la limite de mémoire disponible, c'est à dire le nombre de fichiers pouvant être traités par cette procédure.

Procédure

Lorsque vous cliquez sur **Nouveau...** () dans la fenêtre courante, un onglet apparaît, affichant une cohérence détaillée. Sélectionnez vos listes de mots puis cliquez sur **calculer la cohérence détaillée**.

Chaque colonne peut être triée en cliquant sur son en-tête (**Mot**, **Freq.** etc.). Pour obtenir les mots qui apparaissent dans les 7 textes en haut, j'ai cliqué sur **Textes**.

Rubriques connexes : [Analyse de cohérence \(Simple\)](#), [Affichage de comparaison](#), [Comparaison de listes de mots](#), [Liste de correspondance](#), [Totaux des colonnes](#)

8.8 analyse de cohérence (simple)

Cette fonction (que [Paul Nation](#) a appelée "rang") est automatiquement proposée pour toute liste de mots.

Chaque liste de mots comporte une colonne "Textes". Elle indique, pour chaque mot, le nombre de textes dans lesquels il apparaît (le maximum étant le nombre total de fichiers texte utilisés pour la liste de mots).

Le but

Il s'agit de déterminer quels mots se répètent de façon cohérente dans des textes d'un genre donné. Supposons que le mot **consolider** apparaisse dans de nombreux ensembles de rapports d'entreprise annuels. Il n'apparaît pas très souvent dans chacun des rapports, mais apparaît de manière plus cohérente dans les rapports d'entreprise que dans un ensemble de textes quelconque.

Des mots comme **le** sont cohérents dans la plupart des textes français.

Pour analyser des mots courants comme **le**, il peut être utile de vous servir d'une liste de cohérence. Certains éléments lexicaux peuvent ici encore se répéter de façon significative.

Il s'agit ensuite de d'éliminer ceux qui se répètent souvent dans de nombreux textes et de déterminer lesquels sont caractéristiques du genre ou sous-genre de vos textes en recherchant lesquels ont une cohérence significative. Enregistrez votre liste de mots et utilisez la pour effectuer une [comparaison](#) avec d'autres listes de mots WordList, ou listes de mots-clés Keywords.

Rubriques connexes : [Analyse de cohérence \(détaillée\)](#), [Comparaison de listes de mots](#), [Liste de correspondance](#)

8.9 lemmes

Il peut être utile d'enregistrer plusieurs entrées en même temps : par exemple **want**; **wants**; **wanting**; **wanted**, qui font partie du même lemme.

Regroupement manuel

Pour ce faire, il suffit de faire glisser une entrée vers l'autre. Si votre liste de mots comporte :

WANT
WANTED
WANTING

saisissez simplement **wanting** ou **wanted** avec votre souris et placez-le sur **want**.

Reportez-vous à la section [sélectionner le fichier de lemmes](#) pour les regrouper avec un mot ne se trouvant pas dans la liste.

Les listes alphabétiques ainsi que les listes de fréquence seront correctement mises à jour, bien que la liste de fréquence n'indique pas le bon ordre tant que vous n'avez pas re-trié le fichier en utilisant [supprimer les entrées](#). Une entrée d'en-tête lemmatisée est marquée en rouge dans la marge. Les autres entrées marquées seront mises en couleur comme si vous les aviez supprimées. Les entrées liées qui ont été regroupées avec l'en-tête sont visibles à droite.

N	Word	Freq.	%	Texts	%	Lemmas	Set
23	A GAME OF	5		4	25.00		
24	A GOD UNKNOWN	6		1	6.25		
25	A GOOD DEAL	15		1	6.25	a good deal[5] a great deal[10]	
26	A GREAT DEAL	10		5	31.25		

Voici une liste de mots basée sur des [clusters de 3 mots](#) où **a good deal** avait à l'origine une fréquence de 5, mais a été lié à **a great deal** et a par conséquent augmenté de 10.

Pour voir tous les éléments que vous souhaitez regrouper sur un même écran, utilisez les clés de fonction.

1. Appuyez sur F5 pour marquer une entrée à grouper avec une autre. La première entrée marquée sera "head". Tant que vous n'aurez pas décidé quelles sont les autres entrées à regrouper, le bord de cette ligne s'affichera en vert. Toute les entrées que vous souhaitez lier à l'en-tête (en appuyant sur F5) seront marquées en blanc. Si vous changez d'avis, appuyez à nouveau sur F5 pour faire disparaître le marquage.
2. Appuyez sur F4 pour grouper toutes les entrées que vous avez marquées. Le programme place ensuite les fréquences communes de tous les mots que vous avez marqués avec celui que vous avez marqué en premier (en-tête).

Pour dissocier

Pour dissocier des entrées, cliquez sur un élément dont les lemmes sont visibles à droite et appuyez sur Ctrl+F4.

Regroupement à partir d'un fichier

Il est également possible de regrouper les lemmes en utilisant un [fichier texte](#) qui automatise le

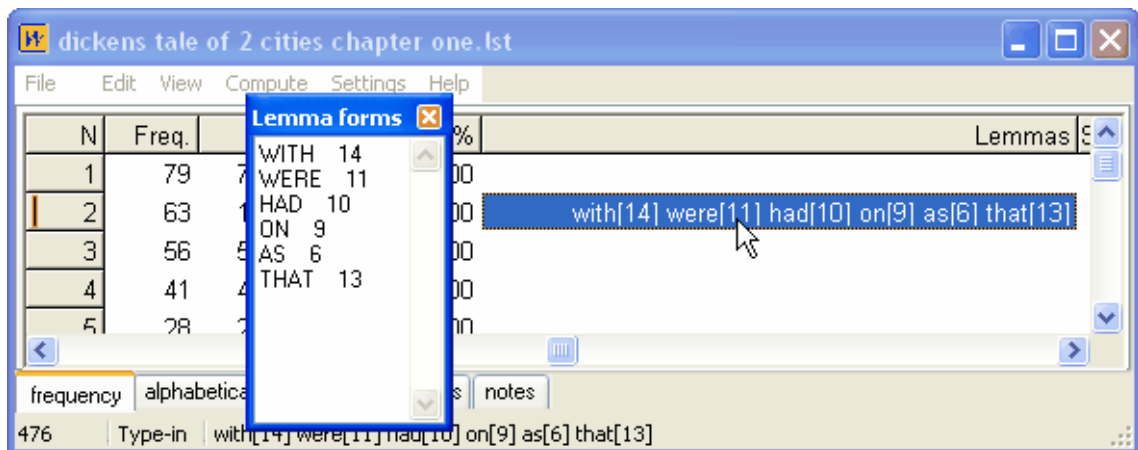
processus de correspondance et de regroupement. Le calcul des listes a lieu lorsque vous cliquez sur l'option de menu *Faire correspondre les lemmes* (☰) dans WordList, Concord ou KeyWords. Chaque entrée de votre liste de lemmes est vérifiée pour déterminer si elle correspond à une des entrées de votre liste de mots. Dans l'exemple, si **am**, **was** et **were** apparaissent, ils seront enregistrés en tant que lemmes de **be**. Si **go** et **went** apparaissent, alors **went** sera groupé avec **go**.

Groupement automatique

Pour accélérer le processus de lemmatisation, vous pouvez effectuer un regroupement automatique de toutes les entrées de votre liste active de mots qui répondent à vos critères de sélection.

Impossible de lire tous les formats de lemmes

Cliquez deux fois sur la colonne des lemmes, comme sur la capture d'écran ci-dessous.



Ainsi, une fenêtre avec les formes de lemmes s'affichera, indiquant les différents composants.

Rubriques connexes : [Regroupement automatique](#), [Utiliser un fichier texte pour la lemmatisation](#)

8.10 listes d'index : utilisation

le but

1. Un index peut entre autres servir à enregistrer les positions de tous les mots de votre fichier texte, pour que vous puissiez voir ultérieurement quel mot apparaît dans quelle partie de chaque texte. L'index permet également d'avoir accès à ces mots plus rapidement, par exemple pour la concordance. Si vous sélectionnez un ou plusieurs mots de l'index et cliquez sur **C**, vous obtenez rapidement une concordance.
2. Une autre utilisation de l'index est de calculer les scores d'"[informations mutuelles](#)".
3. Vous pouvez aussi utiliser un index pour voir les [clusters de mots](#).

Rubriques connexes : [Création d'une liste d'index](#), [Affichage des listes d'index](#), Sommaire de l'aide de WordList.

8.11 listes d'index : affichage

Ouvrez un index dans WordList comme n'importe quelle liste de mots, en cliquant sur Fichier | Ouvrir. Ou, plus simple encore, cliquez sur contrôleur | Listes précédentes et cliquez deux fois sur l'un des index que vous avez créés. L'index apparaîtra comme si c'était une grande liste de mots.

N	Word	Freq.	%	Texts	% emmas	Se
1	THE	57,881	5.90	4,038	99.61	0
2	OF	52,923	2.97	4,037	99.58	0
3	AND	27,839	2.56	4,037	99.58	0
4	TO	05,498	2.54	4,037	99.58	0
5	#	15,482	2.16	2,956	72.92	0
6	A	90,736	2.13	4,039	99.63	0
7	IN	57,924	1.91	4,042	99.70	0
8	THAT	21,060	1.09	4,016	99.06	0
9	IT	57,112	1.03	3,973	98.00	0
10	IS	92,636	0.97	4,016	99.06	0



frequency alphabetical statistics filenames notes

126,550 Type-in

La capture d'écran ci-dessus montre les mots du top 10 du BNC World Corpus. Le numéro 5 (#) représente les nombres ou les mots qui contiennent des nombres comme 50 €. Ces mots très fréquents sont également très cohérents : ils apparaissent dans au moins 99% des 4 054 textes de BNC World.

Sur la capture d'écran ci-dessous, vous voyez les mots classés en fonction du nombre de textes : tous ces mots apparaissent 10 fois dans le corpus, mais leurs fréquences sont différentes.

N	Word	Freq.	%	Texts	%	emmas
64,857	LACANIAN	49	0.00	10	0.25	0
64,858	LACTATE	27	0.00	10	0.25	0
64,859	LAINE	16	0.00	10	0.25	0
64,860	LAKOFF	51	0.00	10	0.25	0
64,861	LAMBIE	29	0.00	10	0.25	0
64,862	LAMINAR	52	0.00	10	0.25	0
64,863	LAMPLIGHTER	15	0.00	10	0.25	0
64,864	LANDFORM	35	0.00	10	0.25	0
64,865	LANDLORDISM	19	0.00	10	0.25	0
64,866	LANDSLIPS	15	0.00	10	0.25	0

Vous pouvez mettre en surbrillance ou marquer un ou plusieurs mots en cliquant sur , puis sur  pour obtenir une concordance rapide.

Rubriques connexes : [Création d'une liste d'index](#), [Clusters de WordList](#), Sommaire de l'aide de WordList.

8.12 création d'un index WordList

Pour créer un index, utilisez le [contrôleur](#) et cliquez sur *Modifier les paramètres | Index*. Vous devez spécifier un nom de fichier pour l'index.

WordList parcourt ensuite les textes que vous avez sélectionnés et garde en mémoire les informations relatives à la position de chaque exemple de chaque type de mots.

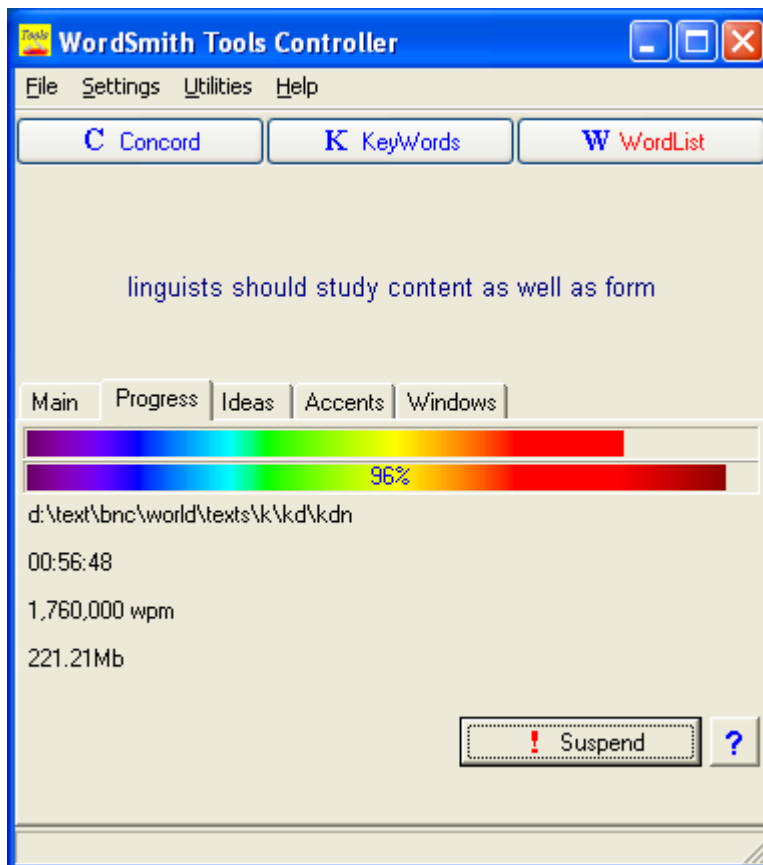
fichiers d'index

Deux fichiers sont créés pour chaque index :

fichier `.tok` : gros fichier contenant les informations relatives à la position de chaque unité lexicale d'un mot dans vos fichiers texte.

fichier `.typ` : connaît tous les types de mots.

Un index permet de calculer les [clusters de mots](#) et les scores des [information mutuelles](#) pour chaque type de mots. La capture d'écran ci-dessous montre les barres de progression pour un index du BNC World corpus. Il a fallu près d'une heure pour effectuer 96% du travail, à un taux de 1,8 millions de mots par minute. Le fichier `BNC words.tok` qui en résulte pèse 1,6 Go et le fichier `BNC words.typ` pèse 26 Mo.



ajouter à un index

Pour ajouter à un index existant, sélectionnez simplement quelques textes supplémentaires et cliquez sur Fichier | Nouveau | Index. Si le nom de fichier est déjà utilisé par un autre index, vous devrez indiquer si vous souhaitez y ajouter des données ("Oui") ou recommencer ("Non").

Rubriques connexes : [Utilisation des listes d'index](#), [Affichage des listes d'index](#), Sommaire de l'aide de WordList.

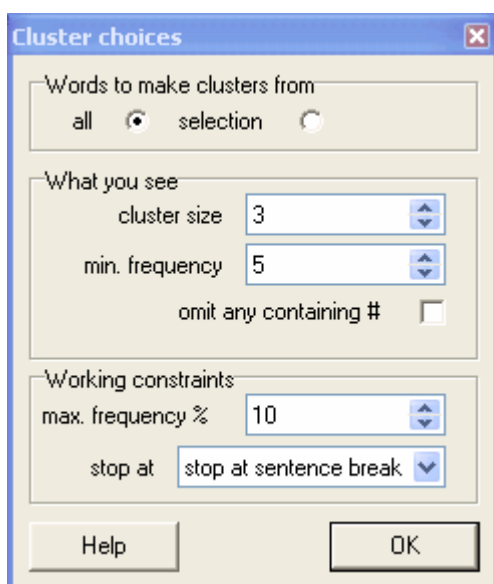
8.13 clusters d'index

Clusters de WordList

Une liste de mots ne se compose pas nécessairement de mots uniques. Vous pouvez créer des listes de mots contenant de 2 à 8 mots par ligne. Pour traiter les clusters dans WordList, il faut d'abord [créer un index](#).

Procédure pour les clusters

[Ouvrez](#) l'index. Cliquez sur *Calculer | Clusters*.



Mots à partir desquels former des clusters

- "tous" : tous les clusters composés de tous les mots ayant un certaine fréquence (cela prendra beaucoup de temps pour un corpus comme celui de BNC World), ou
- "sélection" : uniquement les clusters pour les mots que vous avez sélectionnés (si vous avez par exemple sélectionné BOOK et BOOKS et que vous cherchez les clusters comme **book a table, in my book**).

Pour sélectionner des mots qui ne se trouvent pas les uns à côté des autres, appuyez sur Ctrl et cliquez sur le nombre à gauche, maintenez la touche Ctrl enfoncée et cliquez ailleurs. Le premier sélectionné apparaîtra en vert et les autres en blanc. Dans l'exemple qui suit, qui est basé sur un index du BNC World corpus, les mots **world** puis **life** ont été sélectionnés en cliquant sur les lignes 164 et 167.

N	Word	Freq.	%	Texts	%
161	HOWEVER	59,744	0.06	3,138	77.4
162	MIGHT	59,499	0.06	3,452	85.15
163	ANOTHER	59,196	0.06	3,650	90.03
164	WORLD	59,102	0.06	3,202	78.96
165	YES	58,725	0.06	2,345	57.84
166	PUT	57,809	0.06	3,582	88.36
167	LIFE	57,769	0.06	3,248	80.12
168	AGAIN	57,540	0.06	3,544	87.42
169	PART	56,583	0.06	3,585	88.43
170	NEED	55,399	0.05	3,579	88.26
171	AGAINST	55,337	0.05	3,257	80.34
172	WANT	55,274	0.05	3,213	79.26

frequency alphabetical statistics filenames notes

343,617 Type-in LIFE

Ce processus prend du temps. Dans le cas de BNC World, l'index connaît la position de l'intégralité des 100 millions de mots. Dans l'exemple ci-dessus, il a fallu une minute pour calculer les clusters de 3 mots pour les 115 000 apparitions de **world** et **life** et déterminer 5 719 clusters comme **the world bank** et **of real life**. Il a fallu 36 heures à l'ordinateur de Chris Tribble pour calculer tous les clusters de 3 mots pour l'ensemble du BNC. Il a pu utiliser son PC pendant ce temps, mais ce n'est pas très pratique.

Ce que vous voyez

La taille d'un cluster doit se situer entre 2 et 8 mots.

La "fréquence min." est le nombre minimum pour chacun d'entre eux.

Ici, l'utilisateur a choisi d'afficher tous les clusters de 3 mots qui apparaissent au moins 5 fois.

Contraintes de fonctionnement

Le paramètre "% fréquence max." permet d'accélérer le processus. Il s'agit de la fréquence maximum en % traitée lors du calcul des clusters pour un mot donné. Cela s'explique par le fait qu'il y a de nombreux éléments qui ont une fréquence très élevée et que les clusters qui *commencent* par ces éléments ne sont pas toujours ceux qui vous intéressent. Par exemple, si l'élément **the** représente 6 % dans une liste de mots, les clusters commençant par **the** ne vous intéresseront pas forcément, vous pouvez alors fixer le pourcentage maximum à 0,5 % ou 0,1 % (cela ôte 102 mots de fréquence pour le BNC World corpus). Les clusters contenant des éléments à très haute fréquence au milieu ou à la fin, comme **a** dans **book a table** seront pris en considération, mais pas **in my book**, qui commence par le mot à haute fréquence **in**. Plus vous incluez de mots, plus le processus sera long.

Arrêter à, comme les [clusters Concord](#), présente des contraintes, telles que les marqueurs de fin de phrase et autres ponctuations. En effet, un cluster de 5 mots qui se trouve à cheval sur deux phrases n'est pas très pertinent.

Ce que vous obtenez

N	Word	Freq.	%	Texts	% err
1	EFFECTS OF RABIES	12	0.00	11	0.27
2	IS RABIES AND	11	0.00	11	0.27
3	FROM THE RABIES	7	0.00	5	0.12
4	IS RABIES IN	6	0.00	5	0.12
5	THE RABIES VIRUS	6	0.00	4	0.10
6	SMALLPOX AND RABIES	5	0.00	4	0.10

frequency alphabetical statistics filenames notes

6 Type-in RABIES

Voici quelques clusters de 3 mots issus du BNC World corpus comportant le mot "rabies". Certains d'entre eux sont des unités multi-mots pertinentes. Tous les clusters apparaissant au moins 5 fois sont affichés : pour régler ce paramètre, cliquez sur *Modifier les paramètres / Index* dans le contrôleur et spécifiez une valeur dans "afficher si la fréquence...", comme dans l'exemple ci-dessous :

Index File

file-name E:\WSMITH\wordlist\32\indices\observer

Clusters

size 3 show if frequency at least 5

Rubriques connexes : [clusters dans Concord](#)

8.14 recherche par menu

En utilisant le menu vous pouvez rechercher une sous-chaîne au sein d'une entrée, par exemple tous les mots contenant "constitution" (si vous saisissez ***constitution***, l'astérisque signifie que l'élément peut se trouver au milieu d'un mot, avec ***constitution**, vous trouverez *reconstitution* mais pas *constitutionnellement*, alors que ***constitution*** renverra les deux mots). Ces recherches peuvent être répétées.

Cette fonction vous permet de trouver des parties de mots et donc d'éditer votre liste de mots, par exemple en regroupant deux mots.

Pour trouver la fin ou le milieu d'un mot, on utilise le caractère de remplacement *.

Par conséquent, ***TI*** renverra *anti*, *petit* etc.

ET renverra *muet*, *fluet* etc.

Appuyez sur **F8** pour répéter la dernière recherche.

Les touches de raccourcis pour la recherche sont :

- F8** répéter la dernière recherche (à utiliser en combinaison avec F10 ou F11)
- F10** rechercher à partir de la ligne active
- F11** rechercher jusqu'à la ligne active
- F12** rechercher à partir du début

Cette fonction est utile pour la [lemmatisation](#) (regrouper les mots qui font partie de la même entrée, comme *semble/semblent/semblait/semblant*, etc.)

Rubriques connexes : [recherche d'une entrée par saisie](#)

8.15 scores d'informations mutuelles

le but

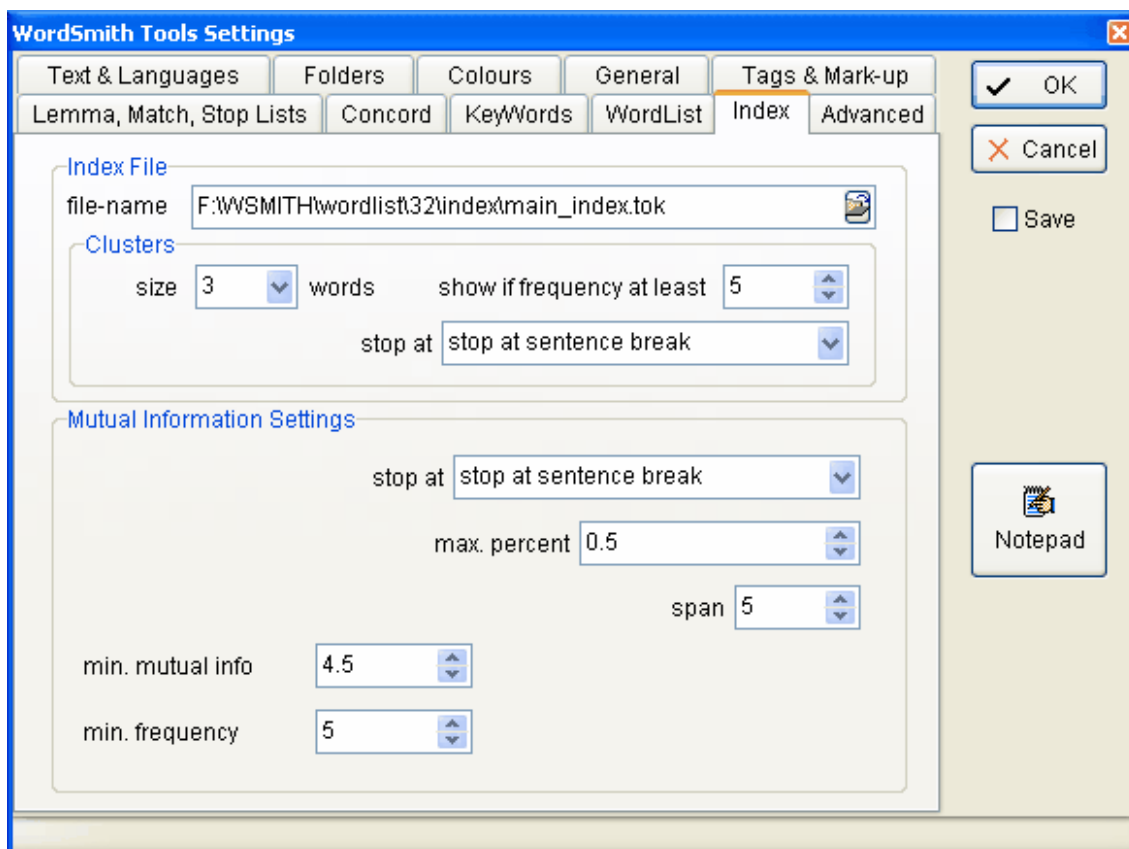
Un score d'Information Mutuelle (IM) met en relation les mots les uns avec les autres. Si *problème* apparaît souvent avec *résoudre*, ils ont certainement des scores d'informations mutuelles élevés. Généralement, on trouve plus souvent *le* à proximité de *problème* que de *résoudre*, c'est pourquoi sont pris en considération lors du calcul des informations mutuelles non seulement les mots qui apparaissent le plus souvent près du mot en question, mais aussi les apparitions de chaque mot ailleurs que près du mot en question. Puisque *le* apparaît fréquemment loin de *problème*, il ne sera pas très lié et aura un faible score d'IM.

Cette relation est bilatérale : dans le cas de *pile* et *face*, aucune distinction n'est établie entre la certitude quasi-totale de trouver *face* près de *pile*, et la probabilité bien moins élevée de trouver *pile* près de *face*.

Il existe différentes formules permettant de calculer le degré des relations de co-occurrences. Les informations mutuelles dans WordSmith ("informations mutuelles spécifiques") sont calculées en utilisant une formule dérivée de Gaussier, Lange et Meunier, décrite dans [Oakes](#), p. 174 ; la probabilité est basée sur la taille totale du corpus en tokens. D'autres aspects de relations collocationnelles sont également calculés, ils sont exposés dans la section [Affichage des informations mutuelles](#).

Paramètres

Les paramètres d'informations mutuelles sont accessibles à partir du [contrôleur](#) sous *Modifier les paramètres* / *Indexage* ou par le biais d'une option de menu dans **WordList**.



arrêter à : vous pouvez indiquer si vous souhaitez prendre en considération les marqueurs de fin ou non. Si c'est le cas, dans "J'ai écrit la lettre. Puis je l'ai postée", **postée** ne serait pas considéré comme un co-occurent de **lettre** car ils sont séparés par un marqueur de fin de phrase.

% max. ignore tous les tokens qui sont plus fréquents que le pourcentage indiqué. Le but est d'éviter de calculer les informations mutuelles pour des mots comme **le** et **de**, qui ont généralement une fréquence très élevée.

étendue : le nombre de mots entre co-occurent et noeud. Avec une étendue de 5, le noeud **écrit** considère **la**, **lettre**, **puis**, **je** et **postée** comme étant des co-occurents potentiels si *arrêter à* est défini sur *aucune limite*.

informations mutuelles min. valeur minimum que doit avoir le score d'IM pour être signalé. 3,0 est une bonne limite. En deçà, le lien entre noeud et co-occurent est généralement faible.

fréquence min. : fréquence minimum que doit avoir un élément pour être pris en considération lors du calcul des informations mutuelles (par défaut = 5). Si un élément n'apparaît qu'une ou deux fois, les informations mutuelles ne seront pas très significatives.

Rubriques connexes : [Affichage des informations mutuelles](#), [Calcul des informations mutuelles](#), [Création d'une liste d'index](#), [Affichage des listes d'index](#), [Sommaire de l'aide de WordList](#).

Référez-vous aux ouvrages de [Oakes](#) pour de plus amples informations sur les informations mutuelles.


8.16 informations mutuelles : calcul

Dans WordList ou Concord

Dans Concord


Par défaut, les informations mutuelles ne sont pas calculées pour une liste de co-occurents. Pour les calculer, il faut une liste de mots qui fournit les données nécessaires. Supposons que vous ayez effectué une concordance de tous les fichiers de `c:\wsmith4\text\shakespeare` et en particulier du mot *amour*. Vous obtenez des co-occurents tels que *Roméo, haine, le, Juliette, nourrice* etc. Leur score IM ou "de relation" est de "??" tant qu'ils ne sont pas calculés.


Si ce n'est pas encore fait, utilisez WordList pour créer une liste de mots avec les mêmes fichiers texte (ou un autre [corpus de référence](#)). Assurez-vous que vous avez spécifié le bon [corpus de référence](#).

Cliquez sur  et Concord utilisera le nom de fichier du corpus de référence. Il recherche tous les co-occurents et calcule les IM à partir du corpus de référence.

Dans WordList

Pour calculer les informations mutuelles (IM), il vous faut un [index WordList](#). Utilisez l'[affichage alphabétique](#) de la liste.

Cliquer sur  vous permet de choisir de calculer les informations mutuelles pour les entrées sélectionnées (en surbrillance) seulement, pour toutes les entrées, ou pour toutes celles comprises entre deux caractères de début, par exemple A et D.

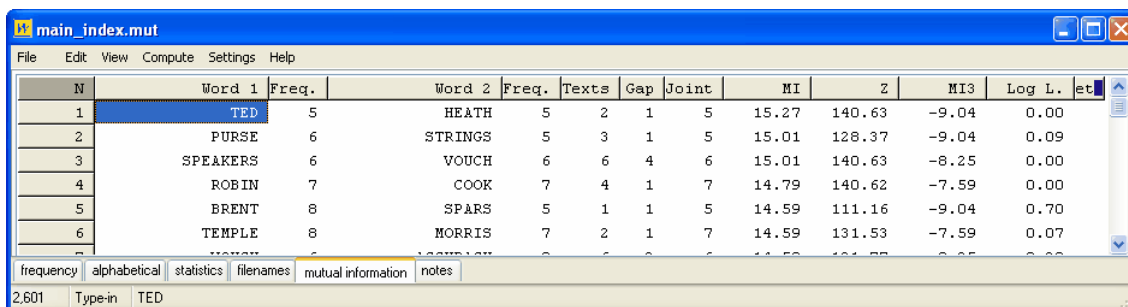
Vous pouvez marquer les éléments pour lesquels vous souhaitez calculer les informations mutuelles (avec ).

Calculer les scores d'informations mutuelles pour chaque entrée d'un index prend beaucoup de temps, soit plusieurs heures pour un index basé sur 10 millions de mots. Pour l'ensemble du BNC, soit 100 millions de mots, cela peut prendre 24 heures, même avec un ordinateur performant. Pensez à enregistrer vos résultats. Vous avez la possibilité de ne le faire que pour une partie de la liste (par exemple de A à D) et de [fusionner](#) ensuite votre liste d'informations mutuelles avec une autre (par exemple de E à H).

Rubriques connexes : [Co-occurents](#), [Mutual Information Settings](#), [Affichage des informations mutuelles](#), [Création d'une liste d'index](#), [Affichage des listes d'index](#), Sommaire de l'aide de WordList.

8.17 Affichage des informations mutuelles

Le sous-programme d'informations mutuelles comporte un certain nombre de colonnes :



N	Word 1	Freq.	Word 2	Freq.	Texts	Gap	Joint	MI	Z	MI3	Log L.	et
1	TED	5	HEATH	5	2	1	5	15.27	140.63	-9.04	0.00	
2	PURSE	6	STRINGS	5	3	1	5	15.01	128.37	-9.04	0.09	
3	SPEAKERS	6	VOUCH	6	6	4	6	15.01	140.63	-8.25	0.00	
4	ROBIN	7	COOK	7	4	1	7	14.79	140.62	-7.59	0.00	
5	BRENT	8	SPARS	5	1	1	5	14.59	111.16	-9.04	0.70	
6	TEMPLE	8	MORRIS	7	2	1	7	14.59	131.53	-7.59	0.07	

2,601 Type-in TED

Mot 1: le mot à gauche, suivi de sa fréquence (sa fréquence dans tout l'index).

Mot 2: le mot à droite, suivi de sa fréquence (sa fréquence dans tout l'index).

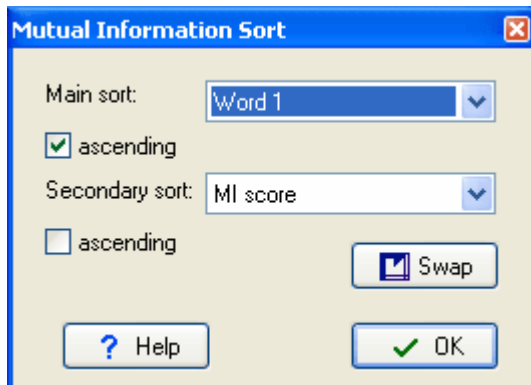
Textes: le nombre de textes dans lequel cette paire de mots a été trouvée (l'index entier en compte 56).

Intervalle: l'intervalle qui sépare généralement le mot 1 du mot 2.

Commun: leur fréquence commune.

À la ligne 2 de cet écran, il est indiqué que PURSE apparaît 6 fois dans l'index, et STRINGS 5 fois. Ils apparaissent 5 fois ensemble. En d'autres termes, dans ce petit corpus, **strings** fait toujours partie du syntagme **purse strings**. L'intervalle est de 1, car **strings** est le mot qui suit **purse**. La paire **purse strings** apparaît dans 3 textes.

Comme toujours, vous pouvez trier les données en cliquant sur les en-têtes. Le tri ci-dessus a été effectué en cliquant d'abord sur "IM", puis sur "Mot 1".



Le tri est double, principal et secondaire, car vous souhaitez parfois voir l'effet qu'a un tri par score d'IM ou Z, ou un autre tri sur l'ensemble de la liste, et parfois vous souhaitez garder les mots triés par ordre alphabétique et simplement les trier par score d'IM ou Z au sein de chaque type de mots. Cliquez sur *Inverser* pour basculer entre tri primaire et secondaire.

Comparez avec l'écran trié par note Z ([Oakes](#) p. 163).

N	Word 1	Freq.	Word 2	Freq.
1	TED	5	HEATH	5
2	SPEAKERS	6	VOUCH	6
3	ROBIN	7	COOK	7
4	BRAINTEASER	14	CROCODILE	14
5	RECORDED	55	TRANSMISSION	55

frequency | alphabetical | statistics | filenames | mutual information | notes

2,601 | Type-in | TED

TED HEATH (un premier ministre anglais des années 1970) est toujours en tête et SPEAKERS ... VOUCH s'affiche toujours, mais d'autres éléments de sont rajoutés.

Voici l'écran trié par score IM3 ([Oakes](#) p. 172) :

The screenshot shows the WordSmith Tools interface with the 'main_index.mut' window. The 'mutual information' tab is selected. The table displays the top 5 most frequent words in a list.

N	Word 1	Freq.	Word 2	Freq.
1	LOOK	330	AT	1,080
2	LET	211	ME	400
3	PER	166	CENT	160
4	HAS	543	BEEN	580
5	TALKING	184	ABOUT	920

At the bottom of the window, the text '2,601 Type-in TED' is visible.

On trouve les éléments bien plus fréquents en tête de liste.

Pour finir, voici le tri par log de vraisemblance ([Dunning, 1993](#)) :

The screenshot shows the WordSmith Tools interface with the 'main_index.mut' window. The 'mutual information' tab is selected. The table displays the top 3 words sorted by log-likelihood.

N	Word 1	Freq.	Word 2	Freq.	Texts	Gap	Joint
1	THERE 'LL	9	BE	1,802	8	1	9
2	RELYING	5	ON	1,621	4	1	5
3	CONCENTRATING	6	ON	1,621	4	1	5

At the bottom of the window, the text '2,601 Type-in THERE'LL' is visible.

Ici, les éléments de la colonne Mot 2 ont une fréquence très élevée et on obtient une colligation (co-occurrence grammaticale).

Rubriques connexes : [Informations mutuelles](#), [Calcul des informations mutuelles](#), [Création d'une liste d'index](#), [Affichage des listes d'index](#), Sommaire de l'aide de WordList.

Référez-vous aux ouvrages de [Oakes](#) pour de plus amples informations sur les informations mutuelles.

8.18 re-trier : listes de cohérence

L'affichage de cohérence par ordre de fréquence peut être re-trié par ordre *alphabétique* (Mot)

total des fréquences (Total, la valeur par défaut)

fréquences dans un fichier donné (les noms de fichiers s'affichent).

Cliquez sur Mot, Total ou nom de fichier.

Le tri peut être croissant ou décroissant, la valeur par défaut étant décroissant.

Rubriques connexes : [Tri des listes de mots](#)

8.19 statistiques

Ils englobent :

- le nombre de fichiers ayant trait à la liste de mots
- la taille des fichiers (en octets, c'est-à-dire en caractères)
- les mots du texte (*tokens*)
- nombre de mots différents (*types*)
- [ratios type/token](#)
- nombre de [phrases](#) dans le texte
- longueur de phrase moyenne (en mots)
- écart moyen de la longueur des phrases (en mots)
- nombre de [paragraphes](#) dans le texte
- longueur de paragraphe moyenne (en mots)
- écart moyen de la longueur des paragraphes (en mots)
- nombre d'[en-têtes](#) dans le texte
- longueur d'en-tête moyenne (en mots)
- nombre de [sections](#) dans le texte
- longueur de section moyenne (en mots)
- écart moyen de la longueur des en-têtes (en mots)
- nombre de mots d'1 lettre

...

nombre de mots à **n** lettres (faites défiler vers le bas dans la liste)

(la longueur de mot maximale est de 14 par [défaut](#). La longueur peut aller jusqu'à 50 lettres, vous pouvez la modifier dans *Paramètres de la liste de mots*, à partir du menu *Paramètres*.) Les mots dépassant cette longueur sont tronqués, et un signe + apparaît à la fin du mot, signalant la troncature.

Le nombre de mots différents (types) est calculé séparément pour chaque texte. Par conséquent, si vous avez créé une liste de mots se basant sur plusieurs textes, vous n'obtiendrez pas le même total en additionnant le nombre de mots dans chaque texte que le nombre de mots dans l'ensemble du corpus.

Rubriques connexes : [Affichage WordList](#) (avec capture d'écran), [Statistiques de résumé](#), [Début et fin des segments de texte](#).

8.20 Importer des mots à partir d'une liste au format texte

le but

Vous serez parfois amené à créer une liste de mots à partir de données obtenues sous forme de liste lorsque vous n'avez pas accès aux textes originaux.

critères

Votre fichier texte peut être dans n'importe quelle [langue](#) (sélectionnez la avant de créer la liste), et peut être en [Unicode](#) ou [ASCII](#).

Le format doit être compatible avec la [liste d'arrêt](#), et chaque mot doit être suivi de <tab> et de la fréquence sous forme de nombre brut (les décimales sont ignorées). N'utilisez pas de virgule pour délimiter les milliers, car ils seraient considérés comme étant des mots. Il n'est pas nécessaire que les mots soient rangés par ordre alphabétique ou par ordre de fréquence.

Exemple

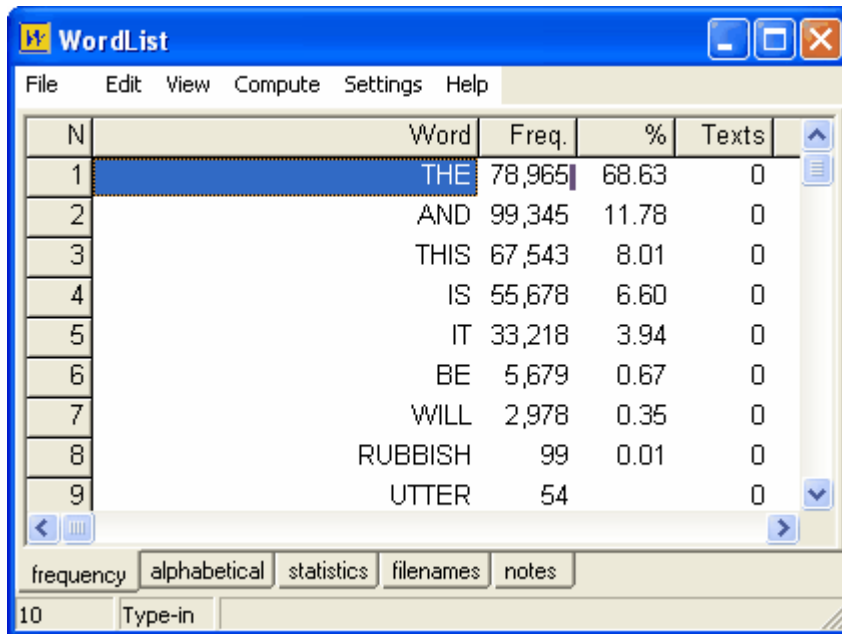
```
; Liste de mots test.  
THIS 67 543
```

```

IT      33 218
WILL   2  978
BE      5  679
COMPLETE 45
AND    99 345
UTTER                    54
RUBBISH 99
THE    578 965
IS     55 678

```

Les résultats se présentent ainsi.



The screenshot shows the WordList application window with a menu bar (File, Edit, View, Compute, Settings, Help) and a table of word statistics. The table has columns for Rank (N), Word, Frequency (Freq.), Percentage (%), and Texts. The word 'THE' is highlighted in the first row.

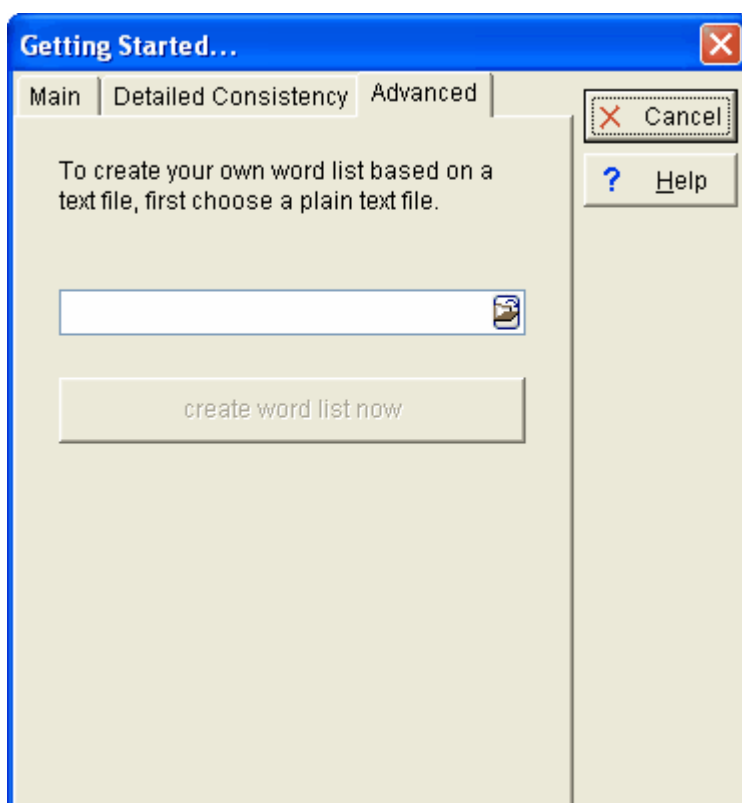
N	Word	Freq.	%	Texts
1	THE	78,965	68.63	0
2	AND	99,345	11.78	0
3	THIS	67,543	8.01	0
4	IS	55,678	6.60	0
5	IT	33,218	3.94	0
6	BE	5,679	0.67	0
7	WILL	2,978	0.35	0
8	RUBBISH	99	0.01	0
9	UTTER	54		0

At the bottom of the window, there are tabs for 'frequency', 'alphabetical', 'statistics', 'filenames', and 'notes'. The 'frequency' tab is selected. Below the tabs, there is a 'Type-in' field with the number '10' next to it.

Les statistiques sont calculés simplement : les longueurs de mots (avec écart moyen et standard) et le nombre de types et de tokens. La plupart des procédures doivent connaître le nombre total de mots (tokens) ainsi que le nombre de types de mots différents pour que vous puissiez utiliser votre liste de mots dans KeyWords ou un autre programme.

procédure

Si vous cliquez sur l'option de menu *Nouveau* (🟢) dans WordList, une fenêtre avec les onglets suivant apparaît : un onglet *Principal* pour les utilisations les plus courantes,



un pour la [cohérence détaillée](#), et un autre (*Avancé*) pour créer des listes de mots en utilisant un fichier de texte brut.

Sélectionnez votre fichier .txt puis cliquez sur *créer la liste de mots*.

8.21 ratios type/token

Si un texte comporte 1 000 mots, on dit qu'il comporte 1 000 "tokens". Cependant, bon nombre de ces mots vont se répéter, et il n'y aura peut-être que 400 mots différents dans le texte. Ainsi, les "types" représentent les différents mots.

Ici, le ratio entre les types et les tokens serait de 40 %.

Mais ce ratio type/token (également appelé TTR) varie énormément en fonction de la longueur du texte (ou du corpus de texte) étudié. Un article de 1 000 mots peut avoir un TTR de 40 %, un texte plus court avoir un ratio de 70 % et un texte de 4 millions de mots, quant à lui, peut avoir un ratio de 2 % et ainsi de suite. Dans la plupart des cas, ces informations n'ont pas vraiment de sens, même si elles sont affichées dans les statistiques de WordList. Le TTR conventionnel donne bien sûr quelques informations si vous traitez un corpus comprenant de nombreux segments de texte de taille égale (par exemple les corpus LOB et Brown). Mais en réalité, et surtout si vos recherches s'orientent sur le texte par opposition à la langue, vous serez amené à traiter des textes de longueurs différentes, et le TTR conventionnel ne vous sera pas d'une grande aide.

Cependant, WordList calcule cela de manière différente. Le ratio type/token standardisé (également appelé STTR) est calculé tous les *n* mots à mesure que WordList parcourt chaque fichier texte. Par [défaut](#), *n* = 1 000. Cela signifie que le ratio est calculé pour les 1 000 premiers mots, puis recalculé pour les 1 000 suivants, et ainsi de suite jusqu'à la fin de votre texte ou corpus. Une moyenne est calculée, ce qui signifie que vous obtenez un ratio type/token basé sur des blocs consécutifs de texte de 1 000 mots. Les textes comprenant moins de 1 000 mots (ou

les textes comprenant un nombre de mots inférieur au nombre n défini) obtiennent un ratio standardisé de 0.

Définition de la limite N

Vous pouvez modifier le nombre n dans la section [Paramètres minimum et maximum](#) et lui affecter une valeur comprise en 100 et 20 000.

Ce que calcule le STTR

Remarque : Le ratio est calculé a) en comptant chaque [forme différente](#) comme un mot (ainsi **manger** et **manges** sont considérés comme deux types différents) b) en utilisant les mots qui ne sont pas dans une [liste d'arrêt](#) c) en prenant en compte ceux dont la longueur est comprise dans les valeurs que vous avez spécifiées, d) en tenant compte de vos préférences en matière de [nombres](#) et [traits d'union](#).

Le nombre affiché est un pourcentage des nouveaux types par tranche de n tokens. Ainsi, vous pouvez comparer les ratios type/token de textes de différentes longueurs. Cette méthode s'oppose à celle de [Tuldava](#) (1995:131-50) qui se base sur la notion de 3 stades d'accumulation. La méthode de WordSmith pour calculer le STTR est de ma propre invention, mais elle est parallèle à l'une des méthodes mise au point par le mathématicien [David Malvern](#) travaillant avec Brian Richards (Université de Reading).


Pour aller plus loin

TTR et STTR sont des mesures relativement grossières même si la plupart des gens supposent qu'elles donnent des informations sur la "densité lexicale". Supposons que vous disposiez d'un texte comprenant 1 000 mots consacrés à une discussion sur **ÉLÉPHANT, LION, TIGRE** etc., puis 1 000 sur **MADONNA, ELVIS** etc., puis 1 000 sur **NUAGE, PLUIE, SOLEIL**. Si vous définissez la limite STTR à 1 000 et que vous obtenez 48 % pour chaque section, les statistiques ne vous diront pas qu'il s'agissait de l'Afrique, de la musique et du temps. Supposons que la limite entre Afrique et Musique se situe au mot 650 au lieu du mot 1 000, il y a fort à parier qu'il n'y aura que très peu de différences dans les statistiques. Mais qu'est-ce qui entraînerait *vraiment* une différence ? Un texte traitant des nuages et écrit par une personne faisant une grande distinction entre les différents types de nuages peut également utiliser **BROUILLARD, BRUME, CUMULUS, CUMULO-NIMBUS**. Le STTR serait plus élevé que pour un texte écrit par un enfant employant principalement le mot **NUAGE** et des adjectifs tels que **HAUT, BAS, ÉPAIS, FONCÉ, FIN, TRÈS FIN** et ferait une utilisation abondante des adjectifs **FONCÉ, MIN** pour décrire les nuages.

Remarque : Shakespeare était connu pour n'employer qu'un vocabulaire relativement limité en termes de mesures telles que celles-ci.

8.22 sensibilité à la casse

On utilise le plus souvent des listes de mots non sensibles à la casse, notamment parce que dans de nombreuses langues, on utilise des majuscules à la fois pour les noms propres et pour marquer les débuts de phrase, d'en-tête etc. Pour obtenir une liste de mots faisant la distinction entre major, Major et MAJOR, activez la sensibilité à la casse (*Modifier les paramètres | WordList | Sensibilité à la casse* dans le [contrôleur](#)).

Il se peut que votre liste de mots sensible à la casse s'affiche entièrement en majuscules. Appuyez sur **Ctrl+L** ou utilisez l'option de menu [Mise en page](#) () pour y remédier.

8.23 paramètres minimum et maximum

Ils englobent :

longueur de mot minimum

Valeur par défaut : 1 lettre. Lorsque vous créez une liste de mots, vous pouvez spécifier une longueur de mot minimum, pour éliminer par exemple tous les mots de moins de 3 lettres.

longueur de mot maximum

Valeur par défaut : 49 lettres. Les mots peuvent avoir une longueur maximale de 50 caractères. Si un mot ne rentre pas dans cette limite et que vous cochez l'option Abréger avec +, WordList place le symbole + à la fin de ce mot pour indiquer qu'il est abrégé. Si Abréger avec + n'est pas activé, les mots longs ne sont pas pris en considération lors de la création de la liste de mots. Si vous indiquez une longueur minimale et maximale de 4 et ne cochez pas l'option Abréger avec +, vous obtenez une liste de mots ne comportant que les mots de 4 lettres.

fréquence minimum

Valeur par défaut : 1. Par défaut, tous les mots sont enregistrés, y compris ceux qui n'apparaissent qu'une fois. Pour ne voir que les mots les plus fréquents, indiquez une valeur comprise entre 0 et 32 000.

fréquence maximale

La valeur maximale par défaut est 2 147 483 647 (2 Go). Il faudrait traiter un nombre considérable de textes pour atteindre cette limite. En indiquant une fréquence maximum de 500 et une fréquence minimum de 50, la liste de mots ne contient que les mots moyennement fréquents.

valeur moyenne type/token (par défaut 1 000)

Permet de calculer aisément le ratio type/token dans une liste de mots. Indiquez un nombre entre 10 et 20 000. Pour une explication détaillée, reportez-vous à la section sur les [informations type/token de WordList](#).

Rubriques connexes : [Caractéristiques de texte](#), [Listes d'arrêt](#), [Définir les paramètres par défaut](#)

8.24 tri

Procédure...

Pour trier, cliquez simplement sur la première ligne d'une liste. Cliquez à nouveau pour basculer entre tri croissant et décroissant.

La fenêtre principale de fréquence n'effectue pas de tri lorsque vous avez une liste de mots à l'écran, mais vous pouvez re-trier dans la fenêtre Alphabétique (voir les onglets au bas de WordList) de différentes façons.

Pour sélectionner un des tris ci-dessous, appuyez sur F6 ou Ctrl+F6 ou Maj+Ctrl+F6.

Autrement, cliquez sur l'option de menu appropriée.

Tri des mots par ordre alphabétique

Avant de trier ou de re-trier, vérifiez que vous avez sélectionné la bonne [langue](#), car de nombreuses langues ont un ordre de tri spécifique. En espagnol, par exemple, l'ordre est le suivant : **A,B,C,CH,D,E,F,G,H,I,J,K,L,LL,M,N,Ñ,O,P,Q,R,S,T,U,V,W,X,Y,Z**.

Inverser le classement des mots

Cela vous permet de classer les mots en fonction du suffixe. L'ordre est déterminé par la fin des mots, et non par le début. Toutes les formes en *-ing* seront ainsi regroupées.

Tri par longueur de mot 

Cela permet de trier des mots en fonction de leur longueur (1 lettre, 2 lettres etc. jusqu'à 50 lettres). Les mots de même longueur sont triés par ordre alphabétique.

Tri de cohérence

Cliquez sur l'en-tête "Textes" pour re-trier les mots en fonction de leur [cohérence](#).

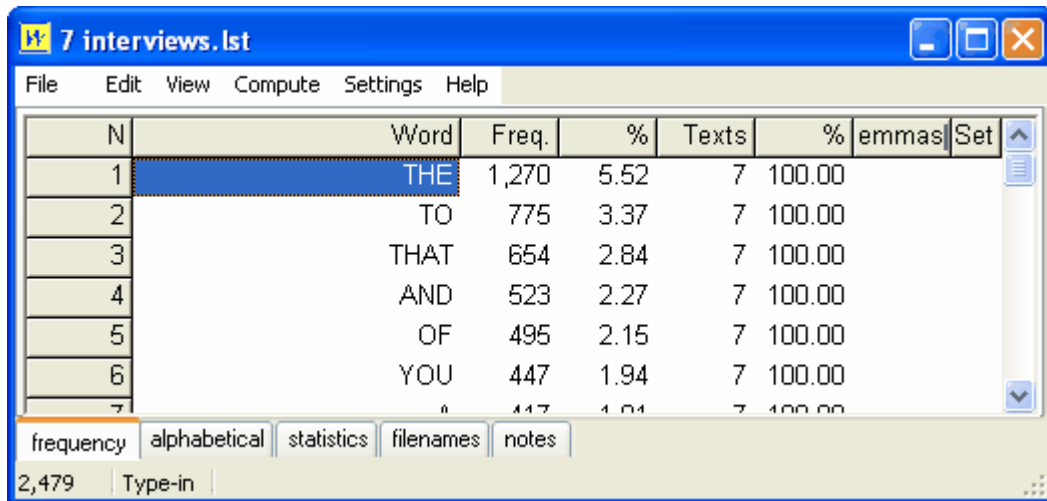
Rubriques connexes : [Tri dans Concord](#), [Tri dans KeyWords](#), [Éditer des entrées](#); [Caractères accentués](#); [Sélectionner une langue](#)

8.25 Affichage WordList

Chaque fenêtre de WordList affiche les éléments suivants :

- le mot
- sa fréquence
- sa fréquence en tant que pourcentage des mots dans les textes à partir desquels la liste a été créée
- le nombre de textes dans lesquels chaque mot apparaît
- ce nombre est un pourcentage du corpus de textes

L'affichage de fréquence ressemble à ceci :



N	Word	Freq.	%	Texts	% emmas	Set
1	THE	1,270	5.52	7	100.00	
2	TO	775	3.37	7	100.00	
3	THAT	654	2.84	7	100.00	
4	AND	523	2.27	7	100.00	
5	OF	495	2.15	7	100.00	
6	YOU	447	1.94	7	100.00	
7	a	417	1.81	7	100.00	

frequency | alphabetical | statistics | filenames | notes

2,479 Type-in

Ici vous pouvez voir les 6 mots principaux d'une liste de mots basée sur 7 interviews. Il y a 2 479 mots au total, mais seuls les premiers sont visibles sur la capture d'écran. La colonne Fréq. affiche le nombre de fois où chaque mot apparaît (**THE** apparaît 1 270 fois dans les 7 textes), et la colonne % nous montre que 1 270 représente 5,52 % des mots présents dans les 7 textes. La colonne Textes montre que le mot **THE** apparaît dans 7 textes, c'est à dire dans 100 % des textes utilisés pour la liste de mots.

N	Word	Freq.	%	Texts	% emmas	Set
4	ABERDEEN	1	0.00	1	14.29	
5	ABILITY	1	0.00	1	14.29	
6	ABLE	18	0.08	5	71.43	
7	ABOLISHING	1	0.00	1	14.29	
8	ABOUT	70	0.30	7	100.00	
9	ABOVE	1	0.00	1	14.29	
10	ABOQUITE	1	0.00	1	14.29	

frequency alphabetical statistics filenames notes
2,479 Type-in

Le tri alphabétique affiche quelques mots mais cette fois, ils sont triés par ordre alphabétique. **ABLE** apparaît 18 fois au total, dans 5 des 7 textes. **ABOUT**, en revanche, apparaît dans les 7 textes.

Étudions à présent les statistiques.

N	0	1	2	3	
text file	overall	31.5.98.txt	alex salmond	15.3.98.txt	4.6.98.txt
file size	151,474	23,489	20,216	37,530	
tokens (running words) in text	22,992	3,485	3,089	5,736	
tokens used for word list	22,992	3,485	3,089	5,736	
types (distinct words)	2,479	824	674	924	
type/token ratio (TTR)	10.78	23.64	21.82	16.10	
standardised TTR	37.15	36.37	37.37	35.06	
standardised TTR std.dev.	58.95	49.00	46.95	53.56	
standardised TTR basis	1,000	1,000	1,000	1,000	
mean word length (in characters)	4.49	4.52	4.55	4.44	
word length std.dev.	2.42	2.48	2.58	2.33	
sentences	928	147	132	233	
mean (in words)	24.78	23.71	23.40	24.21	
std dev	19.73	20.38	15.28	18.79	

frequency alphabetical statistics filenames notes
2,479 Type-in

Comme mentionné ci-dessus, il y a 2 479 types de mots dans les 7 textes. Le nombre total de

mots s'élève à 22 992. Chaque mot comprend en moyenne 4,49 caractères. Il y a 928 phrases au total, chacune comportant 24,78 mots en moyenne. Dans le texte de l'interview d'Alex Salmond, il n'y a que 674 types de mots différents, et cette interview comporte un peu plus de 3 000 mots. Ceci est expliqué plus en détail à la page [Statistiques](#).

Enfin, voici une capture d'écran représentant la même liste de mots triée par ordre alphabétique décroissant. Dans la partie que l'on peut voir, on constate que tous les mots se finissent par -IC

N	Word	Freq.	%	Texts
22	TRAFFIC	3	0.01	2
23	SPECIFIC	4	0.02	4
24	TERRIFIC	1	0.00	1
25	MAGIC	1	0.00	1
26	STRATEGIC	2	0.01	2
27	PUBLIC	31	0.13	4
28	DYNAMIC	1	0.00	1

Pour trier une liste par ordre alphabétique décroissant, rendez la fenêtre Alphabétique visible, puis cliquez sur *Edition | Tri décroissant* dans le menu. Pour revenir à un tri alphabétique ordinaire, appuyez sur F6.

Rubriques connexes : [Cohérence](#), [Lemmatisation](#)

8.26 contrôleur WordSmith : Paramètres de WordList

Ces derniers sont accessibles via le [contrôleur](#) dans *Modifier les paramètres | WordList*. Certains des choix effectués, par exemple les [paramètres maximum et minimum](#), peuvent affecter d'autres outils.

Il y a deux ensembles : Ce que vous obtenez et ce qui s'affiche

CE QUE VOUS OBTENEZ

Longueur de mot et fréquences

Pour de plus amples informations, reportez-vous à la section [Paramètres minimum et maximum](#)

Nombre type/token standardisé

Reportez-vous à la section [Type WordList/Informations sur les tokens](#).

Balises

Si vous souhaitez inclure des balises dans une liste de mots, vous devez d'abord créer un [fichier de balises](#). Assurez-vous ensuite que l'option "Inclure les balises" est cochée pour que les balises soient visibles dans la liste de mots. Les statistiques basées sur le décompte des mots

ou le nombre de types ou de tokens ne comprennent pas les balises. Pour chaque balise s'affichent sa fréquence, cette même fréquence en terme de pourcentage des mots à l'exception des balises, et le nombre de textes présents.

CE QUI S'AFFICHE

Sensibilité à la casse

En règle générale, vous créez des listes de mots qui ne sont pas sensibles à la casse. Si vous souhaitez qu'une liste le soit, cochez [sensibilité à la casse](#).

Rubriques connexes : [Utilisation des listes d'index](#), [Affichage des listes d'index](#), Sommaire de l'aide de WordList, [Calcul des clusters de listes de mots](#).

WordSmith Tools

Viewer et Aligner

Section

IX

9 Viewer et Aligner

9.1 but

C'est un programme permettant d'afficher votre texte ou d'autres fichiers en mettant en surbrillance les mots intéressants. Par défaut, ils s'affichent au format [texte brut](#), mais vous pouvez afficher du texte au format [SGML ou HTML](#), ou afficher les caractères accentués sous forme de codes numériques, par exemple pour visualiser les codes remplaçant les [caractères accentués](#). Il y a un grand nombre de [paramètres](#) et d'[options](#) que vous pouvez modifier.

Viewer permet de parcourir vos textes à une vitesse que vous pouvez définir.

Il vous permet également d'[éditer](#) votre fichier texte avec des symboles tels que € convertis dans le jeu de caractères Windows, de numéroter les phrases ou les paragraphes, puis enfin d'[enregistrer comme fichier texte](#).

Vous pouvez également l'utiliser pour produire une version [alignée](#) de deux textes, avec tour à tour des phrases ou paragraphes de l'un ou de l'autre.

Viewer est un utilitaire indépendant, mais qui reste lié aux autres outils : il est lancé à chaque fois que vous avez besoin de revenir au texte source, par exemple pour afficher vos [mots-clés](#) en contexte.

Rubriques connexes : [Paramètres de Viewer](#), [Options de Viewer](#)

9.2 alignement avec Viewer

Cette fonctionnalité permet d'aligner les phrases issues de deux fichiers. Les traducteurs ont besoin d'étudier les différences entre l'original et la traduction. D'autres linguistes voudront étudier les différences entre deux versions d'un texte dans la même langue. Des étudiants de [langues différentes](#) peuvent l'utiliser comme des lectures bilingues, pour étudier de près les différences dans l'ordre des mots.

Cette fonction vous aide à créer un nouveau texte composé des deux fichiers, dans lequel les phrases sont alternées. Vous pouvez ainsi comparer la traduction et le texte original.

Exemple

Original : *Der Knabe sagte diesen Gedanken dem Schwesterchen, und diese folgte. Allein auch der Weg auf den Hals hinab war nicht zu finden. So klar die Sonne schien, ...* (extrait de *Bergkristall*, de Adalbert Stifter)

Traduction : *Le garçon fit part de cette pensée à sa sœur et elle le suivit. Mais la route était introuvable. Bien que le soleil brillait, ...*

Texte aligné :

<G1> Der Knabe sagte diesen Gedanken dem Schwesterchen, und diese folgte.

<E1> Le garçon fit part de cette pensée à sa sœur et elle le suivit.

<G2> Allein auch der Weg auf den Hals hinab war nicht zu finden.

<E2> Mais la route était introuvable.

<G3> So klar die Sonne schien, ...

<E3> Bien que le soleil brillait, ...

Deux textes alignés vous permettent de repérer les ajouts, oublis, normalisations, changements de style et préférences au niveau de l'ordre des mots. Dans le cas présent, le traducteur a évité les calques.

Rubriques connexes : [Les boutons](#)

9.3 alignement et déplacement

Le premier texte s'affichera dans une couleur et le deuxième dans une autre couleur. Les couleurs peuvent être modifiées dans le menu *Paramètres*.

Vous pouvez être amené à modifier l'ordre des phrases. Il se peut que le traducteur ait utilisé

trois phrases alors qu'il n'y en avait qu'une dans le texte d'origine.

ajustement avec la souris

Pour changer l'ordre des phrases, il suffit de faire glisser une phrase et de la placer au-dessus de celle qui la précède et qui est de même couleur.

Pour finir, il est recommandé d'[enregistrer les résultats \(F2\)](#).

Rubriques connexes : Sommaire de Viewer

9.4 édition

Bien que **Viewer** ne soit pas un logiciel de traitement de texte complet, il possède quelques fonctionnalités d'édition qui ont été intégrées afin de palier aux problèmes de format les plus courants.

Édition ([Ed](#)) : ouvre une fenêtre vous permettant d'éditer le contenu de la phrase ou du paragraphe.

Supprimer les espaces superflus : cette fonction permet de parcourir chaque phrase du texte et de supprimer tout espace superflu. Plusieurs espaces consécutifs sont réduits à un seul.

Rechercher des lignes en minuscules : cela permet d'identifier une phrase ou un paragraphe ne commençant pas par une majuscule ou un nombre. Vous pouvez [regrouper](#) l'élément avec celui le précédant. Ce problème est assez courant lorsque le texte a été enregistré en tant que "texte brut avec sauts de ligne" (où la balise <Entrée> se situe à la fin de chaque ligne, que ce soit une fin de paragraphe ou non).

[Rechercher des lignes courtes](#)

[Insérer des balises](#)

Il est ensuite recommandé d'enregistrer votre texte (F2)

Vous pouvez également :

ouvrir un nouveau fichier pour l'afficher (dans **Viewer** vous pouvez ouvrir jusqu'à 10 fichiers)

copier un fichier texte dans le [presse-papiers](#) (sélectionnez, puis appuyez sur Ctrl+Inser)

imprimer tout le fichier texte actif ou seulement une partie

modifier le [jeu de caractères](#) actuel

rechercher des mots ou des syntagmes (appuyez sur F12)

parcourir vos textes (un écran à la fois). Le curseur change de forme pour montrer que vous êtes en mode "parcourir". Vous pouvez modifier la vitesse grâce à la petite barre de défilement horizontale et arrêter de parcourir les textes à n'importe quel moment en cliquant dans le texte ou en appuyant sur une touche du clavier.

9.5 numérotation des phrases et des paragraphes

Viewer permet de faire une copie de votre texte dont toutes les phrases ou tous les paragraphes sont balisés avec <S> et <P>.

Pour ce faire, ouvrez le fichier texte, sélectionnez *Édition | Insérer des balises*, puis [enregistrez-le au format texte brut](#).

Pour sélectionner le mode phrase ou le mode paragraphe, reportez-vous à la rubrique

[Paramètres](#).

Rubriques connexes : Sommaire de Viewer

9.6 options

Mode : phrase/paragraphe

Ceci permet de basculer entre le mode Phrase (par défaut) et le mode paragraphe. Cela signifie que vous pouvez afficher vos fichiers texte de sorte à avoir une phrase ou un paragraphe par ligne. De la même manière, vous pouvez créer un texte aligné en alternant phrases et paragraphes provenant de chacun des textes source. Les autres fonctions ([regrouper](#), [dissocier](#))

fonctionnent de la même façon, quel que soit le mode.

Si vous ré-enregistrez le texte, le mode paragraphe s'avère peut-être le plus approprié. Cela permet d'insérer une marque <Entrée> à la fin de chaque paragraphe. En mode phrase, <Entrée> est présent à la fin de chaque phrase, présentant ainsi le texte comme une longue liste de phrases.

Affichage : Normal/Accents et leurs codes

Le texte peut s'afficher en mode normal ou avec les codes remplaçant les caractères accentués correspondants. Par exemple, **caf<233> au lait** s'affiche au lieu de **café au lait**.

Balises ignorées : masqué/visible

Vos [balises à ignorer](#) peuvent être rendues visibles ou invisibles.

Format : Texte brut/HTML/SGML



Le paramètre par défaut est le [texte brut](#). Cependant, si votre texte est au format [HTML](#), [XML](#) ou [SGML](#), vous pouvez modifier ce paramètre, et les [balises](#) s'afficheront séparément dans la colonne de gauche.

En-tête : masqué/visible


Si votre texte est [balisé](#) et qu'il comporte un en-tête, il est possible d'afficher ou de supprimer ce dernier.

9.7 regroupement et dissociation des phrases

Regroupement

La manière la plus simple de regrouper deux phrases est de faire glisser la phrase que vous souhaitez déplacer et de la placer sur celle située au-dessus. Vous pouvez également les marquer avec , puis cliquer sur . Un avertissement s'affichera si vous risquez de désorganiser votre texte.

Dissociation en deux

Pour scinder une phrase, cliquez sur . Une liste des mots s'affichera. Cliquez sur le mot **terminant** la phrase, puis cliquez sur OK.

exemple

"It
is
good!"
Mary
wanted
more.

Cela permet d'insérer les mots *Mary wanted more.* dans une nouvelle ligne dessous.

Rubriques connexes : [Sommaire de Viewer](#)

9.8 paramètres

Les paramètres sont les paramètres standard de la plupart des outils :

[Couleurs](#)

[Police](#)

[Impression](#)

[Caractéristiques de texte](#)

[Vérifier tous les paramètres](#)

9.9 aspects techniques

Quand est-ce qu'une phrase n'est pas une phrase ?

Il n'existe aucun moyen mécanique parfait permettant de délimiter les fins de phrases. Un en-tête peut par exemple ne comporter aucun point final, mais il n'est pas considéré comme partie intégrante de la phrase qui lui fait suite. Il est fréquent qu'une phrase ne comporte pas de point final, dans le cas d'une liste d'éléments par exemple.

L'algorithme utilisé par **Viewer** est le suivant : une phrase se termine si un point final, un point d'interrogation ou d'exclamation est immédiatement suivi d'un ou plusieurs [séparateurs de mots](#) et si le symbole suivant (à l'exception des signes de ponctuation) est une lettre majuscule, une majuscule accentuée, un nombre ou un symbole monétaire. **WordList**, utilise la même routine, à la différence qu'il essaye de faire la distinction entre les phrases et les en-têtes. Ainsi le nombre de phrases dans les deux outils a peu de chance d'être identique.

Prenons l'exemple de cet extrait de *A Tale of Two Cities* :

"*Wo-ho!*" said the coachman. "So, then! One more pull and you're at the top and be damned to you, for I have had trouble enough to get you to it! - Joe!"

Viewer considère (à tort) que - Joe! est une phrase indépendante, mais traite "*Wo-ho!*" said the coachman. comme une seule phrase. Cependant, le programme séparerait la phrase en deux si le mot qui suit *ho!* comportait une lettre majuscule (par exemple *Wild Bill, the coachman, said.*).

Ainsi, vous constatez que **Viewer** ne peut pas gérer les phrases comme vous le feriez. (La phrase *I saw Mr. Smith.* serait considérée comme deux phrases séparées. En revanche, plusieurs en-têtes peuvent être groupés et considérés comme une seule phrase. C'est pourquoi vous pouvez choisir de *rechercher des phrases courtes* afin de [repérer](#) les phrases ne comportant qu'un seul mot.

Quelle est la longueur d'une phrase ?

L'espace de stockage pour chaque phrase ou paragraphe est de 10 000 caractères. Viewer peut afficher jusqu'à 16 368 phrases ou paragraphes. Si on estime que chaque phrase comporte en moyenne 10 mots, cela représente un total de 160 000 mots.

Gestion du disque, accents, etc.

Pour obtenir de meilleurs résultats, il est préférable d'utiliser des versions [ascii](#) ou [ansi](#) de vos deux textes.

Le texte joint est enregistré en utilisant un [jeu de caractères](#) Windows.

Rubriques connexes : Sommaire de Viewer

9.10 non correspondances de traduction

Viewer permet de rechercher les endroits où l'alignement ne correspond pas (par exemple lorsqu'une phrase source a été traduite par plusieurs phrases cibles). Il suffit pour cela d'aller dans le menu *Correspondance par lettres majuscules*. Cela permet de rechercher les noms propres dans les deux versions : si par exemple Paris est mentionné dans la phrase 25 du texte source et non dans la phrase 25 de la traduction, mais dans la phrase 27, c'est qu'un décalage s'est produit.


Viewer effectue une recherche descendante à partir de la phrase active du texte et pointe vers l'endroit où se trouve le décalage. À vous ensuite de remonter dans le texte pour trouver le point de divergence des phrases. Il peut être utile de grouper par 10 ou par 20 pour accélérer la recherche de décalage.

Une fois le problème isolé, [dissociez](#), [regroupez](#) ou éditez les phrases comme il convient, puis enregistrez le texte.


Rubriques connexes : [Le côté technique](#), [Recherche de phrases inhabituelles](#), Sommaire de Viewer

9.11 résolution des problèmes

Impossible de voir la phrase ou le paragraphe en entier

Cliquez sur  pour redimensionner automatiquement les lignes de l'affichage. Ceci permet d'ajuster la hauteur de ligne en fonction de la colonne de données actuellement en surbrillance.

Impossible d'afficher le fichier texte en entier

Cliquez sur  pour actualiser l'affichage. Viewer devra relire le fichier texte, dans une limite de 16 368 phrases.

Les couleurs ne conviennent pas

Modifiez les couleurs dans le menu *Paramètres*. Les couleurs utilisées pour chaque langue dans la fenêtre double sont les mêmes que les couleurs utilisées pour le tri primaire et le tri secondaire dans **Concord**.

Rubriques connexes : Sommaire de Viewer

9.12 phrases inhabituelles

Comme **Viewer** utilise les points finaux, points d'exclamation et d'interrogation comme indicateurs de fin de phrase, vous constaterez qu'une chaîne telle que "Hello! Paul! Come here!" est considérée comme trois courtes phrases. En fonction de ce que vous cherchez à faire, vous pouvez considérer cette chaîne comme une seule phrase, comme dans le cas d'une traduction (par exemple "Salut Paul, viens ici!").

Il peut s'avérer utile de rechercher des phrases très courtes pour voir si vos phrases originales ont été traitées comme vous le souhaitez.

Cette fonction permet aussi de trouver des lignes en minuscules : lorsqu'une phrase ou un paragraphe ne commence pas par une lettre majuscule ou un nombre, vous pouvez regrouper l'élément avec celui le précédant. Ce problème est assez courant lorsque le texte a été enregistré en tant que "texte brut avec sauts de ligne" (où la balise <Entrée> se situe à la fin de chaque ligne, que ce soit une fin de paragraphe ou non).

recherche

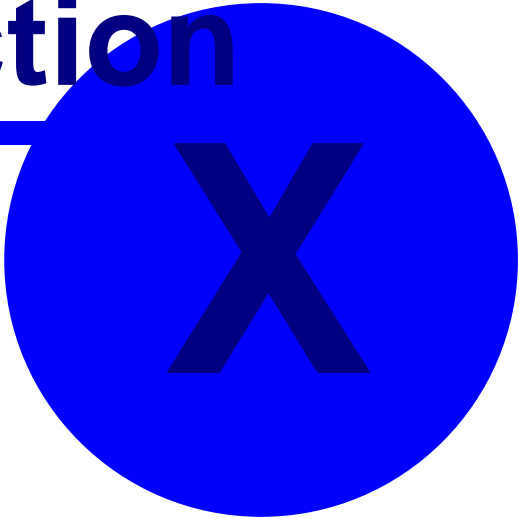
F8 ou *Recherche de lignes courtes*. **Viewer** parcourt les phrases ou paragraphes suivants, potentiellement problématiques, et vous pourrez [regrouper](#) chaque élément avec celui le précédent.

Rubriques connexes : [Le côté technique](#), [Recherche de non correspondance de traduction](#), Sommaire de Viewer

WordSmith Tools

Utilitaires

Section



X

10 Utilitaires

10.1 Convertir les données provenant de versions antérieures

10.1.1 Convertir les données provenant de versions antérieures

Au fur et à mesure que Oxford WordSmith Tools s'est développé, il est devenu nécessaire de stocker des quantités plus importantes de données associées à une liste de mots donnée ou une concordance donnée. Par exemple des données à propos desquelles une ou plusieurs [which langues](#) ont été sélectionnées pour une concordance, des [notes](#) stockées avec chaque type de fichier de résultats. Il est donc devenu nécessaire de fournir un outil permettant de convertir des données à partir des formats utilisés dans les versions 1.0 à 3.0 vers le nouveau format utilisé dans la version actuelle.

16

32

C'est l'outil de Conversion de données

Si vous tentez d'ouvrir un fichier créé avec un version précédente, une fenêtre s'affichera, vous permettant de procéder à la conversion de ce fichier.

10.2 WebGetter

10.2.1 présentation générale

Le but

Télécharger des pages web à l'aide d'un moteur de recherche vous permet de construire votre propre corpus de textes.

Que faire ?

Saisissez simplement un mot ou un syntagme puis cliquez sur OK ou appuyez sur <Entrée>.

Procédure

WebGetter parcourt le moteur de recherche spécifié dans la deuxième boîte et télécharge les 100 premières sources. Le programme utilise le moteur de recherche comme vous le feriez vous-même, récupérant une liste de références utiles. Il lance ensuite un robot qui visite chaque adresse web et télécharge la page Web (à partir du site Internet original et non pas à partir du cache du moteur de recherche). Il est possible de lancer plusieurs robots en même temps. L'avantage est qu'un téléchargement lent ne bloque pas tous les autres.

Après avoir téléchargé une page web, le robot WebGetter vérifie si elle correspond à vos critères (définis dans [Paramètres](#)). Si la page est d'une taille suffisante, un fichier portant un nom très similaire à l'adresse web est enregistré sur votre disque dur.

Lorsqu'il n'y a plus de références, WebGetter visite à nouveau le moteur de recherche pour en obtenir plus.

Rubriques connexes : [Paramètres](#), [Affichage](#), [Limites](#)

10.2.2 paramètres

Les différents paramètres sont les suivants :

- emplacement où les textes doivent être stockés. Le dossier que vous spécifiez sera la *racine*. C'est-à-dire que si vous spécifiez `c:\temp` and et que vous recherchez "besteirol", les résultats seront stockés dans le dossier `c:\temp\besteirol`. Si vous effectuez une autre recherche, "Oxford WordSmith Tools" par exemple, les résultats seront stockés dans `c:\temp\WordSmithTools`.
- délai d'expiration : nombre de secondes après lequel le robot WebGetter arrête d'essayer d'accéder à une page Web s'il n'y a aucune réponse. Valeur suggérée : 20 secondes.
- nombre max. d'éléments simultanés : Le principe de fonctionnement de WebGetter est d'activer des robots simultanément, chacun d'entre eux effectuant une requête à une page Web différente. Valeur suggérée : 20. C'est à dire que 20 pages peuvent être téléchargées à la fois.
- langue : vous spécifiez la langue que vous souhaitez.
- longueur de fichier minimum (20 Ko recommandés) : c'est la taille minimum pour chaque fichier texte téléchargé à partir d'Internet. Les fichiers les plus petits peuvent ne contenir que quelques images et rien de plus.
- nombre de mots minimum (300 recommandés) : après chaque téléchargement, WebGetter parcourt le fichier texte téléchargé en comptant le nombre de mots et n'enregistre le fichier que s'il y a suffisamment de mots.
- mots requis : vous pouvez saisir quelques mots dont vous souhaitez la présence dans chaque téléchargement. Vous pouvez spécifier que tous les mots doivent être présents ou n'importe lequel d'entre eux.

Moteurs de recherche

Téléchargez une sélection de moteurs de recherche en cliquant sur Moteurs. Ceci vous permet de récupérer les dernières informations à propos de chaque moteur de recherche à partir de www.lexically.net/downloads/searchengines.htm.

Options avancées

Si vous travaillez dans un environnement avec un "serveur proxy", WebGetter le reconnaîtra automatiquement et utilisera le proxy sauf si vous décochez la case concernée. Si vous avez le moindre doute, consultez votre administrateur réseau.

La grille de paramètres

Elle comporte les éléments suivants :

nom	C'est le nom qui apparaît dans la partie supérieure, dans la liste de moteurs de recherche
ignorer	Sites Web à ne pas visiter lors du téléchargement (par opposition à une requête de liste). Lorsque WebGetter récupère une page à partir de Google, il ne demande que la liste de Google, et non pas d'autres pages Web Google.
URL	URL où le moteur de recherche est trouvé.
Chaîne de recherche	Syntaxe du terme de recherche
Max.	Nombre d'essais pour chaque contact
Suivant	
Langue	Langue requise :
Autre	

Le terme de recherche est spécifié tout comme lorsque vous utilisez le même moteur de

recherche vous-même. Peu de paramètres avancés sont utilisés pour chaque moteur de recherche. Vous pouvez essayer de les saisir dans la grille, dans la colonne Chaîne de recherche. Vous pouvez mesurer les paramètres de chaque moteur de recherche simplement en l'essayant puis en adaptant la chaîne de recherche comme il convient. Certains moteurs de recherche tentent de placer des cookies sur votre PC, et ceci peut faire échouer le téléchargement.

Vous pouvez afficher la ligne d'adresse dans l'onglet Avancé. WebGetter tente d'indiquer le terme de recherche au moteur de recherche, le nombre maximum de résultats à afficher par requête, la langue à utiliser, et comment en obtenir plus.

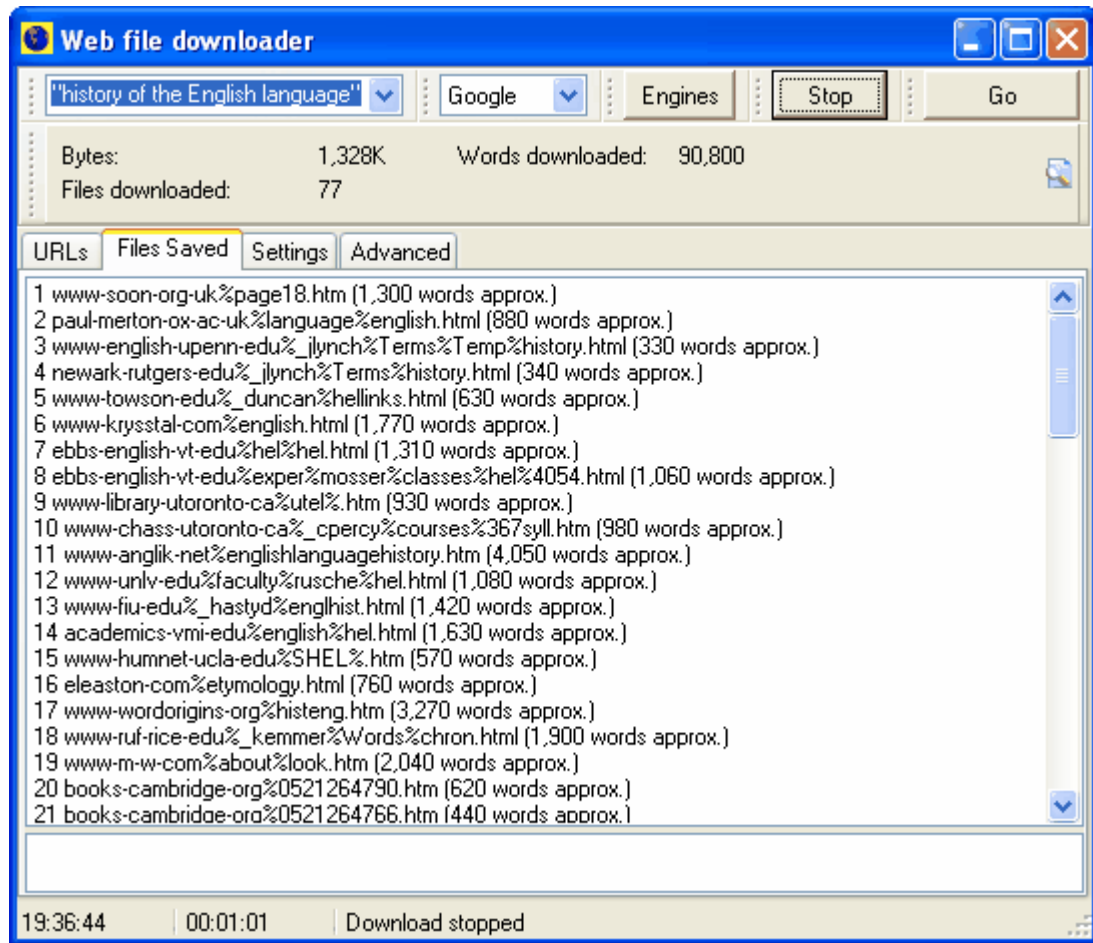
Rubriques connexes : [Affichage](#), [Limites](#)

10.2.3 affichage

Lorsque Webgetter fonctionne, les URL visités s'affichent. S'ils sont en grisés, c'est qu'ils sont trop petits pour être utilisés ou qu'ils n'ont pas encore été contactés. S'ils sont en bleu foncé, c'est qu'ils ont été enregistrés sur le disque. Dans la partie supérieure s'affiche le nombre d'octets traités, et chaque fois qu'un fichier correspondant à vos critères est enregistré, vous verrez le nombre de fichiers s'afficher et le nombre de mots augmenter. En bas s'affichent l'heure et le temps écoulé.

Un onglet donne accès à une liste des fichiers téléchargés avec succès.

Voici une liste partielle de ce que j'ai obtenu avec une connexion large bande en une minute et une seconde, avec le terme de recherche "history of the English language" (NdT : Histoire de la langue anglaise) (avec des guillemets).



Comme vous pouvez le voir, près d'1,3 Mo de pages Web ont été examinées, et 90 000 mots (1,1 Mo) ont été enregistrés, avec les paramètres par défaut (chacune des pages doit faire au moins 10 Ko et comporter 300 mots). J'ai eu quelques temps morts, car vraisemblablement, 20 secondes n'est pas un délai suffisamment long pour certains sites Internet ou serveurs qui sont lents.

Rubriques connexes : [Paramètres](#), [Limites](#)

10.2.4 limites

Tout dépend du moteur de recherche ainsi que des termes de recherche utilisés. Internet est un gigantesque panneau d'affichage. La plupart des éléments présents sont des publicités, des prix de catalogue, etc. Les termes de recherche sont collectés par les moteurs de recherche en analysant les termes insérés par l'auteur de la page web. Il n'y a aucune garantie que les pages web correspondent exactement au terme que vous avez spécifié. Ceci dit, il y a toujours un lien.

Utilisez l'option [Paramètres](#) pour filtrer ce que vous avez téléchargé.

Rubriques connexes : [Affichage](#)

10.3 Sélecteur de langue

10.3.1 Présentation générale



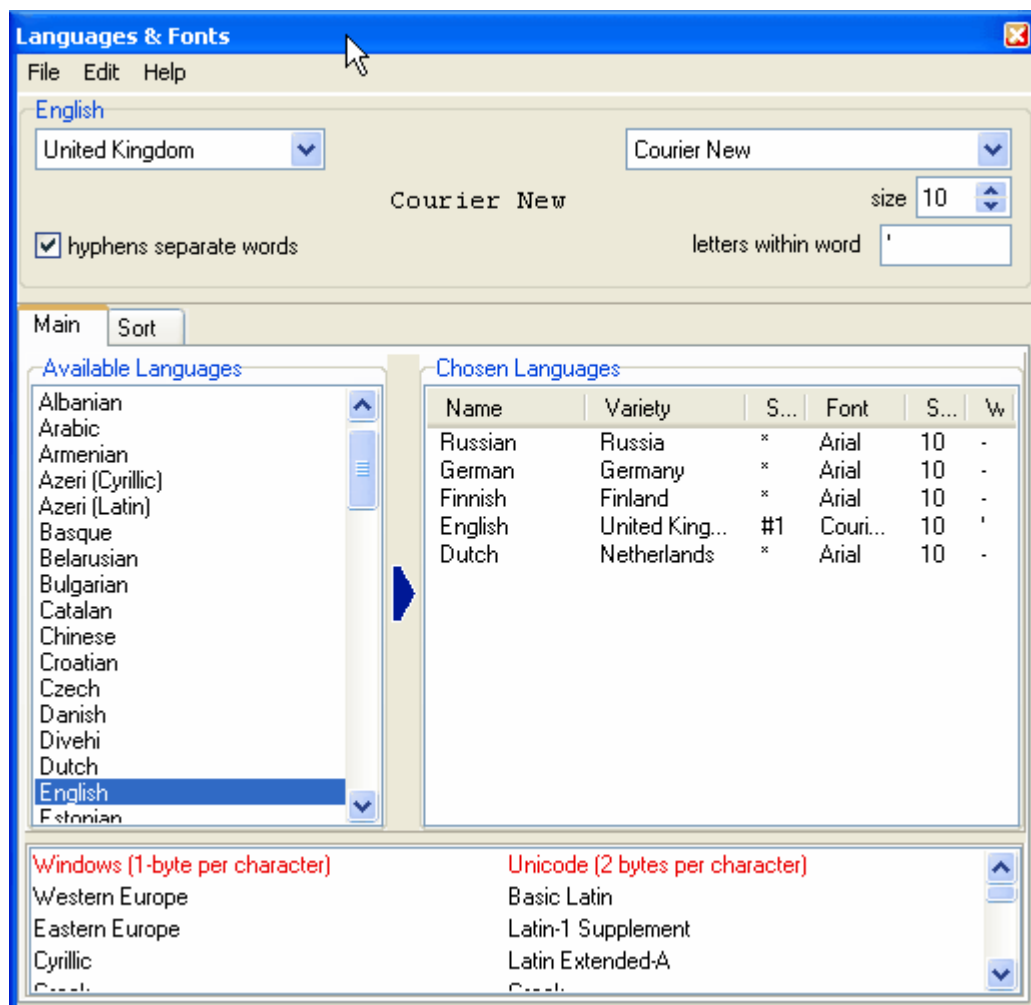
C'est un outil permettant de sélectionner les langues à utiliser.

En règle générale, cette sélection n'est effectuée qu'une fois, lors de votre première utilisation de WordSmith Tools.

Comment y accéder

Le sélecteur de langue est accessible depuis le menu principal du contrôleur WordSmith : *Paramètres | Modifier les paramètres | Texte et Langues | Autres Langues*.

Ce qui s'affiche ressemble à ceci :



5 langues ont déjà été sélectionnées.

Dans la partie inférieure s'affichent les éléments pris en charge par la police actuelle, en termes de texte ANSI Windows ou Unicode. La police Courier New de l'ordinateur utilisée pour cette tâche peut prendre en charge des caractères dans Windows pour l'Europe Occidentale et

Orientale, le cyrillique, etc., ainsi que plusieurs couches du standard Unicode.

Rubriques connexes : [Langue](#), [Police](#), [Tri](#), [Autres langues](#), [Enregistrement de vos préférences](#)

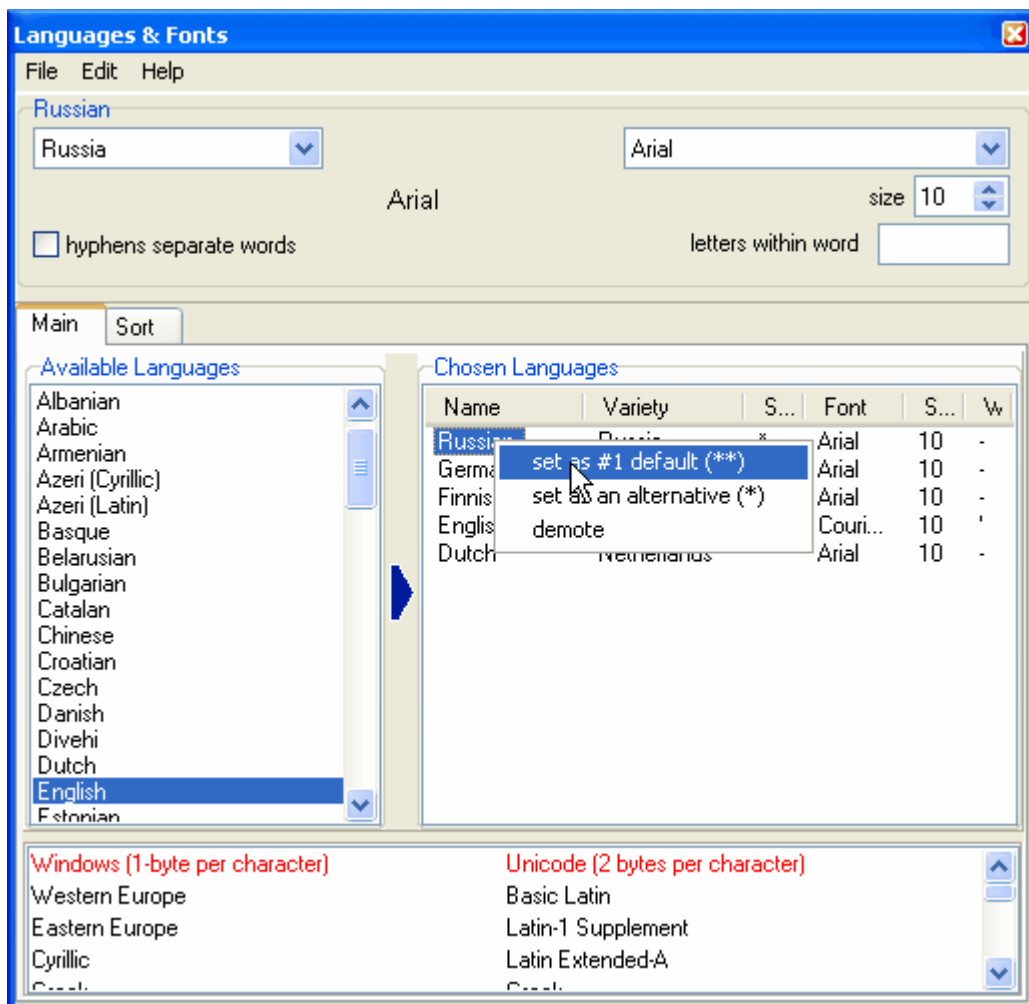
10.3.2 Langue

Comment y accéder

Le sélecteur de langue est accessible depuis le menu principal du contrôleur WordSmith :
Paramètres | Modifier les paramètres | Texte et Langues | Autres Langues.

Fonctions

La liste de langues sur la gauche montre toutes celles prises en charge par votre ordinateur. Si certaines d'entre elles sont grisées, c'est qu'elles n'ont pas été installées dans Windows, bien qu'elles soient prises en charge par votre version de Windows. (Pour installer davantage de prise en charge multilingue, vous aurez besoin du CD-ROM Windows original ou pouvez trouver de l'aide sur Internet.)



Sur la droite sont affichées les langues actuellement utilisées par WordSmith. La langue par défaut doit être marquée par #1 et celles que vous êtes susceptible d'utiliser par un astérisque : *. Pour chaque langue choisie, vous pouvez spécifier tous les symboles qui peuvent être

compris dans un mot, par exemple l'apostrophe en anglais : il est plus logique de considérer "don't" comme étant un seul mot plutôt que "don" et "t" séparément. Vous pouvez aussi spécifier si le trait d'union est un séparateur ou non, par exemple si "abat-jour" doit être considéré comme deux mots ou comme un seul.

Cliquez avec le bouton droit de la souris pour changer le statut de la langue que vous avez choisie. Cet utilisateur va définir le russe comme étant la langue par défaut #1. Pour supprimer une langue, cliquez avec le bouton droit de la souris et sélectionnez "effacer". Pour ajouter une langue, faites-la glisser de la fenêtre de gauche vers celle de droite, sélectionnez ensuite le pays et la police pour cette langue.

Chaque fois que vous changez la langue, la liste des [polices](#) disponibles change, ainsi que l'apparence des [mots triés](#). La fenêtre en bas montre quels sont les caractères pris en charge par ces langues, en Unicode ou en jeu à octet unique.

Certaines langues ne marquent pas les [séparateurs de mots](#).

Rubriques connexes : [Autres langues](#), [Enregistrement de vos préférences](#)

10.3.3 Police

La fenêtre Police affiche les polices disponibles pour chaque langue en fonction de celles qui sont installées. Il vous faudra une police capable d'afficher les caractères dont vous aurez besoin : on trouve de nombreuses polices spécifiques sur Internet. Les polices [Unicode](#) affichent un grand nombre de caractères différents, mais pour ce faire, votre texte doit être enregistré au format Unicode. La liste des caractères disponibles varie selon la police.

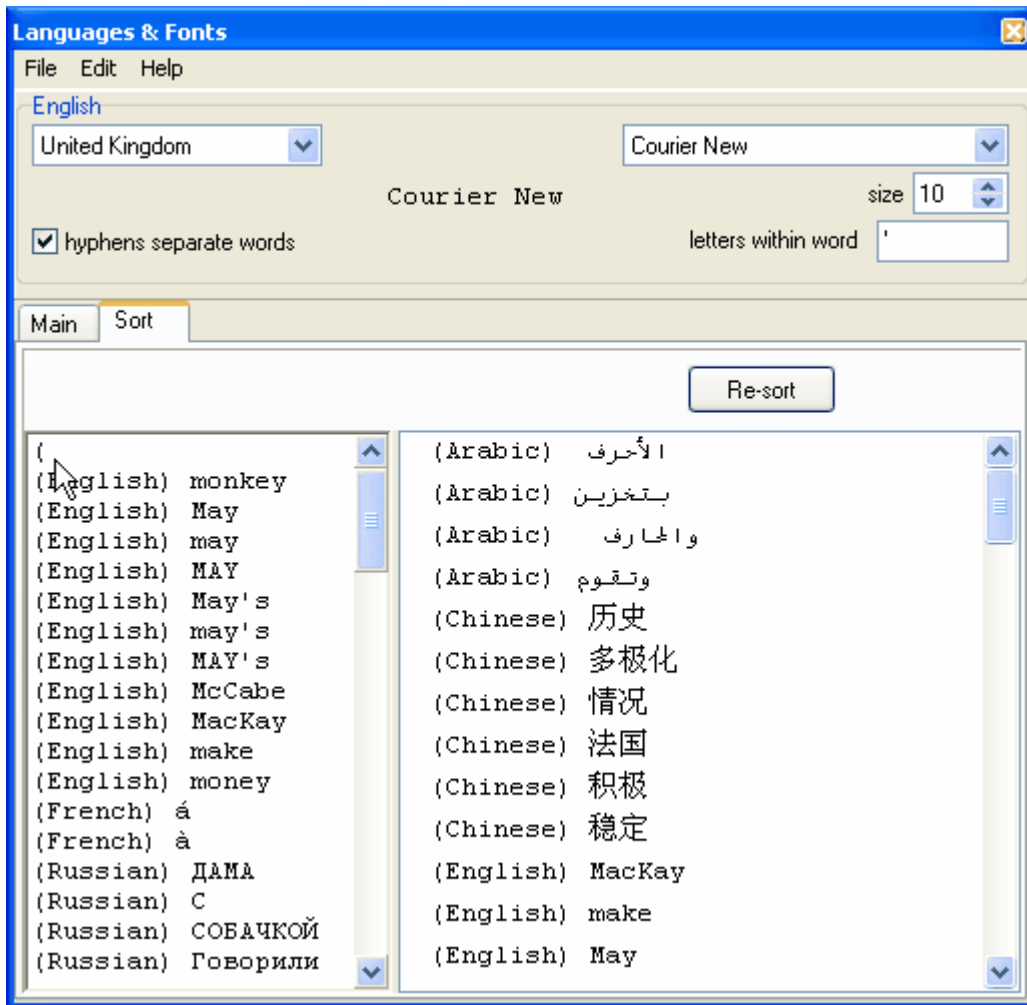
Pour plus d'informations sur Unicode, cliquez [ici](#).

Rubriques connexes : [Langue](#), [Tri](#), [Autres langues](#), [Enregistrement de vos préférences](#)

10.3.4 Tri

Le classement est fonction de la langue choisie. On ne classe pas de la même manière l'anglais, l'espagnol, ou le danois.

Affichage



- Vous voyez deux fenêtres s'afficher sous le bouton "re-trier". Celle de gauche contient quelques mots en diverses langues, vous pouvez en ajouter. Le curseur qui apparaît sur la capture d'écran montre l'endroit où l'utilisateur saisit le mot, après avoir déjà saisi "(". Si vous n'y parvenez pas via le clavier, faites un copier-coller depuis vos textes.
- La fenêtre de droite affiche le classement de ces mots en fonction de la langue sélectionnée.

Rubriques connexes : [Langue](#), [Police](#), [Autres langues](#), [Enregistrement de vos préférences](#)

10.3.5 Autres langues

Pour utiliser une langue qui ne figure pas dans la liste, cliquez sur *Édition* et basez le nom de la nouvelle langue sur une des langues existantes. Choisissez une police capable d'afficher les caractères et symboles que vous souhaitez inclure. Le tri est géré comme dans la langue sur laquelle vous basez votre nouvelle langue.

Rubriques connexes : [Langue](#), [Police](#), [Tri](#), [Enregistrement de vos préférences](#)

10.3.6 Enregistrement de vos préférences

Enregistrez vos résultats avant de quitter de sorte que Oxford WordSmith Tools se souvienne, lors de la prochaine utilisation, de vos préférences concernant la police, votre langue par défaut n° 1 et votre langue par défaut alternative. Les résultats se trouveront dans `\wsmith4\language_choices.ini`.

Rubriques connexes : [Langue](#), [Police](#), [Tri](#), [Autres langues](#)

10.4 Minimal Pairs

10.4.1 fonction



C'est un programme permettant de repérer de possibles erreurs de frappe ou des paires de mots qui ne diffèrent que très peu (paires minimales). Vous pouvez par exemple obtenir une liste affichant ceci : ALEADY 5 et ALREADY 461. Cela signifie que vos textes comportent 5 occurrences mal orthographiées et 461 occurrences correctement écrites. Ce programme permet de repérer de possibles variantes ainsi que des fautes de frappe ou des anagrammes.

Rubriques connexes : [critères](#), [sélectionner vos fichiers](#), [affichage](#), [règles et paramètres](#), [exécution du programme](#).

10.4.2 critères

Une liste de mots est au format texte. Chaque ligne contient un mot et sa fréquence, séparés par des tabulations, par exemple

```
LE           75 432
ÉTAIT       9 895
```

ou

```
1           LE           75 432
2           ÉTAIT       9 895
```

Pour créer une liste comme celle-ci, utilisez [WordList](#). Sélectionnez (mettez en surbrillance) les colonnes contenant le mot et sa fréquence, cliquez sur le bouton ".txt", puis

- Décochez "Numéroter chaque ligne"
- Ligne à enregistrer = "toutes" (si 0-xxx s'affiche, remplacez 0 par 1)
- Colonnes à enregistrer = "colonnes en surbrillance"

Rubriques connexes : [fonction](#), [sélectionner vos fichiers](#), [affichage](#), [règles et paramètres](#), [exécution du programme](#).

10.4.3 sélectionner vos fichiers

- Sélectionnez votre liste de mots d'entrée (qui doit être au format texte brut) en cliquant sur le bouton à droite du champ d'édition et en sélectionnant le fichier *.txt qui est votre liste de mots.
- Si les lignes sont numérotées, cochez la case "le fichier *.txt est pré-numéroté".
- Si le fichier comporte un en-tête (WS3 insère par défaut trois lignes d'informations en en-tête), assurez-vous d'avoir défini correctement la valeur dans le champ "nombre de lignes d'en-tête à ignorer".
- Vous devez spécifier l'emplacement où les résultats sont enregistrés. Les résultats affichent toutes les fautes de frappes et les paires minimales que le programme trouve.
- Spécifiez également si :
 - la liste de résultats doit être numérotée
 - la fréquence des fautes de frappe potentielles doit être affichée
 - la règle ayant généré le résultat doit être affichée.

Rubriques connexes : [fonction](#), [critères](#), [affichage](#), [règles et paramètres](#), [exécution du programme](#).

10.4.4 affichage

Un exemple d'affichage serait :

```
418   ALTHOUGHT   (7)   ALTHOUGH(37975)
```

Ici, les lignes sont numérotées, et les nombres entre parenthèses signifient que ALTHOUGHT apparaît 7 fois et ALTHOUGH 37 975 fois.

Voici un autre exemple, en utilisant un texte médical en néerlandais (mots en minuscules) :

```
136   aplasie     (1)   aplasia(1)[L]
137   apyogene    (1)   apyogeen(1)[S]
138   arachnoideales (1)   arachnoidales(1)[I]
```

Ici, la ligne 136 a généré un rapport de différence d'une lettre, la ligne 137 une inversion, et la ligne 138 une insertion.

Autre exemple, la recherche d'anagrammes dans le journal britannique "Guardian" :

```
35   AUDIE   (7)   ADIEU(43)[A]
36   ABASS   (6)   ASSAB(16)[A]
37   AGUIAR  (6)   AURIGA(11)[A]
38   ALRED'S (6)   ADLER'S(18)[A]
39   ANDOR   (6)   ADORN(128)[A]
```

Rubriques connexes : [fonction](#), [critères](#), [sélectionner vos fichiers](#), [règles et paramètres](#), [exécution du programme](#).

10.4.5 règles et paramètres

Règles

Insertions (abxcd / abcd)

Cette règle recherche 1 lettre supplémentaire qui aurait pu être insérée, par exemple
HOWWEVER

Lettres inversées (abcd / acbd)

Cette règle recherche les lettres inversées, par exemple HOVEWER

Différence d'1 lettre (abcd / abxd)

Cette règle recherche une différence d'1 lettre, par exemple HOWEXER

Inclure les anagrammes (abcd / adbc)

Cette règle recherche les mêmes lettres que le mot de départ, mais dans un ordre différent, par

exemple HWVROEE

Paramètres :

lettres à ignorer si elles se trouvent en fin de mot :

Cette règle permet de spécifier les lettres devant être ignorées si elles se trouvent en fin de mot ; par exemple, si vous indiquez "s", vous ne verrez pas qu'il y a une erreur de frappe en comparant ÉLÉPHANT et ÉLÉPHANTS.

correspondance minimum de début de mot

Ce paramètre est réglé sur 1 par défaut, car lorsque vous recherchez des paires minimales, il y a forcément une partie au début qui correspond parfaitement... Par exemple, si on considère ALEADY, le programme ne s'intéressera qu'aux mots commençant par A pour la recherche des paires minimales. Si le paramètre est réglé sur 1, BLEADY sera pas considéré comme une paire minimale. Pour inclure tous les mots, réglez le paramètre sur 0. En conséquence, le programme sera alors 26 fois plus lent.

longueur de mot minimum

Ce paramètre indique la longueur de mot minimum que le programme prend en compte en recherchant d'éventuelles fautes de frappe. La longueur minimum est de 5 par défaut, ce qui signifie que les mots de 4 lettres seront ignorés. D'une part, cela permet d'accélérer le processus, et d'autre part, la plupart des fautes de frappe se trouvent dans des mots plus longs.

tous les mots commençant par...

Si vous sélectionnez cette option, le programme ignorera le paramètre suivant (fréquence de mots max.). Si vous saisissez une séquence telle que "F,G,H", le programme traitera tous les mots commençant par F ou G ou H, quelle que soit leur fréquence, et recherchera les paires minimales suivant les règles et paramètres évoqués ci-dessus.

fréquence de mots max

(ce paramètre n'est pas accessible si vous cochez "tous les mots commençant par") Quelle est la fréquence d'une faute de frappe ? Cela dépend de la taille du texte à partir duquel la liste de mots a été générée. La valeur par défaut est 10, ce qui signifie que tout mot apparaissant 11 fois est considéré comme étant correct, et non une faute de frappe.

Réglages à la sortie d'usine (restaure les valeurs par défaut)

Enregistrer les paramètres actuels (enregistre vos préférences de fichiers et de règles)

Récupérer les paramètres enregistrés (restaure les dernières préférences enregistrées)

Rubriques connexes : [fonction](#), [critères](#), [sélectionner vos fichiers](#), [affichage](#), [exécution du programme](#).

10.4.6 exécution du programme

- Cliquez sur "Calculer".

Vous voyez normalement quelques lignes du texte source. Certaines lignes et colonnes sont grisées, d'autres sont blanches : déplacez les numéros des lignes et colonnes pour que les véritables données s'affichent en blanc et les en-têtes ou numéros de ligne en grisé.

Pour interrompre le processus, cliquez sur "Stop".

La barre d'état qui se trouve en bas de l'écran montre le nombre de mots trouvés dans la liste de mots, le temps écoulé, ainsi que le temps restant estimé jusqu'à la fin de la tâche.

Cliquez sur "Résultats" pour voir le fichier de résultats une fois que vous avez terminé. Pour finir, cliquez sur "Quitter".

Rubriques connexes : [fonction](#), [critères](#), [sélectionner vos fichiers](#), [affichage](#), [règles et paramètres](#)

10.5 File Utilities

10.5.1 index



Ce sous-programme fournit quelques utilitaires pour une utilisation générale :

[Comparer deux fichiers](#)

[Découpeur de fichiers](#)

[Rechercher des doublons](#)

[Renommer](#)

Recherche de trous : recherche de "[trous](#)" dans des fichiers texte

Splitter

[Regroupeur](#)

10.5.2 Splitter

10.5.2.1 fonction de Splitter

C'est un sous-programme qui découpe les gros fichiers en de nombreux petits fichiers. Il y a divers éléments que **Splitter** doit connaître :

Séparateur de fin de texte

Le symbole qui marque la fin d'un texte : par exemple **[FF]** ou **<fin de l'histoire>** ou **</Text>** ou **!#** ou **[FF*]** ou **[FF?????]**

Limites :

- 1 Le marqueur de fin de texte doit se trouver en début de ligne dans le grand fichier original.
- 2 Sensible à la casse : **</Text>** ne renverra pas **</text>**.
- 3 Le premier caractère d'un séparateur de fin de texte ne peut être un [caractère de remplacement](#) comme **#**, ***** ou **?**.
- 4 ***** et **#** ne peuvent apparaître qu'une fois chacun dans un séparateur de fin de texte.

Splitter crée un nouveau fichier chaque fois qu'il rencontre le séparateur de fin de texte spécifié.

Dossier cible

Dossier dans lequel vous souhaitez copier les petits fichiers. Si vous travaillez en réseau, vous devez avoir la permission en écriture pour y accéder.

Tailles requises

Nombre de lignes minimum et maximum que peuvent comporter les petits fichiers (par défaut 2 et 30 000). Seuls les fichiers qui rentrent dans ces limites seront enregistrés. Cette fonctionnalité est utile pour extraire des fichiers à partir de CD-ROM. Si vous souhaitez avoir de petits fichiers texte, par exemple issus d'un journal, indiquer comme paramètre 1 plutôt que 2.

Une ligne correspond à un espace compris entre deux [Entrée](#).

Mettre la première ligne entre parenthèses

Si vous souhaitez ou non que la première ligne de chaque nouveau fichier texte soit placée entre **< >**. Ceci s'explique par le fait que très souvent, la ligne qui suit votre marqueur de fin de texte est en quelque sorte un en-tête. Si vous ne souhaitez pas insérer **<** et **>** dans la ligne, ne cochez pas la case.

Ligne de titre

Si vous savez quelle ligne de vos textes contient toujours le titre du sous-texte en question, indiquez-le ici, sinon laissez 0.

Rubriques connexes : [Regroupeur](#), [Noms de fichiers](#), [Caractères de remplacement](#), [Les boutons](#), Index de Text Converter.

10.5.2.2 Splitter : noms de fichiers

Splitter créera de nombreux petits fichiers à partir de vos gros fichiers.

Il crée de nouveaux [noms de fichiers](#) de la manière suivante :

Un dossier basé sur le nom du fichier source est créé. Des sous-dossiers sont créés si un dossier contient trop de fichiers.

Si un titre est détecté, chaque fichier contiendra le titre plus un nombre et enfin l'extension .txt.

S'il n'y a pas de titre, le nom du fichier se composera d'un nombre + .txt comme extension.

Par conséquent, un gros fichier appelé HELLO.DAT sera scindé en plusieurs petits fichiers :

```
\HELLODAT\1.txt
\HELLODAT\2.txt
...
\HELLODAT\1\512.txt
```

etc.

Conseils

1. Splitter commence la numérotation à 1 pour chaque session.
2. Remarque : les petits fichiers prendront probablement plus de place que le gros fichier original. Cela s'explique par le fait que le système d'exploitation de disque définit une taille minimum de fichier. Un fichier texte ne contenant qu'un seul caractère aura cette taille minimale, probablement plusieurs milliers d'octets. Même dans ce cas, il est conseillé de garder vos fichiers texte de façon à ce que chaque fichier soit un texte séparé, en utilisant Splitter. Cependant, lorsque vous créez des listes de mots et des listes de mots-clés, créez-les dans des [lots](#).
3. Les fichiers d'un CD-ROM copiés sur votre disque dur seront en lecture seule. Vous pouvez modifier cet attribut en utilisant Text Converter.

10.5.2.3 Splitter : caractères de remplacement

- # Le symbole dièse, #, est utilisé comme caractère de remplacement pour représenter tout **nombre**, donc [FF#] renvoie [FF3] ou [FF9987] mais pas [FF] ni [FF 9] (car ce dernier comporte un espace), ni [FFhello].
- * L'astérisque représente toute **chaîne**, donc [FF*] renverrait tous les exemples mentionnés ci-dessus. * représente le dernier caractère d'un symbole de fin de texte. Cela renverrait [FF suivi de toute séquence, et ce jusqu'à la prochaine balise <Entrée>.
- ^ Le symbole ^ représente toute **lettre** unique, ainsi [FF??] renvoie [FFZQ] mais pas les autres.
- ? Le point d'interrogation représente tout **caractère** unique (y compris les espaces, les signes de ponctuation, les lettres). Ainsi, parmi les exemples précédents [FF??] renvoie [FF 9] mais pas les autres.

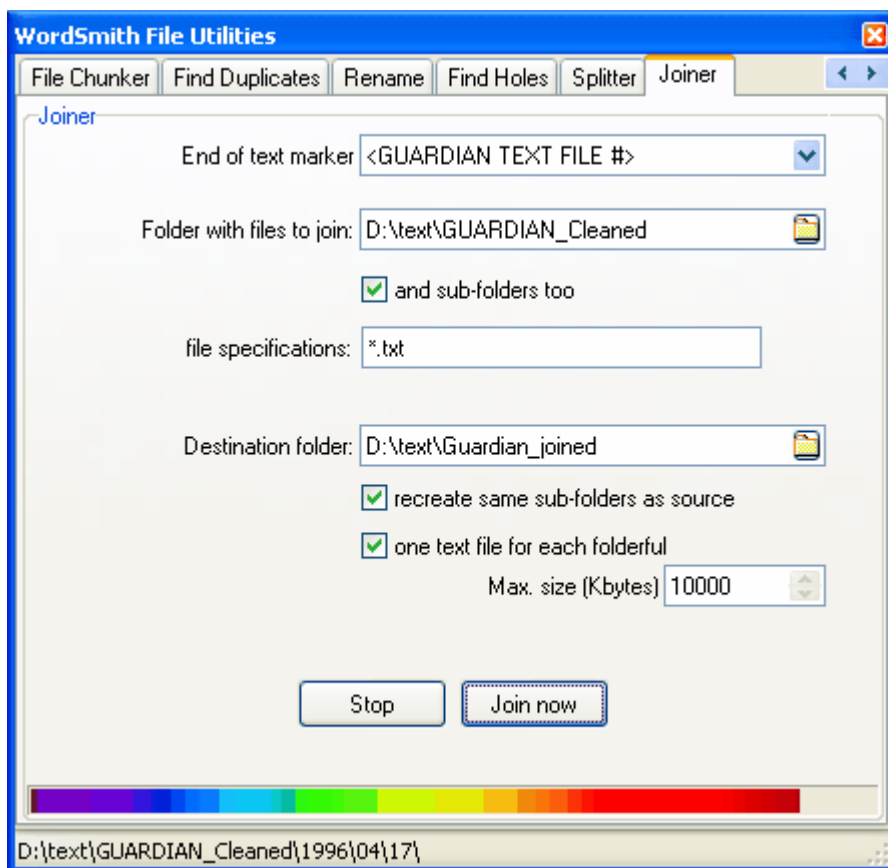
Pour représenter les caractères, et non les caractères de remplacement #, ^, ? ou *, placez-les entre guillemets comme ceci : "?" "#" "^" "*".

Rubriques connexes : [Paramètres](#)

10.5.3 regrouper des fichiers texte

C'est un sous-programme permettant de grouper de petits fichiers texte à des fichiers texte plus gros. Ceci peut vous être utile si vous voulez étudier le schéma d'un ensemble de textes mais que les textes isolés ne vous intéressent pas.

Lorsque vous cliquez sur **Regroupeur**, voici ce qui s'affiche :



Marqueur de fin de texte

Le symbole qui marque la fin d'un texte : par exemple **[FF]** or **<fin de l'histoire>** ou **</Texte>** ou **!#** ou **[FF*]** ou **[FF?????]**. Le marqueur de fin de texte apparaîtra en début de ligne dans le grand fichier original. Le signe # sera remplacé par le numéro du texte au cours du traitement des textes.

Dossier contenant les fichiers à joindre

C'est là que se trouvent actuellement les petits fichiers que vous souhaitez fusionner. Ils ne seront pas supprimés, il vous suffit de les fusionner avec le dossier cible.

sous-dossiers également

Cochez cette case si vous voulez traiter les sous-dossiers du dossier contenant les fichiers à regrouper.

spécifications des fichiers

Type de fichiers que vous souhaitez fusionner, par exemple ***.*** ou ***.txt** ou ***.txt ;*.ctx**.

Dossier cible

Dossier dans lequel vous souhaitez copier et fusionner les petits fichiers. Si vous travaillez en réseau, vous devez avoir la permission en écriture pour y accéder.

recréer les mêmes sous-dossiers que la source

Si vous cochez cette case, l'arborescence de la source sera reproduite. Dans l'exemple, tous les sous-dossiers de `d:\text\guardian_cleaned` seront créés dans `d:\text\guardian_joined`.

un texte par dossier

Si vous cochez cette case, l'ensemble des textes sources d'un dossier sera envoyé vers un seul fichier cible.

Taille maximum (Ko)

Taille maximale en ko pour les fichiers fusionnés. 1 000 signifie que chaque fichier fusionné comportera presque 1 mégaoctet de texte. Cela représente près de 150 000 mots pour un texte en anglais ne comportant pas de balises. Cela n'est valable que si l'option *un texte par dossier* n'est pas cochée.

Bouton stop

Effectue l'action indiquée dans la légende.

Rubriques connexes : Splitter, Index de Text Converter.

10.5.4 comparer deux fichiers

Le but

Cette fonctionnalité permet de vérifier si deux fichiers sont identiques ou pas. Il vous est sûrement déjà arrivé de faire des copies de fichiers, et, quelques semaines plus tard, vous ne vous souveniez plus de ce qu'ils contenaient. Ou alors vous avez utilisé le [découpeur de fichiers](#) pour copier un gros fichier sur des disquettes, et vous voulez vous assurer que la copie est identique à l'original.

Ce programme vérifie si

- a) ils ont la même taille
- b) ils ont le même contenu
(le programme analyse les deux fichiers, octet par octet, et vérifient s'ils correspondent)
- c) ils ont les mêmes attributs
(les attributs d'un fichier sont : "lecture seule" [vous ne pouvez pas modifier le fichier], "système" [un fichier considéré comme central pour le système d'exploitation], "caché" [un fichier tellement important que Bill Gates est peu enclin à ce que vous ayez connaissance de son existence])
- d) ils ont la même date.

Procédure

Spécifiez les deux fichiers devant être comparés, puis cliquez sur "Comparer".

Rubriques connexes : [découpeur de fichiers](#), [rechercher des doublons](#), [renommer](#)

10.5.5 découpeur de fichiers

Le but

Il s'agit de pouvoir couper un gros fichier en plusieurs parties afin de le copier sur plusieurs disquettes ou CD-ROM. Sinon, comment mettre un fichier de 5 Mo sur 3 ou 4 disquettes et le transférer sur un autre ordinateur ?

Plus tard, sur l'autre ordinateur, vous pouvez recompiler les fichiers coupés en un seul fichier.

Procédure : pour copier un fichier

1. Spécifiez le fichier devant être découpé (le gros que vous voulez copier)
2. Spécifiez le lecteur et le répertoire dans lequel vous voulez copier les morceaux. Pour A: il faudra insérer une disquette formatée pour chaque morceau.
3. Spécifiez la taille de chaque morceau (par défaut = 1 400 Ko, qui tient sur une disquette)
4. Spécifiez si vous voulez compresser pendant le découpage (compresse le fichier au fur et à mesure)
5. Cliquez sur copier.

Procédure : pour restaurer un fichier

1. Spécifiez votre premier morceau (le premier morceau créé avec ce programme)
2. Spécifiez dans quel dossier restaurer (à l'endroit où vous voulez les résultats)
3. Spécifiez si vous voulez supprimer les morceaux à la fin (si vous n'en avez pas besoin)
4. Cliquez sur restaurer.

Rubriques connexes : [comparer deux fichiers](#), [rechercher des doublons](#), [renommer](#)

10.5.6 rechercher des doublons

Le but

Cette fonctionnalité permet de vérifier si deux fichiers portant le même nom se trouvent dans deux dossiers différents. Il vous est sûrement déjà arrivé de faire des copies de fichiers, et, quelques semaines plus tard, vous ne vous souveniez plus de leur emplacement.

Ce programme ne fait que vérifier si les fichiers comparés portent le même nom. (Vous pouvez utiliser la fonction [Comparer deux fichiers](#) pour vérifier s'ils sont effectivement identiques). Le programme peut prendre en charge un grand nombre de dossiers à la fois, sa fonction étant de repérer les fichiers en double inutiles ou la réutilisation confuse de deux noms de fichiers.

Procédure

Spécifiez le premier dossier, puis cliquez sur "Rechercher". *Rechercher les doublons* parcourra le dossier et tous les sous-dossiers, puis affichera un compte-rendu des doublons trouvés.

Vous pouvez également spécifier deux dossiers différents (situés sur deux lecteurs différents, par exemple), et le processus compare les deux.

Rubriques connexes : [comparer deux fichiers](#), [découpeur de fichiers](#), [renommer](#)

10.5.7 renommer

Le but

Cette fonction permet de renommer plusieurs fichiers à la fois, dans un ou plusieurs dossiers. Certains fichiers peuvent avoir des noms très longs qui ne conviennent pas à certaines applications. Il est aussi très pénible de renommer de nombreux fichiers un à un.

Procédure

Indiquez les dossiers, ainsi que d'éventuels sous-dossiers à traiter, et les types de fichiers auxquels concernés.

La valeur par défaut est *Tous les fichiers *.**.

Spécifiez également un "masque de nom" ainsi que le nombre de départ.

Par exemple, avec le masque SOLEIL et 0 comme nombre de départ, le premier fichier trouvé, admettons qu'il s'appelle `un_fichier_long_et_tres_complique.txt`, sera renommé `SOLEIL0.txt`. Le suivant sera renommé `SOLEIL1.txt`, etc. Si le fichier suivant était `un_fichier_long_et_tres_compliqué.htm`, il serait renommé en `SOLEIL2.htm`.

Si vous cliquez sur "Rechercher des fichiers", une liste de tous les fichiers correspondant à vos préférences d'affiche. Si vous cliquez sur "Renommer", chaque fichier est renommé en fonction de vos paramètres.

Rubriques connexes : [comparer deux fichiers](#), [découpeur de fichiers](#), [rechercher des doublons](#)

10.6 Text Converter

10.6.1 but

Ce programme permet d'effectuer un "Rechercher et Remplacer", sur un nombre quasi-infini de fichiers.

Il est très utile pour parcourir un grand nombre de fichiers et les reformater comme vous le souhaitez, comme par exemple supprimer les espaces superflus, s'assurer que seuls les paragraphes comportent <Entrée> à la fin, modifier les caractères accentués, s'assurer de disposer des symboles Windows € etc.

conversion de texte

Pour effectuer un simple rechercher et remplacer, il suffit de saisir l'élément de recherche ainsi que l'élément devant remplacer ce dernier ; pour des conversions plus complexes, utilisez un [Fichier de conversion](#) pour indiquer à **Text Converter** quels symboles ou quelles chaînes

doivent être traduites. Il fonctionne sous Windows et enregistre les fichiers en utilisant le [jeu de caractères](#) Windows, mais il convertit le texte en utilisant des jeux de caractères DOS ou Windows. Text Converter permet de rendre vos textes utilisables avec un navigateur Internet tel que Netscape.

Il permet d'effectuer un "rechercher et remplacer" comme le font les logiciels de traitement de texte, mais il effectue cette opération dans un grand nombre de fichiers, les uns après les autres. Au cours de cette opération, il peut également remplacer **n'importe quel nombre** de chaînes, et pas seulement une.

Une fois le fichier de conversion préparé et les [Paramètres](#) spécifiés, **Text Converter** lit chaque fichier source et crée une nouvelle version ou remplace la version précédente, en fonction du [paramètre de réécriture](#) spécifié.

Vous pourrez voir les informations détaillées, c'est-à-dire combien de fois chaque chaîne a été trouvée et remplacée.

filtrage des fichiers

Vous voudrez peut-être vous assurer que les textes répondant à certains critères sont [placés dans les dossiers correspondants](#).

Astuce

Le moyen le plus simple de vérifier que vos fichiers texte sont tels que vous le souhaitez, surtout si vous avez un grand nombre de fichiers à convertir, est d'en copier quelques uns dans un dossier temporaire, et d'essayer votre fichier de conversion avec Text Converter. Vous pouvez alors voir si vous avez oublié de spécifier certaines conversions nécessaires. Une fois que vous vous êtes assuré que tout est en ordre, vous pouvez supprimer les fichiers temporaires.

Rubriques connexes : Sommaire de Text Converter, [Les boutons](#)

10.6.2 Text Converter : paramètres

1. Cliquez sur *Fichiers* (l'onglet supérieur gauche). Indiquez si vous souhaitez que le programme traite les sous-dossiers du dossier sélectionné. Il n'y a aucune limite quant au nombre de fichiers que Text Converter peut traiter à la fois.

2. Cliquez sur l'onglet *Conversion* puis :

3. Indiquez si vous souhaitez faire des copies des fichiers texte ou écraser les originaux. Pour écraser les fichiers originaux, il faut être sûr des modifications apportées ; enregistrer des copies peut entraîner un problème d'espace disque.

4. Spécifiez ce qui doit être converti, quels sont les mots de recherche et par quoi ils doivent être remplacés. Pour une conversion rapide, il est possible de simplement taper un mot à modifier ainsi que le mot devant le remplacer (par exemple *Une seule modification* de manière à ce que **responsable** devienne **responsable**). Vous pouvez également sélectionner votre propre [Fichier de conversion](#).

5. Une autre solution est de simplement mettre à jour les fichiers patrimoniaux dans la section Tous les fichiers, en cliquant sur :

Dos vers Windows ;

Unix vers Windows ;

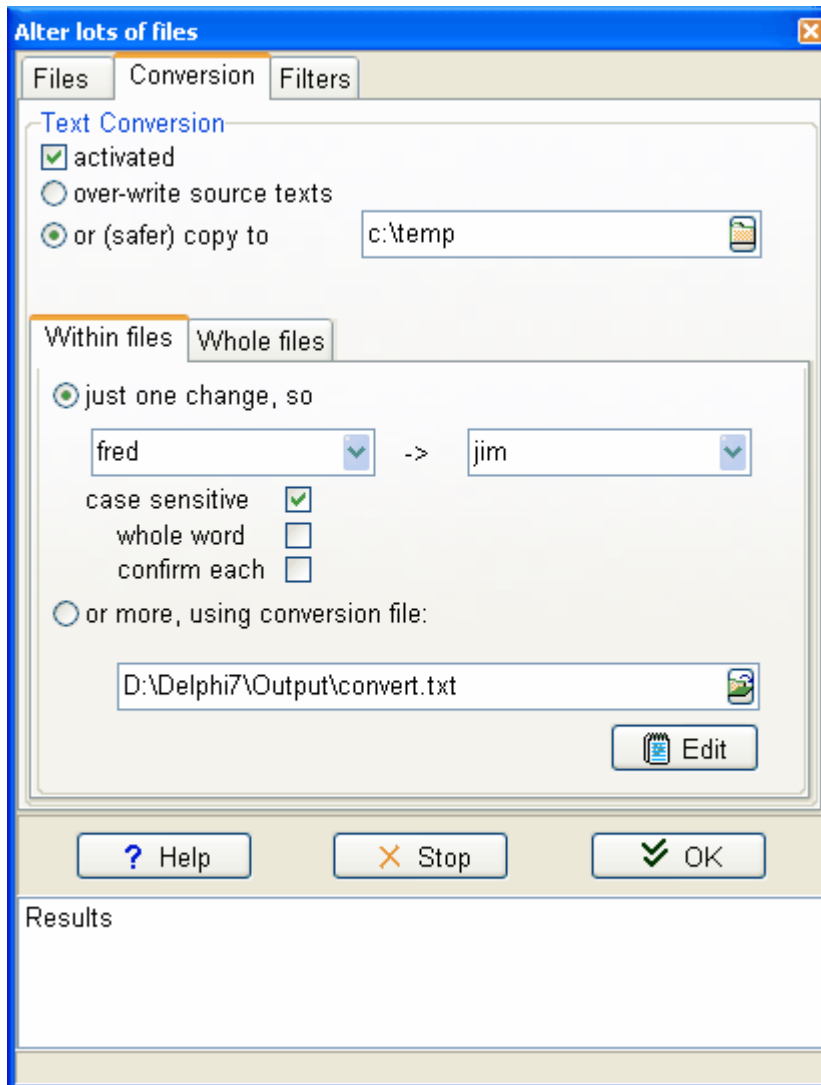
*MS Word *.doc vers *.txt ou*

en Unicode..

6. Si vous souhaitez que certains fichiers ne soient pas convertis ou si vous ne voulez pas de conversion mais simplement placer des fichiers dans des sous-dossiers appropriés, cliquez sur l'onglet [Filtres](#).

Si vous sélectionnez *Écraser les textes source*, Text Converter ira plus vite et utilisera moins d'espace disque. Bien évidemment, il faut que vous soyez sûr de vos codes de conversion de fichier avant de commencer.

Une portion d'espace disque sera tout de même utilisée si vous choisissez l'écrasement des fichiers. Le processus de conversion s'effectue, et si l'opération réussit, le fichier original est supprimé et la nouvelle version copiée. L'espace disque du dossier de destination doit être suffisant pour le plus gros de vos fichiers. Le processus est beaucoup plus rapide si les fichiers convertis se situent sur le même lecteur que les fichiers source. Si ce n'est pas le cas, il faudra préciser si vous acceptez que le même lecteur soit utilisé.



Cliquez sur *OK* pour démarrer; une liste de résultats s'affichera, comme le montre la capture d'écran ci-dessous.

Vous pouvez arrêter **Text Converter** à tout moment en cliquant sur le bouton Annuler ou en appuyant sur la touche Échap.

```
D:\Temp\test.txt converted = c:\temp\test.txt
1 done
```

Cliquez avec le bouton droit de la souris pour consulter le fichier source ou le fichier converti.

Rubriques connexes : Sommaire de Text Converter.

10.6.3 Text Converter : syntaxe

La syntaxe pour un [Fichier de conversion](#) est la suivante :

Seules les lignes commençant par / ou " sont utilisées. Les autres sont ignorées.

Chaque chaîne à convertir se présente sous la forme suivante : "A" -> "B". La chaîne d'origine, celle que vous recherchez, est comprise entre guillemets, suivie d'un espace, d'un trait d'union, du symbole > et de la chaîne de remplacement.

Codes de contrôle

Les codes de contrôle peuvent être symbolisés ainsi : {CHR(xxx)} où xxx est le code.

Exemples : {CHR(13)} correspond à un retour chariot, {CHR(10)} correspond à un saut de ligne, {CHR(9)} correspond à une tabulation, {CHR(12)} correspond à l'alimentation papier. Pour représenter le signe <Entrée> qui apparaît à la fin des paragraphes et parfois à la fin de chaque ligne, il faut saisir {CHR(13)}{CHR(10)}, qui correspond à un retour-chariot immédiatement suivi d'un saut de ligne.

Saisissez {CHR(34)} pour indiquer des apostrophes inversées doubles (utilisées en anglais).

Caractères de remplacement (*,?,# and ~)

* Vous pouvez utiliser l'astérisque comme caractère de remplacement. Ainsi, "<*" -> "" supprimera de votre texte toute chaîne située entre < >. "<head */head>" supprimera toute chaîne commençant par "<head " et se terminant par "/head>", même si des centaines de caractères se trouvent entre ces deux éléments. L'étendue de la recherche par défaut est de 1 000 caractères, et cette valeur peut aller jusqu'à 25 000. Le texte est lu segment par segment, avec une mise en tampon de 30 000 caractères. La valeur maximum peut donc convenir au début du texte, mais après cela, seul un espace de recherche de 1 000 caractères pourra être garanti. Supprimer une grande quantité de texte peut éliminer plus de texte que prévu si le texte n'est pas correctement balisé au départ. Il vous faudra outrepasser l'étendue de recherche par défaut en spécifiant une valeur entre parenthèses, par exemple "<head*(100)/head>". L'astérisque n'est pas toujours le premier ou le dernier symbole de la chaîne de recherche située entre guillemets.

L'astérisque garde en mémoire jusqu'à 1 000 caractères. "<div*(100)>" garde en mémoire tous les caractères jusqu'à > et peut les utiliser en remplacement : ainsi, "<div*(100)>" -> "[section *]" renverra [section 1 They Meet Again] si l'élément original comporte <div1 They Meet Again>. "<div*>" produirait les mêmes résultats, mais permet une recherche allant jusqu'à 1 000 caractères pour le signe >.

correspond à un nombre, quel qu'il soit. "<div#>" renverra <div1>, <div2> , <div468>, etc. Si le signe # est utilisé dans le remplacement, le même nombre sera utilisé dans le remplacement. Ainsi "<div#>" -> "[section #]" renverra [section 468] si l'original comporte <div468>.

? Le point d'interrogation représente tout caractère simple, sauf l'espace. Il est possible d'utiliser jusqu'à dix ? dans la chaîne de remplacement afin de reproduire le caractère désigné par le ?? dans la chaîne de recherche.

~ Le tilde signifie "sauf". ~"<p>" "<*" -> "" signifie qu'il faut supprimer tout ce qui se trouve entre crochets, sauf les occurrences de <p>.

{CHR(42)} désigne *, {CHR(35)} désigne #, {CHR(63)} désigne ? et {CHR(126)} désigne ~.

Mot entier, ignorer la casse, confirmer, espaces superflus

/C arrête le processus pour que vous confirmiez, avant de procéder à tout changement, que vous souhaitez continuer.

`/w` permet d'effectuer une recherche de mot entier (en s'assurant que l'altération n'ait lieu que si un [séparateur de mots](#) est présent de chaque côté) (`/W "rire"` renverra le mot "rire", mais pas "sourire" ni "sourires").

`/I` effectue une recherche sans tenir compte de la casse (`/I "restaurant"` -> "hôtel" remplace "restaurant" par "hôtel", RESTAURANT par HÔTEL et Restaurant par Hôtel, en respectant la casse le plus possible).

Il est possible d'effectuer des combinaisons, par exemple

`/IWC "the" -> "this"`

`/s` supprime tous les espaces superflus. Il réduit toute séquence de deux espaces ou plus à un seul espace. Il supprime également certains problèmes courants de formatage tel qu'un espace isolé situé après un retour chariot ou avant un signe de ponctuation tel qu'un point ou une virgule. `/s` peut être utilisé seul sur une ligne ou combiné à d'autres recherches.

Compléments (/A, /T et {v})

`/A` signifie "ajouter un texte". `/A "Ulan" START` insère Ulan au début, `/A "Bator" END` insère Bator à la fin du texte. Consultez le fichier `\wsmith4\convert.txt` pour en voir une utilisation.

`/T` signifie ajouter un titre. Ainsi `/T "<title>*</title>" -> ""` permet de rechercher `<title> ... </title>` et si l'élément est trouvé, l'expression spécifiée est insérée. Ainsi, votre navigateur affichera le titre dans la barre de titre.

`{v=""}` signifie qu'il faut garder l'élément en mémoire et l'utiliser dans une autre ligne du fichier de conversion lorsque l'élément `{v}` est trouvé.. `"26 Dec." -> "Boxing`

`Day" {v="Xmas"}` garde en mémoire la référence Xmas et `"1 May" ->`

`"Mayday" {v="after Easter"}` stocke `after Easter` pour qu'il soit utilisé dans une ligne ultérieure telle que `"/celebration/" -> "{v}"`. Supposons que votre texte comporte l'élément `26 Dec.` et `1 May`. Lorsque l'élément `"/celebration/` est trouvé, Xmas, est inséré si la mention la plus récente dans le texte est `26 Dec.` et `after Easter` sera inséré si la mention la plus récente dans le texte est `1 May`.

Consultez le fichier `\wsmith4\convert.txt` pour voir des exemples.

Rubriques connexes : Sommaire de Text Converter.

10.6.4 Text Converter : déplacer si...

Cette fonction permet de rechercher un mot ou syntagme donné dans chaque fichier, et, si le fichier contient le mot ou syntagme, il est déplacé vers un nouveau dossier.

Le but

Imaginons que vous disposez d'un ensemble de fichiers, certains contenant des dialogues de films, d'autres des renseignements sur la Grande Muraille de Chine ou l'anatomie des mouches. Vous recherchez les dialogues de films et souhaitez les placer dans un dossier nommé *Attendu*.

Procédure

1. Cliquez sur l'onglet *Filters* (en haut).
2. Cochez *Activé*.
3. Spécifiez un mot ou syntagme que le texte doit contenir. Attention à la casse.
4. Indiquez si vous souhaitez chercher le mot ou syntagme
 - n'importe où dans le texte
 - n'importe où dans le texte avant tel ou tel mot ou syntagme, ou
 - entre deux mots ou syntagmes différents.
5. Indiquez l'action à effectuer si les conditions sont remplies :
 - ne rien faire
 - copier vers un dossier, ou

- déplacer vers un dossier, ou
- supprimer le fichier.

Vous pouvez aussi créer un sous-dossier du nom du mot ou syntagme que vous avez choisi à l'étape 3.

Le programme peut ajouter l'extension `.txt` (utile lorsque qu'un fichier n'a pas d'extension).

Rubriques connexes : Sommaire de Text Converter.

10.6.5 Fichier de conversion Text Converter

Préparez votre fichier de conversion Text Converter avec un éditeur de [texte](#) tel que le Bloc-notes.

Le fichier `\wsmith4\convert.txt` peut vous servir de base.

Si des [caractères accentués](#) sont présents dans vos fichiers originaux, utilisez un éditeur DOS pour préparer le fichier de conversion si ces fichiers ont été écrits sous DOS ou un éditeur Windows s'ils ont été écrits avec un logiciel de traitement de texte Windows. Certains logiciels de traitement de texte Windows peuvent gérer les deux formats.

Le fichier peut contenir n'importe quel nombre de lignes de conversion, et chacune d'entre elles peut comporter deux chaînes délimitées par des guillemets (") et limitées à 80 caractères chacune.

Text Converter effectue toutes les modifications dans l'ordre, ainsi qu'il est spécifié dans le fichier de conversion. Remarque : toute modification peut avoir un impact sur les modifications suivantes.

Modifications augmentant la taille du fichier original

La plupart des modifications réduisent la taille du fichier original. Cependant, Text Converter peut gérer une augmentation de la taille du fichier original, tant qu'il y a suffisamment d'espace disque disponible.

Astuce

Pour supprimer les signes <Entrée> situés en fin de ligne mais pas en fin de paragraphe, étudiez tout d'abord les fins de paragraphe pour voir leurs signes distinctifs. Par exemple, si les paragraphes se terminent par deux <Entrée>, saisissez les lignes suivantes dans votre fichier de conversion :

```
"{CHR(13)}{CHR(10)}{CHR(13)}{CHR(10)}" -> "{%}"
```

(cette ligne permet de remplacer les deux <Entrée> par {%%} .) Ceci dit, il peut s'agir de n'importe quelle autre combinaison unique. Le processus sera plus rapide si vous remplacez la chaîne recherchée par une chaîne de la même longueur (ici, 4 caractères).

```
"{CHR(13)}{CHR(10)}" -> " "
```

Cette ligne permet de remplacer toutes les autres balises <Entrée> par un espace, pour conserver la séparation entre les mots.

```
"{%}" -> "{CHR(13)}{CHR(10)}{CHR(13)}{CHR(10)}"
```

Cette ligne permet de remplacer la combinaison {%%} par <Entrée><Entrée>, restaurant ainsi la structure initiale des paragraphes.

```
/s
```

Cette ligne permet de supprimer tous les espaces superflus

Rubriques connexes : [exemple de fichier de conversion](#), [syntaxe](#), Sommaire de Text Converter.

10.6.6 Text Converter : exemple de fichier de conversion

Vous pouvez copier l'intégralité ou une partie de ce qui suit dans le [presse-papiers](#) et le coller dans le bloc-notes.

```
[ commentaires : vous pouvez y noter ce que vous voulez, cette
partie sera ignorée ]
```

```
[ correction d'une faute d'orthographe ]
"responsable" -> "responsable"
```

```
[ convertir < > en [ ] et { } en ( ) ]
```

```
"<" -> "["
```

```
">" -> "]"
```

```
"}" -> ")"
```

```
"{" -> "("
```

```
/s
```

```
[efface tous les espaces superflus]
```

Le fichier `\smith4\convert.txt` est un échantillon de fichier de conversion destiné à être utilisé avec des fichiers texte du British National Corpus.

Rubriques connexes : Sommaire de Text Converter.

WordSmith Tools

Balises et notes

Section

XI

11 Balises et notes

11.1 présentation générale

À quoi servent les annotations ?

On appelle "texte annoté" du texte qui comporte des informations supplémentaires, intégrées au moyen de *balises*, par exemple "Nous <pronom> aimons <verbe> voyager <verbe>". Vous pouvez effectuer une concordance de mots ou de balises.

Vous pouvez *afficher* ces informations supplémentaires ou les *ignorer* pour n'avoir que le texte brut ("Nous aimons voyager"). **Oxford WordSmith Tools** vous permet de choisir quels éléments afficher et lesquels ignorer.

Vous pouvez *traduire* les balises et entités de référence [HTML ou SGML](#) : si votre texte comporte **É** , il est utile de pouvoir le changer en **É**.

Vous pouvez *sélectionner* au sein de fichiers texte, par exemple supprimer un en-tête ou ne garder que les conclusions, au lieu de vous servir de l'intégralité du texte.

Vous pouvez également demander à **Oxford WordSmith Tools** de ne sélectionner que les fichiers qui répondent à certains critères, par exemple comporter "sexe=f" dans la section d'en-tête dans un fichier texte où le locuteur est une femme.

Vous pouvez visualiser la sélection de balises grâce à l'option "Sélectionner des textes", puis le bouton [Afficher](#). Toutes les balises retenues s'affichent, et les balises ignorées sont remplacées par des espaces.

Rubriques connexes : [Gestion des balises](#), [Création d'un fichier de balises](#), [Affichage des balises les plus proches dans Concord](#), [Son et vidéo dans Concord](#), [Concordance des balises](#), [Types de balises](#), [Affichage des balises](#), [Utilisation des balises comme sélecteurs de texte](#), [Balises dans WordList](#)

11.2 types de balises

Indiquez le début et la fin de chaque type de balise et essayez de rester cohérents. N'utilisez que des symboles qui ne sont pas utilisés ailleurs dans vos textes.

huit marqueurs particuliers

Huit types de marqueurs sont importants pour les listes de mots, soit ceux qui marquent les [début et fin d'en-têtes](#), les [sections](#), les [phrases](#) et les [paragraphe](#)s. Saisissez-les dans les espaces prévus à cet effet dans [caractéristiques de texte](#).

balises entre 2 séparateurs

Ces balises signalent la partie du discours à laquelle un mot se rapporte ; on les utilise également en [HTML, XML, SGML](#) pour des "basculements", par exemple <H1> pour passer en style "titre H1" et </H1> pour repasser en normal. Utilisez les mêmes symboles d'ouverture et de fermeture pour toutes vos balises, la plupart du temps on utilise des chevrons, pour toutes vos balises (comme dans le British National Corpus qui utilise des balises [SGML](#)) : <Noun> , <Verb> , <Pronoun> .

références d'entité

Les formats [HTML, XML et SGML](#) utilisent des entités de référence pour les symboles ne faisant pas partie de l'alphabet, par exemple **é** ; **té** qui représente **été**.

Pour spécifier ces deux types de balisage, sélectionnez Paramètres/Liste de balises ou Paramètres/Caractéristiques de textes/Balises Une boîte de dialogue avec les boutons Texte à Ignorer et Parcourir s'affiche.

L'option [Balises à ignorer](#) permet de spécifier les balises que vous ne souhaitez pas voir apparaître dans vos résultats de concordance ou de listes de mots.

L'option [Balises à inclure](#) permet de spécifier un fichier de balises contenant les balises que vous souhaitez inclure dans vos concordances et listes de mots.

L'option [Balises à traduire](#) permet de spécifier les entités de référence que vous souhaitez convertir, comme `´`.

marqueurs multimédia

Vous pouvez baliser les fichiers texte en guise de référence pour les fichiers son et vidéo.

Admettons qu'un texte contienne : `bla bla bla ... bla bla` etc. Si vous effectuez une concordance de `bla bla`, la balise est incluse, et vous entendez le fichier mp3. Reportez-vous à la section [définition des balises multimédia](#).

Rubriques connexes : [Présentation générale des balises](#), [Gestion des balises](#), [Création d'un fichier de balises](#), [Affichage des balises les plus proches dans Concord](#), [Concordance des balises](#), [Affichage des balises](#), [Utilisation des balises comme sélecteurs de texte](#), [Son et vidéo dans Concord](#)

11.3 gestion des balises

ignorer toutes les balises

Spécifiez tous les symboles d'ouverture et de fermeture dans Modifier les paramètres | Balises | Balises à ignorer pour que ces balises n'apparaissent pas dans les listes de mots et les concordances.

exemple : `<*>` Toute l'expression commençant par le symbole `<` et finissant par le symbole `>` suivant est ignorée (avec un écart maximum de 1 000 entre `<` et `>`).

ignorer et inclure des balises

Si vous souhaitez n'ignorer que certaines balises, vous devez créer un [fichier de balises](#) qui liste toutes celles que vous souhaitez inclure. Elles apparaissent alors dans les listes de mots et les concordances.

Pour que la suite Oxford WordSmith Tools puisse lire ce fichier texte, cliquez sur "Fichier de balises" dans Paramètres. Ces balises seront incluses dans vos listes de mots, concordances, etc. comme n'importe quel mot ou suffixe.

exemple : supposons que vous ayez défini `<*>` dans "balises à ignorer", mais ayez défini `<title>`, `<body>` et `<conclusion>` dans les balises à inclure. WordSmith va alors inclure toutes les occurrences de `<title>`, `<body>` ou `<conclusion>` dans vos données mais ignorera `<introduction>`, `<Ulan Bator>`, `<threat>` etc.

Les balises à inclure ne seront actives que si le [nom de fichier](#) est visible et que vous avez appuyé sur *Charger* ou *Effacer*. Cliquer sur *charger*, permet de voir quelles balises sont issues du fichier de balises.

Si vous spécifiez le nom du fichier dans vos [paramètres par défaut](#) (`wshell.ini`) et incluez `autoload tagfile=YES`, les balises seront automatiquement chargées au lancement de WordSmith.

traduire les références d'entités par d'autres caractères

Si vous utilisez du texte balisé en [SGML ou HTML](#) il peut être utile de traduire ces symboles. Par exemple, les textes en SGML, XML, HTML utilisent `—` au lieu d'un tiret long. Pour ce faire, préparez un [fichier de balises](#) contenant les chaînes à traduire. Cliquez ensuite sur

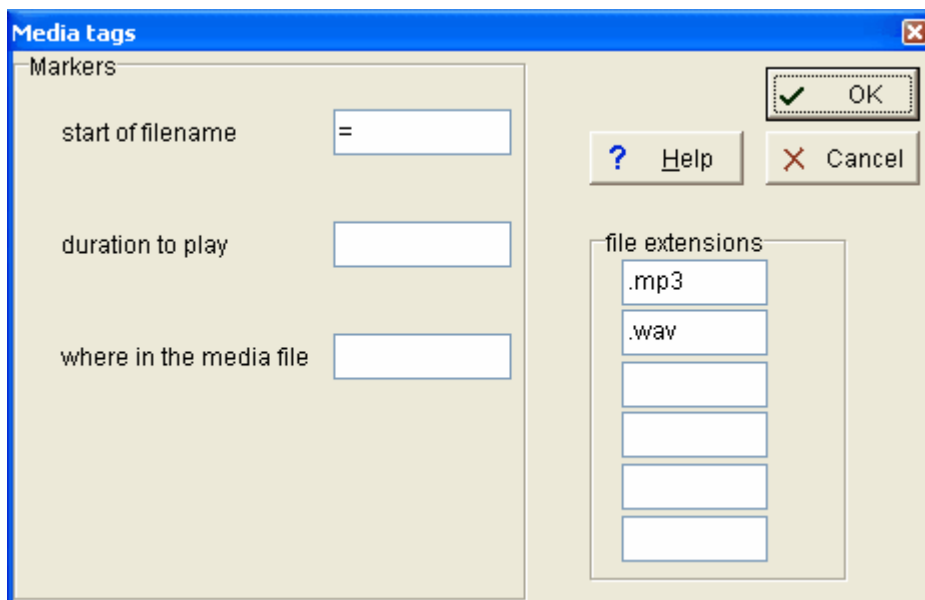
Modifier les paramètres | Balises et Notes| Fichier d'entités (entités à traduire) et sélectionnez votre fichier d'entités. WordSmith traduit alors toutes les références d'entités de ce fichier par les caractères correspondants.

Si vous spécifiez le nom du fichier dans vos [paramètres par défaut](#) (`wshell.ini`) et incluez `autoload tags to translate file=YES`, les entités seront automatiquement chargées au lancement de WordSmith.

Rubriques connexes : [Présentation générale des balises](#), [Création d'un fichier de balises](#), [Affichage des balises les plus proches dans Concord](#), [Concordance des balises](#), [Types de balises](#), [Affichage des balises](#), [Utilisation des balises comme sélecteurs de texte](#), [Enregistrement des valeurs par défaut](#)

11.4 balises multimédia

La capture d'écran ci-dessous donne un exemple de définition de balises multimédia. Pour obtenir la référence `` à partir du texte source,



il suffit d'utiliser le caractère = pour indiquer le début du nom de fichier. Ici, ce qui suit le signe = est une adresse web. Dans un texte contenant des balises comme `<sound$$C:\mysounds\talk.wav>`, indiquez le début du nom de fichier avec \$\$\$. Ce sont les extensions qui définissent les fichiers que votre ordinateur peut lire. Pour ce faire, votre ordinateur doit pouvoir traiter les fichiers son ou vidéo. C'est l'extension du fichier qui permet à Windows de déterminer comment le lire.

Durée et longueur de lecture

La durée de lecture et de l'emplacement de début de lecture sont calculés en secondes.

Si vous ne spécifiez rien, l'ensemble du fichier son ou vidéo sera lu.

Vos fichiers texte peuvent comporter des indications de durée et point de départ, par exemple `<sound$$C:\mysounds\talk.wav start:150 play:5>`. Indiquez la durée de lecture dans **play:** et l'emplacement dans le fichier media dans **start:**

S'il n'y a pas de marqueurs de durée et de position de départ, le premier nombre est, par défaut, la position de départ, et le second, la durée ; `<sound$$C:\mysounds\talk.wav 15 5>` signifie

donc "lire c:\mysounds\talk.wav 15 secondes après le début pendant 5 secondes".

paramètres par défaut

Les paramètres par défaut sont les suivants : lire les fichiers .mp3 et .wav. Puis, [enregistrez vos paramètres par défaut](#) pour la prochaine utilisation.

Rubriques connexes : [Son et vidéo dans Concord](#), [Présentation générale des balises](#), [Création d'un fichier de balises](#), [Gestion des balises](#), [Concordance des balises](#), [Affichage des balises les plus proches dans Concord](#), [Affichage des balises](#), [Types de balises](#)

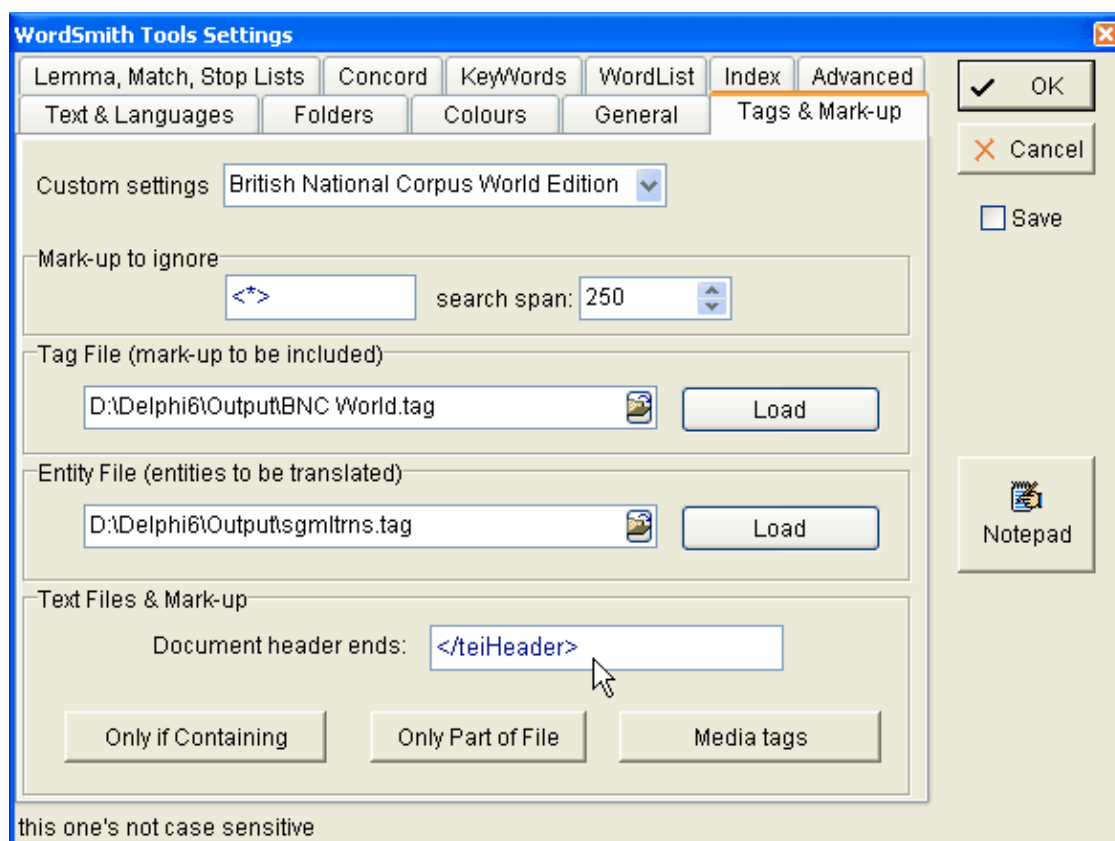
11.5 balises utilisées comme sélecteurs

Paramètres par défaut

Les paramètres par défaut sont les suivants : sélectionner *toutes* les sections de *tous* les textes choisis dans [Sélectionner des textes](#), avec suppression de toutes les balises placées entre crochets.

Cependant, **WordSmith** peut utiliser les balises pour sélectionner une section de texte et ignorer le reste. Il s'agit de la "sélection au sein de textes". Vous pouvez également choisir *parmi* les textes : c'est-à-dire que **WordSmith** peut analyser le début de chaque texte afin de déterminer s'il répond à certains critères.

Pour accéder à ces fonctions, cliquez sur *Paramètres | Modifier les paramètres | Balises | Seulement si les éléments contenus sont ou Uniquement une partie du fichier*.



Paramètres personnalisés

Cette fenêtre offre différentes possibilités qui facilitent les choix suivants. Si vous sélectionnez *British National Corpus World Edition* par exemple, comme sur la capture d'écran, `</teiHeader>` sera automatiquement ajouté dans la boîte "Fin d'en-tête du document".

Balises à ignorer

Pour supprimer des balises, par exemple dans un fichier [HTML](#), saisissez `< >` ou `[]` ou `< >`; `[]` dans *Balises à ignorer*. "Étendue de la recherche" correspond à la distance que doit parcourir WordSmith dans le texte pour trouver un symbole de fermeture comme `>` après un symbole d'ouverture comme `<`. La raison est que ces symboles sont également utilisés en mathématiques.

Fichier de balises, fichier d'entités

Voir [Création d'un fichier de balises](#).

En-tête du document

Pour supprimer l'en-tête d'un document (un en-tête répété, contenant des informations sur le copyright, comme au début de chaque texte du BNC), assurez-vous simplement que vous avez correctement spécifié les balises, comme dans l'exemple de `</teiHeader>`. L'option Paramètres personnalisés au-dessus vous indique automatiquement les choix appropriés. Pour des recherches plus complexes, vous pouvez utiliser les boutons [Seulement si les éléments contenus sont](#) ou [Uniquement une partie du fichier](#) que vous voyez sur l'image ci-dessus.

Dans quel ordre les différentes sélections sont-elles traitées ?

Que vous effectuiez une sélection parmi différents textes ou une sélection au sein d'un texte, WordSmith vérifie que chaque fichier texte répond à vos critères avant d'effectuer une concordance, une liste de mots, etc. WordSmith

1. Effectue une sélection [parmi des fichiers](#) pour voir s'ils contiennent les mots spécifiés
2. Supprime toutes les "[sections à supprimer](#)"
3. Si vous avez spécifié des "[sections à conserver](#)", WordSmith supprime tout ce qui ne s'y trouve pas
4. Coupe le [début de chaque ligne](#), si possible
5. Traite les références d'entités à [traduire](#)
6. [Ignore](#) toutes les balises qui n'ont pas été conservées (reportez-vous à la section "Balises à ignorer" de la capture d'écran ci-dessus).

Rubriques connexes : [Présentation générale des balises](#), [Création d'un fichier de balises](#), [Gestion des balises](#), [Concordance des balises](#), [Affichage des balises les plus proches dans Concord](#), [Affichage des balises](#), [Types de balises](#)

11.6 uniquement si les éléments contenus sont...

Le but

Vous pouvez être amené à ne vouloir traiter que les discours d'orateurs seniors, uniquement des textes de publicité ou des dialogues utilisables en cours. Cette fonction de WordSmith permet d'effectuer une recherche au sein de chaque texte, par exemple dans les en-têtes. Vous obtenez donc les bons textes, et ceux qui ne correspondent pas sont ignorés.

Supposons que vous ayez une large collection de textes (par exemple le British National Corpus) et que vous ne vous souveniez pas lesquels comportent des discours en langue anglaise prononcés par un orateur senior.

Sachant que le BNC utilise `stext>` pour les textes parlés, `sex=m` pour les hommes, `age=5` pour les gens âgés de plus de 60 ans, WordSmith peut filtrer votre sélection de textes. Le programme effectuera une recherche dans l'intégralité de tous les textes (et pas seulement au

niveau des balises ou des sections d'en-tête mais au niveau des [2 premiers Mo](#) du fichier) pour vérifier qu'il correspond à vos critères.

Vous pouvez spécifier jusqu'à 12 balises dans une limite de 80 caractères chacune. Attention, les balises sont sensibles à la casse (vous n'obtiendrez par exemple aucun résultat si vous tapez Age=5 au lieu de age=5).

Horizontalement, les options représentent les combinaisons liées par "ou". Verticalement, les combinaisons sont des liens "and". L'ensemble du bas représente les combinaisons "mais pas". Une fois vos textes traités, ceux correspondant à vos critères s'affichent dans la [fenêtre Sélectionner les fichiers texte](#). Vous pouvez enregistrer la liste dans vos [favoris](#) pour la réutiliser ultérieurement.

Exemples :

Vous ne voulez que des textes contenant à la fois le mot **chats** et le mot **chiens**: saisissez **chats** dans la première boîte et **chiens** en dessous.

Vous voulez que le texte comporte **roses** ou **violettes**, et **fleurs** doit également être présent : saisissez **roses** et **violettes** l'un à côté de l'autre dans les deux premières boîtes. Saisissez ensuite **fleurs** dans la boîte la plus à gauche de la ligne du dessous.

Vous voulez **réserver** ou **hôtel** mais uniquement s'ils ne se trouvent pas dans un fichier texte contenant **tourisme** ou **Europe** : saisissez **réserver** dans la première boîte, **hôtel** dans la boîte située à côté, et **tourisme*** et **Europe*** dans les deux premières boîtes de la ligne du bas.

Rubriques connexes : [Balises utilisées comme sélecteurs](#), [Sélection au sein de textes](#), [Filtrer vos fichiers texte](#) avec [Text Converter](#).

11.7 sélection au sein de textes

Couper le début de chaque ligne/paragraphe

Ceci est utile car certains corpus ont une quantité de codage fixe pour les détails de lignes à chaque début de ligne. Vous pouvez les couper (c'est-à-dire après chaque <Entrée>).

Déterminez le nombre de caractères à couper (max. 100), la valeur par défaut étant 0.

Saisissez -1 si vous souhaitez couper l'intégralité de ce qui se trouve avant le premier caractère alphabétique en début de ligne, et -2 pour tout couper jusqu'au premier retrait.

Sections à couper

Si vous utilisez des fichiers texte comportant des en-têtes en [SGML, XML ou HTML](#) (par exemple le British National Corpus), vous pouvez les supprimer de vos listes de mots, concordances, etc. comme indiqué dans [l'exemple d'en-tête](#).

Dans le cas de sélections plus compliquées, vous pouvez spécifier ce qui doit être coupé, en indiquant le début (par exemple <HEAD>) et la fin (par exemple </HEAD>). Vous pouvez couper jusqu'à 3 sections distinctes (de <HEAD> à </HEAD> ou de <BODY> à </BODY>). Cette fonction permet de couper une section autant de fois qu'elle apparaît dans un texte.

Sections à conserver (contextes)

Admettons que vous vouliez sélectionner une section de texte et couper le reste. Spécifiez la balise d'ouverture et de fermeture, par exemple <Intro> et <Body>

(seules les introductions sont analysées), ou **Marie:** à **Pierre:** (vous conservez alors uniquement le discours de Marie).

Il faut que chaque section soit délimitée par un symbole unique comme < ou > . Dans le cas de **Marie:** et **Pierre:**, il faut être sûr que pour tout ce que dit Marie, son nom est immédiatement

suivi de deux points et que ses paroles sont suivies de **Pierre:**. Attention à la casse (impossible de trouver **MARIE:**).

Si vous utilisez `<H1>` et `</H1>` appliquez cette fonction à un texte [HTML](#) vous obtenez les entêtes principaux, mais rien de plus.

Vous pouvez utiliser deux sections différentes, par exemple `<Intro>` et `</Intro>` pour obtenir l'introduction et `<Conclusion>` et `</Conclusion>` pour la conclusion. Le symbole de fin ne doit pas nécessairement ressembler au symbole d'ouverture de la section : vous pouvez par exemple garder `<INTRO>` et `</BODY>` et par conséquent couper la conclusion si elle apparaît après `</BODY>`.

Rubriques connexes : [Balises utilisées comme sélecteurs](#), [Uniquement si l'élément <x> contenu](#).

11.8 création d'un fichier de balises

Syntaxe des balises

Chaque balise est sensible à la casse.

Par convention, les balises commencent par `<` et se terminent par `>` mais le premier et le dernier caractère peuvent être différents.

Vous pouvez utiliser les caractères suivants :

* pour n'importe quelle séquence de caractères

? pour n'importe quel caractère seul

pour n'importe quel caractère numérique

N'utilisez pas le caractère "[" pour insérer des commentaires dans un fichier de balises car il peut servir comme symbole de balise. La caractère # représente un nombre (par ex. `<h#>` renvoie les balises `<h5>`, `<h1>`, etc.). Le caractère ? représente tout caractère unique (`<?>` renvoie `<s>`, `<p>`, etc.). Le caractère * représente n'importe quel nombre de caractères (par ex. `<u*>` renvoie `<u who=Fred>`, `<u who=Mariana>`, etc.). Autrement, préparez votre fichier de liste de balises de la même façon que les [listes d'arrêt](#).

Pour créer un nouveau fichier de balises `*.tag`, vous pouvez utiliser **Bloc-notes** ou **n'importe quel éditeur de texte**. **Saisissez une entrée par ligne**.

Il est possible de stocker autant de balises prédéfinies que vous le souhaitez. Cependant, plus vous en utilisez, plus la tâche sera conséquente pour WordSmith. Le processus s'en trouvera ralenti et utilisera une grande quantité de mémoire.

Balises à retenir (balises à inclure)

Un fichier de balises contient une simple liste de toutes les balises à retenir. Des fichiers de balises d'exemple (e.g `bnc world.tag`) sont inclus lors de l'installation pour le traitement des fichiers du BNC : vous pouvez créer un nouveau fichier de balises en ouvrant un fichier existant, en le modifiant, et en l'enregistrant sous un autre nom.

Les balises sont par défaut affichées dans la [couleur](#) standard (gris) mais vous pouvez spécifier le premier plan et l'arrière-plan pour les balises que vous souhaitez afficher différemment en saisissant

```
/colour="foreground on background"
```

```
par ex. <noun> /colour="jaune sur rouge"
```

Couleurs disponibles :

'Noir', 'Blanc', 'Crème',

'Rouge', 'Marron',

'Jaune'

'Navy', 'Blue', 'Light Blue', 'Sky Blue',

'Green','Olive','Dollar Green','Grey-Green','Lime',
'Purple','Light Purple',
'Grey','Silver','Light Grey','Dark Grey','Medium Grey'.

Les noms de couleurs ne sont pas sensibles à la casse (mais les balises le sont). Remarque: "grey" et "colour" doivent être écrits sous leur forme britannique.

Vous pouvez également saisir "/play media" si vous souhaitez qu'une balise donnée, lorsqu'elle est trouvée dans un fichier, entraîne [l'ouverture d'un fichier son ou vidéo](#). Considérons la balise suivante :

```
<sound *> /colour="blue on yellow" /play media
```

ainsi qu'une occurrence telle que celle-ci :

```
<sound c:\windows\Beethoven's 5th Symphony.wav>
```

ou

```
<sound http://www.political_speeches.com/Mao_Ze_Dung.mp3>
```

vous pourrez choisir d'[écouter le fichier *.wav ou *.mp3](#).

Enfin, vous pouvez saisir un libellé descriptif avec /description "libellé" :

```
<w NN*> /description "nom" /colour="Cream on Purple"
```

```
<EXTRAIT> /description "section"
```

```
<INTRODUCTION> /description "section"
```

```
<SECTION 1> /description "section"
```

Voici un exemple de ce qui s'affiche une fois que vous avez sélectionné un fichier de balises et appuyé sur "Charger" : Comme le montre l'icône, la première balise permet de jouer un média : "play media". On peut voir également que le paramètre "cream on purple" est appliqué aux noms.



Fichier d'entités (entités à traduire)

Un fichier de balises destiné à la traduction d'une entité par une autre utilise la syntaxe suivante : entité à rechercher + espace + entité remplaçant celle recherchée. Par exemple :

É; É

é; é

Vous trouverez un fichier de balises pour la traduction dans le dossier où est installé WordSmith (\wsmith4\sgmltrns.tag). Ce dernier est inclus lors de l'installation. vous pouvez créer un nouveau fichier de balises en ouvrant ce fichier exemple, en le modifiant, et en l'enregistrant sous un autre nom.

Rubriques connexes : [Présentation générale des balises](#), [Gestion des balises](#), [Affichage des balises les plus proches dans Concord](#), [Concordance des balises](#), [Types de balises](#), [Affichage des](#)

[balises](#), [Utilisation des balises comme sélecteurs de texte](#)

11.9 début et fin des segments de texte

WordSmith sait reconnaître 4 types de segments au sein d'un texte : phrases, paragraphes, en-têtes et sections. Attention, ce processus est sensible à la casse. Vous pouvez utiliser `<Entrée>` et `<Tab>` pour représenter la fin d'un paragraphe ou une tabulation dans vos textes. Pour les fins de phrases, vous pouvez également utiliser l'option `auto`.

Phrases

On peut avoir par exemple `<s>` pour indiquer le début d'une phrase et `</s>` pour la fin. Si vous conservez l'option `auto`, les fins de phrase sont déterminées par les points finaux, d'interrogation ou d'exclamation suivis d'une lettre majuscule.

Paragraphes

`<p *>` ou `<p>` peuvent marquer le début d'un paragraphe et `</p>` la fin, par exemple.

En-têtes

On peut avoir par exemple `<head>` pour indiquer le début et `</head>` pour la fin. Remarque : le British National Corpus marque les phrases avec des en-têtes. Par exemple :

```
<head>  
<s n="2"><w NN1>Introduction  
</head>
```

dans du texte `HXL`. Il ne serait pas utile de considérer `Introduction` comme une phrase, **WordSmith** ignore donc les balises de phrases au sein des en-têtes.

Sections

Par exemple, `<section *>` peut marquer le début et `</section>` la fin.

Ces différents types de segments sont comptabilisés à chaque fois qu'une balise de fin comme `</s>`, `</p>`, etc. apparaît dans le texte. Si le texte ne comporte aucune balise de fin `</p>`, les paragraphes seront comptés à chaque fois qu'une balise d'ouverture de paragraphe est trouvée.

Rubriques connexes : [Présentation générale des balises](#), [Gestion des balises](#), [Affichage des balises les plus proches dans Concord](#), [Concordance des balises](#), [Types de balises](#), [Affichage des balises](#), [Utilisation des balises comme sélecteurs de texte](#)

WordSmith Tools

Référence

Section

XIII

12 Référence

12.1 file types

The standard file-extensions used in WordSmith are

.cnc	concordance file
.lst	word list
.mut	mutual information list
.dcl	detailed consistency list
.tokens	word list index file
.kws	key words file
.kdb	key word database file
.ali	aligner list

WordSmith does not affect your Windows Registry, unlike most other programs. The reason is because this can make a system slow down and become unstable, and it also means that to remove WordSmith you can simply delete the folder it is in.

In the Controller's General settings, or on installing, however, you can if you wish associate the current file-types with WordSmith in the Registry. The advantage of this is that Windows should know what Tool to open your data files with.

12.2 international versions

WordSmith can operate with a series of interfaces depending on the language chosen.



If you choose French this is what you see in all of WordSmith.



See also: [acknowledgements](#)

12.3 version 32 bits

Cette version de **Oxford WordSmith Tools** est une réécriture complète de l'ancienne version 16 bits, incluant de nombreux changements. Parmi eux :

- [noms de fichiers](#) longs
- meilleure gestion des [basles et des entités](#), y compris la [Concordance des balises](#)
- réutilisation des anciens travaux (mais il est conseillé de les réenregistrer au format 32 bits). Le logiciel vous proposera de "mettre à jour" les fichiers de données qui sont à l'ancien format.
- [gestion des fichiers zip](#)
- export de données plus aisé vers Microsoft Word et [Excel](#)
- gestion du texte au format Unicode, permettant ainsi de traiter davantage de [langues](#)
- possibilité de [modifier les données](#) au fur et à mesure, par exemple dans le cas d'une lemmatisation spécifique à une langue
- le nombre de lignes de données n'est plus limité à 16 000. Une liste de données admet une limite théorique dépassant les 134 millions de lignes.

Rubriques connexes : [Adresses de contact](#).

12.4 remerciements

Oxford WordSmith Tools s'est considérablement développé au fil des années. À l'origine, chaque outil fut créé car j'avais besoin d'un outil pour une tâche spécifique dans mon travail de linguiste. Les premières versions ont été écrites pour DOS, puis Windows™ est arrivé sur le marché.

L'outil **Concord**, en revanche, a une histoire un peu différente. Il a été développé à partir de *MicroConcord* que Tim Johns et moi-même avons conçu pour DOS et que Oxford University Press a publié en 1993. **Concord** comporte de nombreuses fonctionnalités supplémentaires dans cette version Windows et l'ensemble du code a été réécrit. Cependant, les fonctions essentielles de conception étaient déjà présentes dans *MicroConcord*.

La première version a été écrite avec Borland™ Pascal et Assembler pour les sections nécessitant une exécution rapide. Les programmes ont par la suite été convertis en Delphi™ 16 bits ; cette version est une version 32 bits écrite avec Delphi 6. Les chaînes urgentes ont quant à elles été écrites avec Assembler.

Je tiens à remercier les générations d'étudiants et de collègues du [département d'anglais](#), University of Liverpool, ainsi que le MA Programme in Applied Linguistics at the Catholic University of São Paulo, pour leurs commentaires sur les différents aspects de la suite logicielle (y compris les bugs) ainsi que leurs suggestions sur les fonctionnalités à ajouter. Les chercheurs de nombreux autres pays ont également joué le rôle de testeurs alpha et beta. Je les remercie de leur patience et de leurs commentaires. Je tiens également à remercier Nell Scott ainsi que tous les autres membres de ma famille qui m'ont toujours apporté un immense soutien ainsi que leurs commentaires et suggestions.

Mike Scott

Pour toute suggestion de développement pour **WordSmith Tools**, vous pouvez m'envoyer un [courrier électronique](#)

12.5 bibliographie

- Aston, Guy, 1995, "Corpora in Language Pedagogy : matching theory and practice", in G. Cook & B. Seidlhofer (eds.) *Principle & Practice in Applied Linguistics: Studies in honour of H.G. Widdowson*, Oxford: Oxford University Press, 257-70.
- Aston, Guy & Burnard, Lou, 1998, *The BNC Handbook*, Edinburgh: Edinburgh University Press.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan, 2000, *Longman Grammar of Spoken and Written English*, Harlow: Addison Wesley Longman.
- Clear, Jeremy, 1993, "From Firth Principles: computational tools for the study of collocation" in M. Baker, G. Francis & E. Tognini-Bonelli (eds.), 1993, *Text and Technology: in honour of John Sinclair*, Philadelphia: John Benjamins, 271-92.
- Dunning, Ted, 1993, "Accurate Methods for the Statistics of Surprise and Coincidence", *Computational Linguistics*, Vol 19, No. 1, pp. 61-74.
- Fillmore, Charles J, & Atkins, B.T.S, 1994, "Starting where the Dictionaries Stop: The Challenge of Corpus Lexicography", in B.T.S. Atkins & A. Zampolli, *Computational Approaches to the Lexicon*, Oxford: Clarendon Press, pp. 349-96.
- Katz, Slava, 1996, Distribution of Common Words and Phrases in Text and Language Modelling, *Natural Language Engineering* 2 (1), 15-59
- Murison-Bowie, Simon, 1993, *MicroConcord Manual: an introduction to the practices and principles of concordancing in language teaching*, Oxford: Oxford University Press.
- Nakamura, Junsaku, 1993, "Statistical Methods and Large Corpora: a new tool for describing text types" in M. Baker, G. Francis & E. Tognini-Bonelli (eds.), 1993, *Text and Technology: in honour of John Sinclair*, Philadelphia: John Benjamins, 293-312.
- Oakes, Michael P. 1998, *Statistics for Corpus Linguistics*, Edinburgh: Edinburgh University Press.
- Scott, Mike, 1997, "PC Analysis of Key Words - and Key Key Words", *System*, Vol 25, No. 2, pp. 233-45.
- Sinclair, John M, 1991, *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.
- Stubbs, Michael, 1986, "Lexical Density: A Technique and Some Findings", in M. Coulthard (ed.) *Talking About Text: Studies presented to David Brazil on his retirement*, Discourse Analysis Monograph no. 13, Birmingham: English Language Research, Univ. of Birmingham, 27-42.
- Stubbs, Michael, 1995, "Corpus Evidence for Norms of Lexical Collocation", in G. Cook & B. Seidlhofer (eds.) *Principle & Practice in Applied Linguistics: Studies in honour of H.G. Widdowson*, Oxford: Oxford University Press, 245-56.
- Tuldava, J. 1995, *Methods in Quantitative Linguistics*, Trier: WVT Wissenschaftlicher Verlag Trier.
- Youlmans, Gilbert, 1991, "A New Tool for Discourse Analysis: the vocabulary-management

profile", *Langue*, Vol 67, No. 4, pp. 763-89.

12.6 bugs

Tous les programmes informatiques comportent des bugs. Vous avez peut-être déjà vu le message "Erreur de protection générale" lorsque vous utilisez des suites de logiciels de graphisme ou de traitement de texte onéreuses. Si vous obtenez ce message en utilisant un des programmes, un rapport s'affichera, vous donnant des informations "générales" sur votre ordinateur et des "détails" sur l'erreur. Ces informations me permettront de résoudre le problème et seront enregistrées dans un petit fichier texte nommé `wshell.elf`, `concord.elf`, `wordlist.elf`, etc. Lorsque vous quittez le programme, vous pourrez m'envoyer un courrier électronique pour faire rapport du problème.

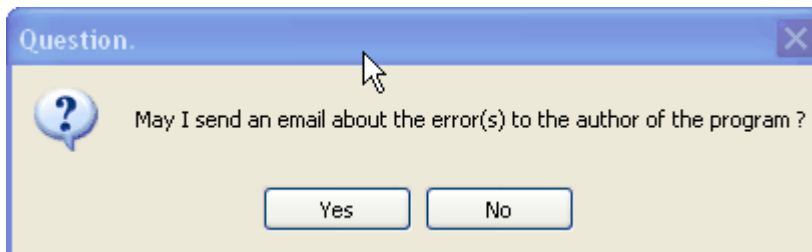
Vous verrez ceci s'afficher en premier lieu :



Vous devrez quitter après avoir cliqué sur OK ou bien WordSmith peut continuer en dépit du problème.

En général, le programme à l'origine du problème peut continuer son exécution. Vous pouvez y revenir sans avoir à fermer le [contrôleur](#) Oxford WordSmith Tools, récupérer vos [résultats enregistrés](#) à partir du disque et recommencer. Si cela ne fonctionne pas, essayez de complètement fermer Oxford WordSmith Tools ou Windows, puis de redémarrer.

Lorsque vous quittez, vous devriez obtenir ceci :



Si vous cliquez sur *Oui*, votre client de messagerie se lance et un texte apparaît, contenant les informations que je peux utiliser pour essayer de corriger le problème à l'avenir. ***Ce courrier électronique ne sera envoyé qu'après avoir cliqué sur Envoyer dans votre client de messagerie. Le message ne sera envoyé qu'à moi et ne sera transmis à personne d'autre. Vérifiez le texte du message avant de l'envoyer. Il contient des informations sur le système d'exploitation, la mémoire vive, l'espace disque, la version de WordSmith ainsi que des détails sur les routines utilisées au moment où le problème est survenu.***

[messages d'erreur](#)

Ils vous préviennent lorsque des problèmes surviennent au cours de l'utilisation du programme. Il s'agit par exemple de messages lorsqu'il n'y a plus d'espace disque disponible ou que vous avez saisi un [nom de fichier](#) non valide ou un nombre contenant une virgule.

Rubriques connexes : [connexion](#), [résolution des problèmes](#).

12.7 Jeux de caractères

12.7.1 présentation générale

Vous avez besoin de "[texte brut](#)" pour utiliser WordSmith. Les fichiers Microsoft Word *.doc ne sont pas pris en charge car ils contiennent du texte mais également de nombreuses informations qui ne sont pas visibles pour l'utilisateur.

Pour traiter un texte avec un ordinateur, l'encodage du texte doit pouvoir être détecté par les programmes. Lors du traitement, le logiciel ne voit qu'une longue chaîne de nombres, et ces derniers doivent correspondre à ce que l'être humain reconnaît comme "caractères". Pour de nombreuses langues comme l'anglais dont l'alphabet est restreint, l'encodage peut être géré en n'utilisant qu'un seul octet par caractère. En revanche, pour une langue telle que le chinois, qui utilise une large gamme de caractères, un système à un seul octet ne suffit pas. C'est la raison pour laquelle les systèmes à plusieurs octets ont été créés. Il est évident que si un texte anglais est encodé dans un jeu de caractères à plusieurs octets, cela créera un fichier plus lourd que s'il était encodé à un seul octet par caractère, et cela utilise inutilement de l'espace disque. C'est pourquoi à l'heure actuelle, les jeux de caractères à un octet sont toujours largement répandus.

En pratique, vos textes seront encodés dans un système Windows à un seul octet, les textes plus anciens dans un système DOS à un octet, et les plus récents, surtout ceux en chinois, japonais et grec, seront encodés en Unicode. Ce qui importe pour l'utilisateur, c'est l'apparence de chaque caractère. Cependant, WordSmith ne peut pas trier les caractères correctement, ou même reconnaître le début ou la fin d'un mot si l'encodage n'est pas correct. WordSmith doit savoir (ou pouvoir détecter) quel est le système d'encodage de vos fichiers. Il peut effectuer certaines tâches en arrière-plan. Mais comme le logiciel ne "comprend" pas les mots qu'il rencontre, il est préférable de [définir le jeu de caractères à l'avance](#), surtout si vous traitez des textes rédigés en allemand, espagnol, russe, grec, polonais, japonais, farsi, arabe, etc.

Les trois jeux de caractères principaux, chacun ayant des caractéristiques propres, sont [Windows](#), [DOS](#), et [Unicode](#).

Astuce

Pour vérifier les résultats après modification du code, cliquez sur [Sélectionner des textes](#) puis affichez le fichier. Lors de l'affichage, vous pouvez modifier les caractéristiques du texte jusqu'à ce que l'apparence convienne. Si vous ne parvenez pas à obtenir un affichage correct, c'est que votre texte n'est pas un fichier [texte brut](#) mais qu'il provient directement d'un traitement de texte. Dans ce cas, ouvrez à nouveau votre fichier dans un logiciel de traitement de texte puis enregistrez-le en [texte brut](#) au format Windows (lequel est plus récent que les formats DOS).

Rubriques connexes : [Sélectionner des accents et des symboles](#), [Caractères accentués](#); [Sélectionner une langue](#)

12.7.2 accents et symboles

Cette fonctionnalité des [Paramètres de concordance](#) et des autres boîtes de dialogue vous permet d'insérer des symboles et des caractères accentués dans vos termes de recherche, d'exclusion ou de contexte etc.

Sélectionnez simplement le symbole et faites-le glisser avec la souris pour le placer où vous le souhaitez. Sinon, cliquez deux fois dessus.

Vous pourrez afficher les caractères accentués disponibles en fonction de divers paramètres actuels tels que le [jeu de caractères](#) et les [caractéristiques du texte](#), ainsi que de nombreux

autres symboles tels que les guillemets français. Certains de ces symboles (®,©,™ et les apostrophes arrondies) ne peuvent être recherchés que si vos textes ont été sauvegardés au format [ANSI](#) Windows.

Rubriques connexes : [Sélectionner une langue](#)

12.7.3 ansi et ascii

ASCII, ANSI, Texte seulement et Texte DOS se réfèrent tous à du texte brut.

La plupart des logiciels de traitement de texte insèrent des codes masqués dans les fichiers texte pour marquer les numéros de page, les caractères en gras, etc. Oxford WordSmith les prend en charge, cependant vous obtiendrez des résultats plus "propres" si vous utilisez du texte brut sans codes masqués.

Si vos textes sources ont été enregistrés au format "Texte seulement avec retours à la ligne" il y aura un <Entrée> tous les 70 ou 80 caractères, à la fin de chaque ligne de texte. Si vos textes ont été enregistrés au format "Texte seulement", chaque <Entrée> sera en fait un saut de paragraphe. Il est conseillé de sauvegarder au format "Texte seulement".

L'utilitaire de Windows, **Bloc-notes** (*Démarrer | Programmes | Accessoires*), permet de créer des fichiers texte brut au format *.txt. Ce format utilise des jeux de caractères de base : les lettres de a à z, les nombres, les signes de ponctuation usuels etc. La principale différence réside dans les caractères accentués. Pour de plus amples informations à ce sujet, reportez-vous à la rubrique [jeux de caractères](#).

Rubriques connexes : [HTML,SGML et XML](#).

12.7.4 DOS

DOS (format texte prédécesseur de Windows) offrait une gamme de jeux de caractères nommés "pages de code". Ils partageaient tous les mêmes codes pour l'alphabet anglais standard (**a, par exemple, correspond au code 97**) ainsi que les signes de ponctuation courants, mais **comprenaient divers symboles pour le dessin des boîtes, les accents pour les langues étrangères, etc.**

Si vous traitez des textes en allemand, espagnol, russe, grec, polonais ou autres, vous devrez probablement rechercher la page de code utilisée lorsque les textes ont été tapés.

Par exemple, le caractère ã est encodé d'une certaine manière en page de code 850 (multilingue) et d'une autre en page de code 860 (portugais). En revanche, il n'est pas du tout disponible dans le codepage 437 (la page de code par défaut au Royaume-Uni et aux États-Unis). Pour modifier ou consulter les pages de code, référez-vous à un manuel DOS ou effectuez une recherche à ce sujet sur Internet.

Au moment du chargement, **Oxford WordSmith Tools** détecte la page de code DOS actuelle. Ainsi, la page de code ne devra être modifiée que si vous utilisez des textes créés lorsqu'une autre page de code était active.

12.7.5 Windows

Les codes des jeux de caractères Windows sont différents des jeux DOS ou Unicode. Par exemple, le symbole £ correspond au code 156 dans DOS et 163 dans Windows. Dans Windows 95 ou supérieur, les polices non occidentales peuvent être activées grâce à Microsoft Plus. Si vos textes ont été écrits avec un logiciel de traitement de texte Windows puis [enregistrés au format texte](#) dans Windows, les caractères accentués obéissent aux codes Windows. Windows vous donne accès à quelques symboles supplémentaires par rapport à DOS (par exemple ®,©,™ et les apostrophes arrondies).

Le format Windows **Occidental** (1252) comprend :

Anglo-saxon, anglais, anglais moyen, basque, catalan, danois, espagnol, finnois, français, allemand, islandais, italien, néerlandais, norvégien, vieux norois, portugais, suédois

Le format Windows **Baltique** (1257) comprend :

Estonien, Letton, Lituanien

Le format Windows **Europe Centrale** (1250) comprend :

Albanais, bosniaque, croate, hongrois, polonais, roumain, serbe, slovaque, slovène, haut sorabe, bas sorabe, tchèque

Le format Windows **Cyrillique** (1251) comprend :

Biélorusse, bulgare, macédonien, russe, serbe (1251), ukrainien

Le format Windows **Grec** (1253) prend en charge le grec

et le format Windows **Turque** (1254) prend en charge le turc.

12.7.6 Unicode

Un format texte standard utilise 2 octets par caractère. Cela permet d'afficher plus de 65 000 caractères et symboles différents, et rend possible l'affichage de langues telles que le chinois, le japonais, le cherokee et bien d'autres.

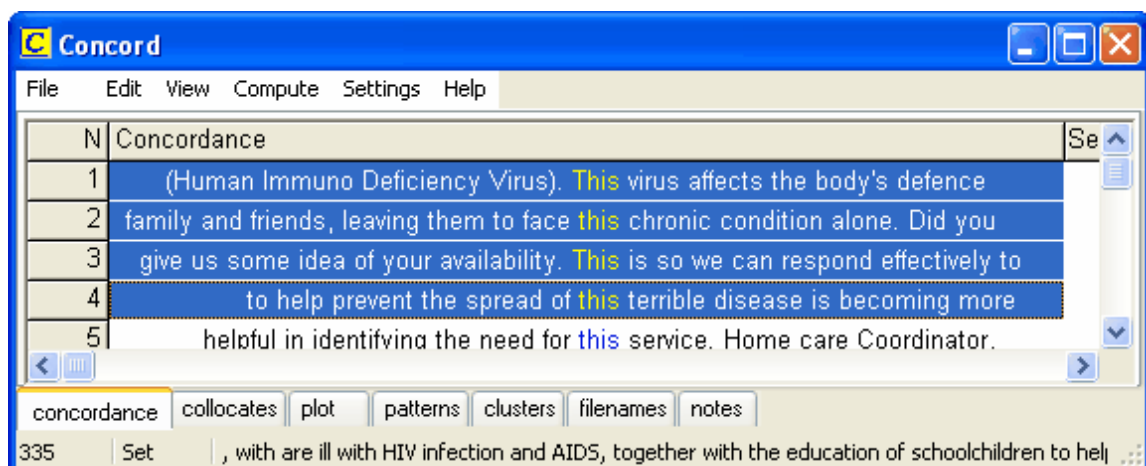
Lorsque vous sélectionnez des textes, vous pouvez cliquer sur un bouton pour tester si les textes sont encodés en Unicode.

12.8 presse-papiers

Vous pouvez sélectionner un bloc de données en utilisant les touches directionnelles et la touche Maj, ou le pointeur de la souris. Appuyez ensuite sur Ctrl+Inser ou Ctrl+C pour les copier dans le presse-papiers. Si vous ouvrez un logiciel de traitement de texte, vous pouvez coller (ou faire un "collage spécial") de la zone sélectionnée dans votre texte. Ce processus est généralement plus simple que [enregistrer en tant que fichier texte](#) (ou [imprimer](#) dans un fichier) et il peut également prendre en charge de nombreuses balises graphiques.

Exemple

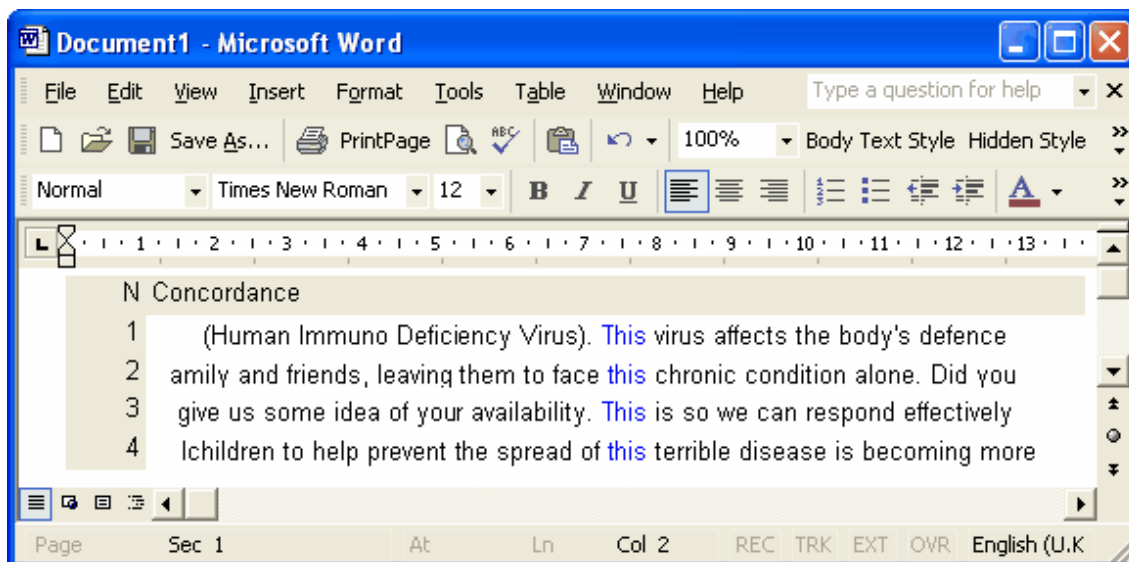
1. Sélectionnez des données. Ici, j'ai sélectionné les 4 premières lignes (sur un total de 335) d'une concordance, mais uniquement le texte visible, aucune information de classe ou de noms de fichiers.



2. Maintenez la touche Ctrl enfoncée et appuyez sur Inser ou C. Les données sont donc placées dans le "presse-papiers" de Windows et peuvent être collées dans n'importe quelle application telle que Word, Excel ou Bloc-notes.

Les données sont automatiquement placées dans le presse-papiers sous deux formats différents :

en tant qu'image



Vous utiliserez probablement ce format pour votre dissertation et y serez obligé dans le cas de données schématisées. Dans cette concordance, vous n'obtiendrez que les mots visibles dans votre ligne de concordance (et pas la ligne entière).

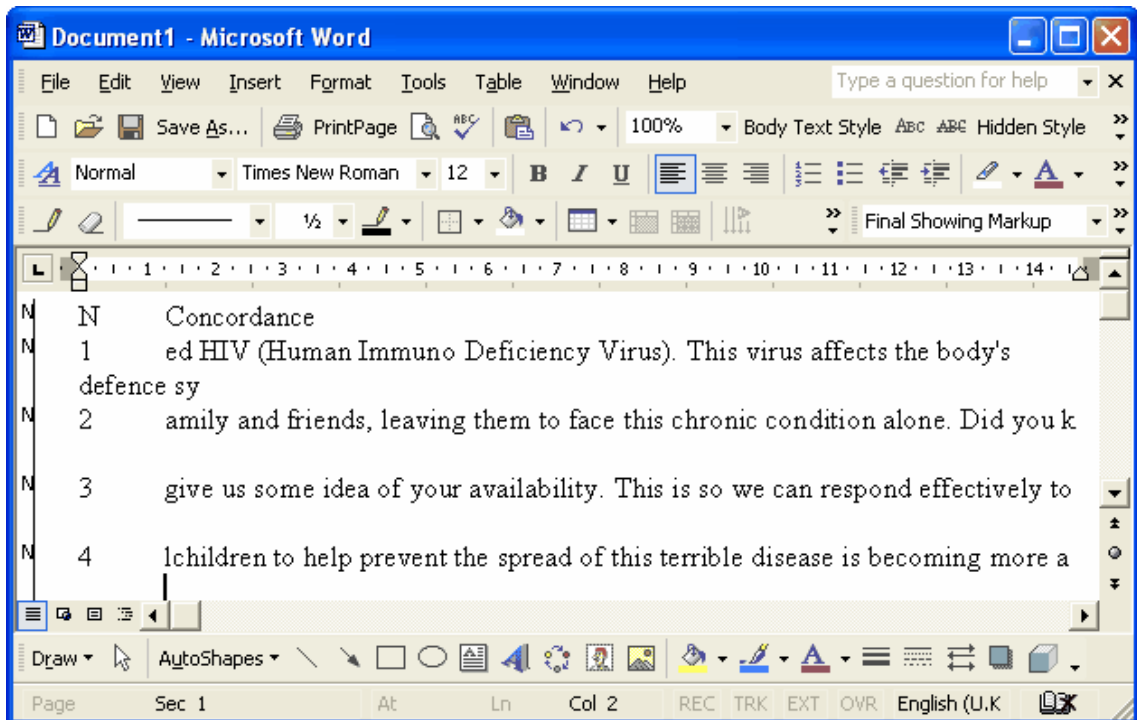
C'est un graphique qui inclut des couleurs ainsi que des données graphiques. Si vous cliquez ensuite sur le graphique, vous pouvez modifier sa taille globale et éditer chaque mot ou ligne graphique le composant (le processus n'est cependant pas très simple).

Pour obtenir ceci dans Word, j'ai cliqué sur *Edition | Collage spécial | Image (Métafichier optimisé)*. Ce qui s'affiche dans Word est très comparable à ce qui s'affiche dans Concord.

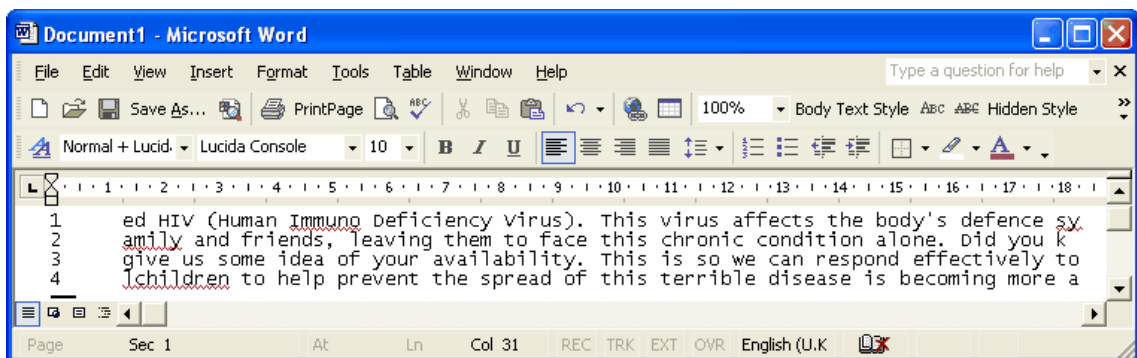
en tant que texte brut

Vous pouvez coller les données en tant que texte brut afin d'éditer les lignes de concordance, pour les utiliser en cours par exemple, ou encore pour les placer dans une feuille de calcul de MS [Excel](#)TM (ce qui est la solution la plus optimale si vous avez des données graphiques telles que des [Schémas de dispersion Concord](#) ou des [schémas KeyWords](#)).

Ici, la concordance ou les autres données sont copiées en tant que texte brut, avec une tabulation séparant chaque colonne. L'éditeur de texte Bloc-notes de Windows ne peut prendre en charge que ce format de données. Microsoft Word colle (avec Maj+Inser ou Ctrl+V) les données en tant que texte. Il colle le nombre de caractères que vous avez défini dans les [paramètres](#) pour l'enregistrement en tant que texte, la valeur par défaut étant de 80.



Les lignes de concordance sont copiées, mais il est évident qu'elle ne sont pas bien alignées et qu'il est difficile d'afficher le terme de recherche (*this*). Pour que le terme de recherche s'aligne correctement, il faut utiliser une police de caractères non proportionnelle telle que Courier ou Lucida Console, et le résultat s'affichera ainsi :



Remarque : un texte de 80 caractères, en police Lucida Console taille 10, avec des nombres situés à gauche aura une largeur dépassant les 18 cm. Pour éviter les retours à la ligne trop nombreux, il est préférable de définir le format de page sur paysage. Une autre solution est de réduire le nombre de caractères par ligne à 50 ou 60.

12.9 adresses de contact

Téléchargements

Vous pouvez obtenir une version plus récente sur [mon site Internet](#). Vous y trouverez également quelques éléments supplémentaires (programmes, listes de mots, etc.) en téléchargement libre ainsi que des liens vers des sources de corpus libres.

La version d'[Oxford University Press](#) la plus récente n'est pas à jour.

Captures d'écran

accédez à la [page des captures d'écran WordSmith 3](#) ou <http://www.lexically.net/wordsmith/version4/screenshots/index.html> pour afficher des captures d'écran sur les fonctionnalités de la suite Oxford WordSmith Tools. Ceci peut vous donner des idées pour vos propres recherches et vous permet également de mieux comprendre les limites de WordSmith.

Achat

Accédez à la page <http://www.lexically.net/wordsmith/purchasing.htm> pour obtenir des détails sur les fournisseurs.

Plaintes et suggestions

Si vous ne disposez pas de la version officielle de Oxford University Press (OUP) mais d'une version provenant de mon site Internet, ne contactez pas OUP mais envoyez-moi directement un courrier électronique (Mike.Scott@liv.ac.uk). Donnez une description aussi complète que possible du problème à régler, ainsi que des informations sur le matériel utilisé. Veuillez ne pas envoyer de pièces jointes dont la taille dépasse 200 Ko. J'essaie d'aider à résoudre les problèmes, mais malheureusement, je ne peux rien promettre...

12.10 format de date

Format de date

Le format de date japonais est année_mois_jour_heure_minute. Au moins, c'est un ordre logique, on part de l'entité la plus grande pour aller jusqu'à la plus petite. Pourquoi les URL ne sont-ils pas organisés de manière aussi logique ?

12.11 Définitions

12.11.1 définitions

mots

Le mot est défini comme étant une *séquence de caractères valides avec un [séparateur de mot](#) à chaque extrémité*. Les caractères valides comprennent toutes les lettres de A à Z, les caractères accentués qui peuvent être utilisés dans le [jeu de caractères](#) actuel, ainsi que des caractères définis par l'utilisateur comme pouvant être inclus au sein d'un mot (l'apostrophe ou le [trait d'union](#) par exemple).

Un mot peut avoir n'importe quelle longueur. Cependant, pour qu'un mot soit stocké dans une liste de mots, vous pouvez définir la longueur que vous souhaitez (50 caractères maximum). Tout mot dépassant cette longueur aura un + qui remplacera tous les caractères au-delà de la limite fixée. Vous pouvez décider d'inclure des mots comprenant des nombres (par exemple 35,50 €) dans les [caractéristiques de texte](#).

clusters

Un cluster est un *groupe de mots contigus*. Le terme *syntagme* n'est ici pas employé car son sens en linguistique implique qu'il y a une relation grammaticale entre les mots qu'elle contient. Dans le [traitement des clusters par WordList](#) ou [Concord](#), il ne peut y avoir aucune certitude quant à ce sujet, bien que les clusters correspondent souvent à des locutions ou à des syntagmes. Rubriques connexes : [general cluster information](#).

phrases

Une phrase est définie par *un point final, un point d'interrogation ou un point d'exclamation immédiatement suivi d'un ou plusieurs [séparateurs de mot](#) puis d'une majuscule dans la langue courante, d'un nombre ou d'un symbole monétaire*. Pour de plus amples informations à ce sujet, reportez-vous à la section [Début et fin des segments de texte](#) ou [informations techniques sur Viewer](#).

paragraphes

Les paragraphes sont définis par l'utilisateur. Reportez-vous à la section [Début et fin des segments de texte](#) pour de plus amples informations.

en-têtes

Les en-têtes sont également définis par l'utilisateur – reportez-vous à la section [Début et fin des segments de texte](#).

Rubriques connexes : [Définition des caractéristiques de texte](#), [Pertinence](#), [Mot-clé clé](#), [Associé](#)

12.11.2 séparateurs de mots

La convention dit qu'un mot est séparé des autres par la présence d'espaces à chacune de ses extrémités. Cependant, **Oxford WordSmith Tools** inclut également, au sein des séparateurs de mots, certains codes standard utilisés par la plupart des logiciels de traitement de texte : saut de page (12), tabulations (9), retour chariot (13) et retour à la ligne (10), fin de texte (26). De plus, les [trait d'unions](#) peuvent aussi être considérés comme séparant des mots tels que *auto-école* en deux mots.

Remarque : en chinois et en japonais, où les mots ne sont pas séparés ainsi, toute fonction de WordSmith exigeant une séparation des mots ne fonctionnera pas tant que vos textes n'aient pas été balisés avec des séparateurs de mots.

12.12 version de démonstration

La version de démonstration de **Oxford WordSmith Tools** offre *toutes* les fonctionnalités de la suite complète. La différence est que le nombre de lignes affichées et pouvant être imprimées (les mots d'une liste, les lignes de concordances, etc.) est limité. L'intégralité des données est enregistrée, mais vous ne pouvez pas tout afficher dans la version de démonstration.

Mise à jour

Pour mettre à jour votre version de démonstration, accédez à la page <http://www.lexically.net/wordsmith/purchasing.htm> pour obtenir les coordonnées de fournisseurs.

Une fois le code d'inscription reçu, exécutez **Oxford WordSmith Tools** comme à l'accoutumée. Une option de menu, nommée *Mettre à jour à partir de la version de démonstration*, s'affichera à l'écran. Cliquez sur cette option puis saisissez le nom d'inscription ainsi que les informations demandées telles que votre lieu de travail et le code d'inscription à 20 lettres. Le nom ainsi que les autres détails doivent comporter au moins huit lettres au total.

Copiez ou collez le nom *exact* d'inscription fourni lorsque vous avez acheté le code pour **Oxford WordSmith Tools**. Ce nom s'affiche dans la fenêtre principale et est toujours disponible à partir de l'option de menu *À propos* (F9). Votre logiciel sera alors totalement activé et l'option de *Mise à jour à partir de la version de démonstration* ne s'affichera plus. Le programme `updater.exe` reste cependant disponible dans le dossier `\wsmith4` et peut être réutilisé en cas de besoin.

Utilisation en réseau : cochez cette option si vous travaillez sur un réseau et que vous souhaitez rendre le logiciel accessible à d'autres ordinateurs connectés à ce réseau. Cette opération n'est disponible que si votre licence le permet (impossible avec une licence utilisateur unique).

Mettre une icône sur le Bureau : cochez cette option si vous souhaitez placer une icône pour WordSmith afin d'accéder facilement au logiciel.

Associer les fichiers WordSmith avec le programme : cochez cette option si vous souhaitez que votre ordinateur ouvre les fichiers créés avec Oxford WordSmith Tools avec le programme

approprié. Un fichier *.cnc doit être ouvert en utilisant Concord, un fichier *.lst avec WordList, etc. Des droits d'administrateur peuvent être nécessaires pour effectuer cette opération.

Si vous commettez une erreur et que votre inscription échoue, vous pouvez retenter l'opération. Si l'inscription réussit mais que vous décidez de modifier d'autres informations, exécutez **updater.exe**, qui se trouve dans le dossier **lwsmith4**.

Rubriques connexes : [Informations sur la version](#), [Adresses de contact](#).

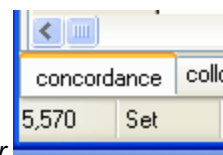
12.13 mode édition et mode saisie

La plupart des fenêtres permettent

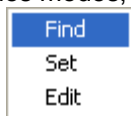
- d'éditer vos données (mode édition), ou
- d'accéder rapidement à un emplacement d'une liste (mode saisie).

Dans une fenêtre de concordance, les touches vous permettent également de paramétrer les [catégories](#) relatives aux données ou de [masquer](#) le terme de recherche.

En mode saisie, les touches du clavier vous permettent d'[accéder rapidement](#) à l'élément de liste qui vous intéresse ; vous pouvez par exemple taper `theocr` pour obtenir `théocracie` (ou vous en approcher) dans une liste de mots. Si vous avez saisi 5 lettres et qu'une correspondance est trouvée, la recherche s'arrête.



Pour basculer entre les modes, effectuez un clic droit sur *Définir* et



sélectionnez dans le menu qui apparaît.

Rubriques connexes : [catégories définies par l'utilisateur](#).

12.14 rechercher des textes source

Pour certains calculs, il est nécessaire que les textes sources originaux soient disponibles. Par exemple, pour que Concord puisse afficher [plus de contexte](#) qu'il n'avait sauvegardé pour chaque ligne, il devra relire le texte source. Pour que KeyWords puisse calculer le [schéma de dispersion](#), il doit se reporter au texte source pour déterminer comment se placent les mots-clés les uns par rapport aux autres et calculer les positions de chaque mot-clé dans le texte et les [liens](#) des mots-clés.

Cela n'est pas possible si vous avez déplacé ou effacé les fichiers source.

Rubriques connexes : [Modifications de noms de fichiers](#), [Sélection de fichiers source](#).

12.15 folders\directories

L'option est disponible à partir du menu Paramètres de chaque outil. Les dossiers par défaut peuvent être modifiés dans Oxford WordSmith Tools ou définis par [défaut](#) dans `wshe11.ini`.

- Dossier de concordance : pour vos fichiers de concordance.
- Dossier KeyWords : pour vos fichiers de listes de mots-clés.
- Dossier WordList : c'est là que vous [enregistrez](#) généralement vos fichiers de listes de mots.
- Dossier de textes : emplacement où sont stockés vos fichiers texte.
- Médias téléchargés : emplacement où vos [fichiers son et vidéo](#) sont stockés après avoir été téléchargés à partir d'Internet.
- Paramètres : emplacement où sont conservés vos fichiers de paramètres (fichiers *.ini ou autres).

Si vous saisissez le nom d'un dossier qui n'existe pas, Oxford WordSmith Tools le créera si cela est possible. Sur un réseau, cela dépend de vos droits de création et d'[enregistrement](#) de fichiers.

Si vous modifiez votre dossier Paramètres, vous devez laisser WordSmith copier tous les fichiers *.ini ainsi que tout autre fichier de paramètres ayant été créé afin de conserver les informations telles que les préférences de langues, entre autres.


Remarque : sur un réseau, les noms de lecteurs tels que **G:**, **H:**, **K:** changent en fonction de l'ordinateur que vous utilisez. Ce qui est désigné par **G:\texts\my text.txt** sur un terminal peut être **H:\texts\my text.txt** sur un autre terminal. Cependant, les lecteurs réseau ont également des noms structurés comme suit : [\\nom_ordinateur\nom_lecteur](#). **WordSmith** peut utiliser ces noms. Ils présentent également l'avantage de rendre les mêmes fichiers accessibles ultérieurement.

Astuce

Utilisez des dossiers différents pour les différentes fonctions de la suite Oxford WordSmith Tools. Vous pouvez par exemple être amené à créer de nombreuses listes de mots et de mots-clés si vous créez des [bases de données](#) de mots-clés. Il est théoriquement possible de placer n'importe quel nombre de fichiers dans un dossier, mais leur accès se trouve ralenti lorsqu'il y en a plus que 500 par dossier. La fonctionnalité "lot" vous permet de créer de nombreux fichiers de listes de mots ou de mots-clés. Il est conseillé d'utiliser un dossier `\keywords` pour stocker les fichiers *.kdb et `\keywords\type1`, `\keywords\type2`, etc. pour les fichiers *.kws pour chaque type.

Rubriques connexes : [rechercher des textes source](#).

12.16 HistoryList

Liste d'historique : dans WordSmith, de nombreuses zones combinées comme celle-ci, permettant de sélectionner un terme de recherche , se souviennent de vos saisies, de sorte que vous puissiez les faire défiler en cliquant sur la flèche.

12.17 HTML, SGML et XML

Ce sont des formats destinés à l'échange de textes. Le plus connu d'entre eux est le format HTML, Hypertext Markup Language, utilisé pour la diffusion de textes sur Internet. SGML signifie Standard Generalized Markup Language et est utilisé par les éditeurs et le BNC ; XML (Extensible Markup Language) est intermédiaire entre HTML et SGML.

Tous ces standards utilisent du [texte brut](#) avec des balises supplémentaires généralement mises entre crochets, comme <h1> et </h1>. On insère ces balises pour apporter des informations supplémentaires au texte :

- 1 un en-tête (<head>) donnant des détails sur l'auteur & l'édition
- 2 des informations sur l'affichage (par exemple <gras>, <italique>)
- 3 quelles sont les sections importantes (<h1> délimite un titre, <body> contient le corps du texte)
- 4 affichage des symboles particuliers (é correspond à é)

Rubriques connexes : [Présentation générale des balises](#)

12.18 traits d'union

Il s'agit du caractère utilisé pour séparer les mots. L'élément "auto-formation" peut être considéré comme étant 2 mots ou 1 mot, selon les paramètres des textes.

12.19 Internet

Disponible dans Modifier les paramètres | Général, dans le contrôleur WordSmith. Indiquez le nom complet de votre navigateur Internet ou cliquez sur parcourir pour le rechercher. WordSmith pourra ainsi accéder à l'aide au format .htm et aux captures d'écran incluses au logiciel.

12.20 limites

Les programmes de la suite **Oxford WordSmith Tools** peuvent traiter un nombre quasi-illimité de textes. Ils peuvent lire des textes provenant de CD-ROM, permettant ainsi de traiter des corpus de plusieurs millions de mots. En pratique, les limites sont a) la [mémoire](#) et b) la patience.

Vous pouvez avoir un nombre illimité de sessions ouvertes pour chaque outil. Chacune d'entre elles vous permet de traiter un ensemble de données.

Les [balises à ignorer](#) ou celles qui contiennent un astérisque peuvent contenir jusqu'à 1 000 caractères.

Seuls les 2 premiers Mo de texte sont examinés lorsque vous recherchez les balises, et ce afin de déterminer si vos [fichiers texte correspondent à vos critères](#). En [Ascii](#), cela représente 2 millions de caractères, en [Unicode](#), 1 million.

Astuce

Appuyez sur F9 pour afficher la fenêtre "À Propos" ; elle indique la date de la version ainsi que la [mémoire](#) disponible. Si la mémoire est insuffisante, a) fermez des applications, b) fermez WordSmith Tools et relancez-le.

Rubriques connexes : [Limites spécifiques à chaque outil](#)

12.21 limites spécifiques des outils

Limites dans Concord

Vous pouvez calculer un nombre illimité de lignes de concordance avec **Concord**.

Concord supporte un maximum de 80 caractères dans [un mot ou un syntagme de recherche](#), mais vous pouvez spécifier un nombre illimité de termes de recherche de concordance dans [fichier de termes de recherche](#).

Chaque concordance peut enregistrer un nombre illimité de co-occurents, avec un [voisinage](#) maximum de 25 mots de part et d'autre de votre mot de recherche.

Limites dans WordList

Une entrée principale peut comporter des milliers de [lemmes](#), mais vous ne pouvez pas regrouper plus de 20 éléments à la fois en utilisant F4. Répétez l'opération aussi souvent que nécessaire.

Une liste de [cohérence détaillée](#) peut prendre en charge jusqu'à 50 fichiers.

Limites dans KeyWords

Un schéma de mots-clés par écran de mots-clés. Pour en afficher davantage, appelez le même fichier dans une nouvelle fenêtre d'affichage.

nombre de fenêtres de [liens](#) par affichage de [schéma](#) de mots-clés : 20.

nombre de fenêtres d'[associés](#) par affichage de mot-clé clé : 20.

Limites dans Splitter

Le longueur maximale d'une ligne d'un gros fichier texte est de 10 000 caractères. Les "Entrée" permettent d'éviter ce genre de problèmes.

Limites dans Text Converter


Vous pouvez rechercher et remplacer jusqu'à 500 chaînes pour chaque texte.



Chaque chaîne de recherche et chaque chaîne de remplacement peut comporter un maximum de 80 caractères.

Un astérisque ne peut pas être le premier ou le dernier caractère d'une chaîne de recherche.

Lorsque l'astérisque est utilisé pour garder en mémoire des informations, la limite est fixée à 1 000 caractères.

Limites dans Text Viewer

Si vous activez l'option Affichage  lors de la sélection des textes, **Viewer** appelle les 10 premiers fichiers texte source sélectionnés.

Lorsque vous sélectionnez des textes ou allez directement au milieu d'un texte (après avoir sélectionné  in Concord), **Viewer** ne traite que 10 000 caractères de chaque fichier pour accélérer le processus dans le cas de gros fichiers. Vous pouvez cependant "relire" le fichier en cliquant sur  rafraîchir l'affichage et le texte sera alors lu en entier.


Vous ne pouvez travailler que sur un texte aligné à la fois.

Rubriques connexes : [Limites générales](#)

12.22 liens entre les outils

Couplage avec logiciels de traitement de texte, feuilles de calcul, navigateurs, etc.

Vous pouvez copier des informations à partir de toutes les fenêtres de listes de mots ou de textes dans le [presse-papiers](#). Pour ce faire, appuyez sur Ctrl+Inser.

 Lorsque ce symbole est présent, vous pouvez copier directement les données sélectionnées dans un nouveau document Microsoft Word™.

Les URL sont cliquables (comme <http://www.lexically.net>) et s'ouvrent dans votre navigateur.

Liens entre les différents outils

Les programmes de la suite **Oxford WordSmith Tools** sont liés par le fichier [wshell.exe](#) (celui dont la barre de titre affiche "Oxford WordSmith Tools [Controller](#)" et se trouve en haut à gauche de votre écran). Il gère les [paramètres par défaut](#) comme la couleur, les dossiers, les polices, les listes d'arrêt etc.

Si vous appuyez sur Ctrl+C dans **WordList** ou **KeyWords**, vous obtenez une concordance du mot actif à partir des fichiers utilisés.

Appuyez sur Ctrl+W dans **Concord** ou **KeyWords** permet de démarrer une liste de mots avec les fichiers actifs.

Tous les outils envoient des informations à l'outil qui a été appelé. Il s'agit du mot actif (mis en surbrillance dans la fenêtre de défilement) et des fichiers texte d'où provient l'information.

Exemple: vous créez une liste de mots à partir de 3 fichiers texte se rapportant au monde des affaires et vous constatez que le mot *espérant* est plus fréquent que ce à quoi vous vous attendiez. Vous pouvez effectuer une concordance de ce mot à partir des mêmes textes. Mettez *espérant* en surbrillance et appuyez sur Ctrl+C. Cela vous permet de voir si *espérant* fait partie d'un [cluster](#) de trois mots et d'afficher le schéma de dispersion.

Exemple: vous avez créé une [base de données](#) de mots-clés en utilisant 300 textes se rapportant au monde des affaires et vous constatez que le mot *offre* semble être un mot-clé clé et qu'il est associé à *entreprise*, *parts*, etc. Mettez *offre* en surbrillance, appuyez sur Ctrl+C pour effectuer une concordance en utilisant ces mêmes 300 fichiers. Ceci vous permet de vérifier le contexte du mot *offre* : s'agit-il d'une offre de rachat ou d'emploi ?

Exemple: vous avez effectué une concordance du mot *vert*. Appuyez sur Ctrl+W pour générer une liste de mot à partir de ces fichiers texte. Appuyez sur Ctrl+K pour comparer cette liste de mots avec un corpus de référence pour voir quels sont les mots-clés dans ces fichiers texte.

12.23 liste des options des menus et des boutons

Ces fonctions peuvent être visibles ou non dans chaque outil, en fonction de la capacité de ce dernier ou de la fenêtre actuelle de données (celle dont la barre de titre est active).



conseils

ouvre une fenêtre affichant un plan de la suite Oxford WordSmith Tools, donnant ainsi un aperçu de l'emplacement où vous vous trouvez et où vous pouvez aller. Cette fenêtre peut également afficher des astuces, mais cela dépend de l'outil.



associés

ouvre une nouvelle fenêtre affichant les [associés](#).



regrouper automatiquement

regroupe ([lemmatise](#)) automatiquement.



redimensionnement automatique

redimensionne chaque ligne de l'affichage de façon à ce que chacune affiche la quantité de données nécessaire. La plupart des fenêtres comportent des lignes de taille fixe, mais certaines vous permettent de modifier la hauteur de ligne (dans Viewer par exemple). Ceci permet d'ajuster la hauteur de ligne en fonction de la colonne de données en surbrillance.



groupes

calcule les [groupes](#) dans une base de données de mots-clés



rassembler des groupes

[rassemble](#) les groupes



clusters

calcule les [clusters](#) de concordance.



co-occurents

affiche les [co-occurents](#) en utilisant des données de concordance.



calculer

calcule une [nouvelle colonne de données](#) en se basant sur des fonctions de calcul et/ou des données existantes.

 **recalculer les co-occurents**

recalcule les co-occurents que vous avez par exemple supprimé des lignes de concordances.

totaux des colonnes

calcule l'écart total, minimum, maximum, moyen et standard pour chaque [colonne](#) de données numériques.

 **concord**

permet d'ouvrir Concord lorsque KeyWords ou WordList est ouvert et d'effectuer une concordance des mots mis en surbrillance en utilisant les textes source originaux.

 **copier**

permet de [copier](#) vos données vers différents emplacements (l'imprimante, [un fichier texte](#), le [presse-papiers](#), etc.).

 **doubler les colonnes**

vous permet de doubler le nombre de colonnes afin d'économiser du papier lors de l'impression.

 **éditer**

permet d'[éditer](#) une liste de recherche ([recherche par saisie](#)).

 **mode édition et mode saisie**

permet de basculer entre le mode édition et le mode saisie.


 **noms de fichiers**

ouvre une nouvelle fenêtre affichant les [noms de fichiers](#) d'où proviennent les données actuelles. Vous pouvez les [éditer](#) si vous le souhaitez.

 **rechercher des fichiers**

recherche les fichiers qui comportent tous les mots que vous avez marqués.

 **agrandir**

augmente la hauteur de toutes les lignes à une taille fixe. Cf. Rétrécir  ci-dessous :

 **aide (F1)**

ouvre l'aide de WordSmith (ce fichier) qui comporte une aide contextuelle.

 **regrouper**

regroupe deux entrées dans Viewer, deux mots dans WordList (cf. [lemmatisation](#)).

 **mise en page**

vous permet de modifier les paramètres de [mise en page](#) : la couleur de chaque colonne, si une colonne de données doit être masquée, les polices de caractères et les largeurs de colonnes.

 **liens**

calcule les [liens](#) entre les mots dans un schéma de mots-clés.

 **marque**

marque une entrée pour le [regroupement](#) ou pour la [recherche de fichiers](#).

 **faire correspondre les lemmes**

compare chaque élément de la liste avec ceux provenant d'un fichier texte comportant des formes lemmatisées et [regroupe](#) les éléments qui correspondent.

 **liste de correspondance**

compare les entrées de la liste active avec celles provenant d'un ["fichier de listes de correspondance" ou d'un modèle](#), marquant celles trouvées du signe (~).

 **informations mutuelles**

calcule les scores d'[informations mutuelles](#) dans une [liste d'index WordList](#).

 **nouveau**

[vous permet de démarrer](#) dans les divers outils : effectuer une concordance, créer une liste de mots ou de mots-clés.

ouvrir

vous permet de sélectionner un ensemble de résultats sélectionnés.

-  **schémas**
calcule les [schémas](#) des collocations.
-  **lire les fichiers médias**
permet de lire un [fichier média](#).
-  **schéma**
ouvre une nouvelle fenêtre affichant un [schéma de dispersion Concord](#) ou un [schéma KeyWords](#).
-  **imprimer (F3)**
affiche un aperçu de vos données avant impression. Il permet également d'imprimer dans un fichier, ce qui est équivalent à "[enregistrer au format texte](#)".
-  **actualiser**
actualise l'affichage de vos fichiers texte (dans Viewer) ou actualise l'écran (dans l'aperçu avant impression).
-  **remplacer**
permet par exemple de remplacer les données d'un lecteur ou d'un dossier lors de [l'édition des noms de fichiers](#) lorsque les textes source ont été déplacés.
-  **re-trier**
permet de re-trier les listes (par exemple, par ordre de fréquence et non par ordre alphabétique) dans [Concord](#), [KeyWords](#) ou [WordList](#).
-  **règle**
affiche/masque les divisions verticales des listes ainsi que les divisions de texte dans les [schémas KeyWords](#). Cliquez sur "règle" dans un menu pour l'afficher ou la masquer ou encore pour modifier le nombre de divisions d'un [schéma](#).
-  **enregistrer (F2)**
[enregistrez vos données](#) en utilisant le nom de fichier existant. S'il s'agit d'un nouveau fichier, vous devrez saisir un nom.
-  **enregistrer sous**
enregistre après saisie du nom de fichier.
-  **enregistrer au format texte**
enregistre le fichier au format *.txt : texte brut Il ne s'agit pas du même processus consistant à enregistrer vos données actuelles pour y revenir ultérieurement dans Oxford WordSmith Tools.
-  **rechercher (F12)**
[recherche](#) au sein d'une liste.
-  **rétrécir**
réduit la hauteur de toutes les lignes à une taille inférieure fixe. Cf. Agrandir  ci-dessous :
-  **parcourir**
permet de parcourir un texte dans Viewer.
-  **statistiques**
ouvre une nouvelle fenêtre affichant les [statistiques détaillées](#).
- barre d'état**
affiche ou masque la barre d'état (dans la partie inférieure de la fenêtre - affiche les commentaires et le statut de la tâche en cours).
-  **statistiques de résumé**
ouvre une nouvelle fenêtre affichant les [statistiques de résumé](#), par exemple la proportion de lemmes par rapport aux types de mots.
-  **échanger les colonnes et les lignes**
permet d'échanger les colonnes et les lignes. Les [statistiques](#) de WordList s'affichent par défaut avec les données du fichier dans chaque colonne. Cliquez sur ce bouton pour échanger les données de lignes avec les données de colonnes.
- barre d'outils**

masque ou affiche une barre d'outils comportant les mêmes boutons que ceux que vous avez sélectionnés lorsque vous avez [personnalisé les menus contextuels](#).

dissocier

permet de dissocier des entrées regroupées, par exemple des entrées [lemmatisées](#).

affiche le texte source

[affiche le texte source](#) et met en surbrillance les mots actuellement sélectionnés dans la liste.

Microsoft Word™

envoie des données formatées vers Word.

liste de mots

[crée une liste de mots](#) dans KeyWords en utilisant les données actuelles.

ignorer

[ignore](#) les entrées supprimées.

Rubriques connexes : [Raccourcis clavier](#), [Personnalisation des menus contextuels](#).

12.24 raccourcis clavier

fenêtres de défilement :

Ctrl+Orig vers le haut de la liste

Ctrl+Fin vers la dernière ligne de la liste

si elle est classée par ordre alphabétique, [saisissez votre mot de recherche](#)

si elle défile horizontalement :

Orig vers la gauche

Fin vers la droite

Ctrl+Droite aller au mot à droite

Ctrl+Gauche aller au mot à gauche


raccourcis clavier :

Ctrl+C [appeler Concord](#) à partir d'un autre outil


Ctrl+W [appeler WordList](#) à partir d'un autre outil


Ctrl+Inser copier le bloc sélectionné vers le [presse-papiers](#)


Touches MAJ+clic sélectionner un bloc


F1 aide 

F2 [enregistrer les résultats](#) 


F3 imprimer les résultats 

F4 regrouper des entrées 

F5 marquer des entrées pour les grouper 

F6 re-trier 

Ctrl+F6 inverser le [tri](#) des mots 

F7 afficher le texte source 

F8 rechercher de courtes phrases (**Viewer**)

F9 Rubrique à propos (affiche la date de la version et la mémoire disponible)

F12 effectuer une recherche au sein d'une liste



Ctrl+M [Fusionner](#) 2 listes de mots ou bases de données KeyWords

Alt+H accéder aux sous-menus de l'**Aide**

Alt+W accéder aux sous-menus des **Paramètres**

Alt+X accéder aux sous-menus de la **Fenêtre**

Alt+X quitter l'outil
Ctrl+Z [Ignorer](#) les lignes supprimées

Rubriques connexes : [Éléments de menus et boutons](#)

12.25 noms de fichiers longs

Cette version de **WordSmith** prend correctement en charge les noms de fichiers longs.

12.26 configuration minimale requise

Cette version de **Oxford WordSmith Tools** est conçue pour les systèmes comportant :

- au moins 256 Mo de RAM (128 peuvent suffire, mais pas avec Windows XP)
- au moins 40 Mo d'espace disque
- Windows 98, NT, 2000, XP ou une version plus récente, ou un émulateur de l'un d'entre eux si vous travaillez sur un système Apple Mac ou Unix. La suite fonctionnera peut-être sous Windows 95 seconde édition, à voir...

Vous constaterez que WordSmith Tools fonctionne mieux sur une machine [plus rapide](#), surtout si elle dispose d'une quantité importante de mémoire [RAM](#).

12.27 manuel de WordSmith Tools

Ce fichier d'aide existe sous forme de manuel que vous obtenez lorsque vous procédez à l'[installation](#). Le fichier (**wshell.pdf**) est au format Adobe Acrobat™. Il comporte une table des matières et un index relativement détaillé, pour la création duquel j'ai utilisé **WordList** et **KeyWords**. La plupart des gens préfèrent la version papier aux fichiers d'aide.

Vous jugerez peut-être utile de voir des captures d'écran de **WordSmith** en cours d'utilisation : reportez vous aux [Adresses de contact](#).

12.28 nombres

En fonction des paramètres de texte, vous pouvez décider d'inclure ou d'exclure les nombres des listes de mots.

12.29 valeur de dispersion du schéma

Le but

La valeur de dispersion est le degré d'uniformité de la répartition d'un ensemble de valeurs. On peut comparer cela à la pluie. Au Royaume-Uni par exemple, les moments pluvieux se répartissent de façon relativement uniforme au cours de l'année. On peut maintenant comparer avec les pays qui ont une saison des pluies.

En termes linguistiques, on peut souhaiter connaître la distribution des occurrences d'un mot tel que *skull* dans la pièce Hamlet. WordSmith permet, depuis la version 1, d'afficher ces résultats sous forme de schéma. Les statistiques sur la valeur de dispersion offrent une prise en charge mathématique et permettent des comparaisons plus aisées.

Calcul

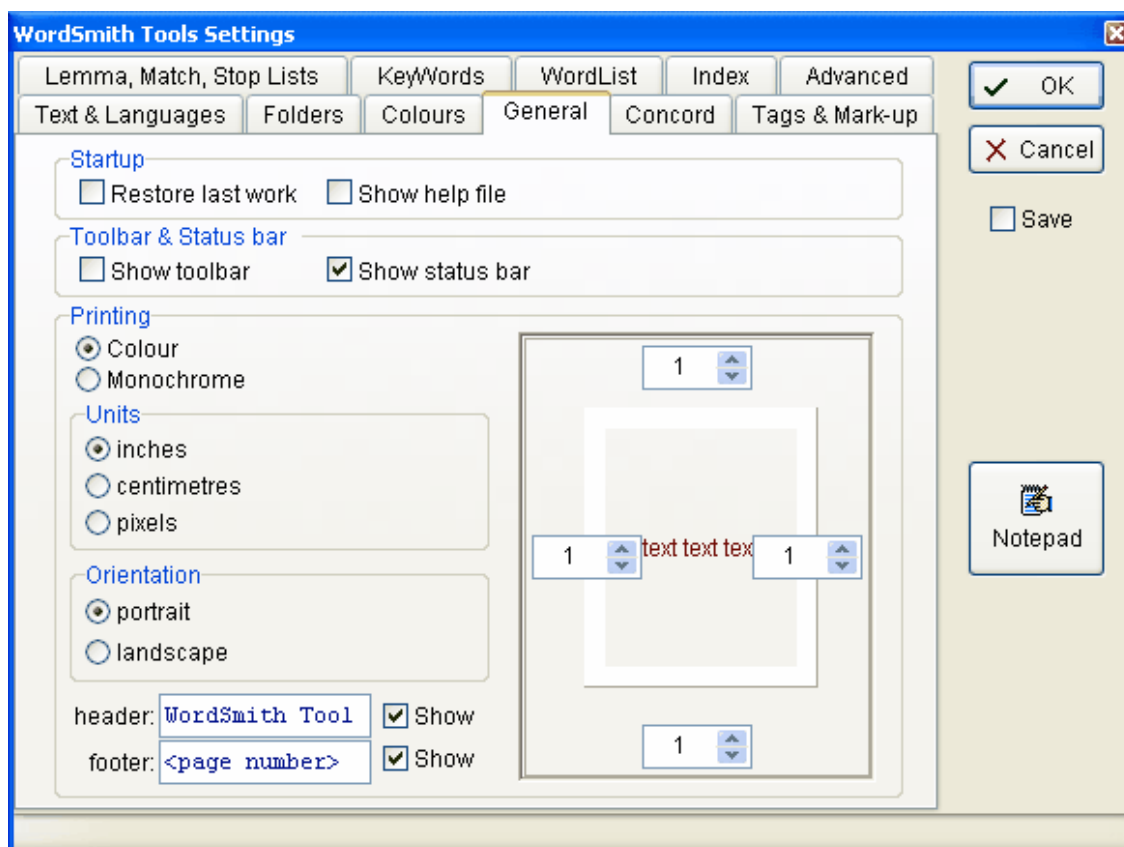
Le schéma de dispersion calculé dans KeyWords et Concord utilise la première des trois

formules énoncées par [Oakes](#) (1998 : 190-191), qu'il désigne comme étant la plus fiable. Tout comme avec la [règle](#), le schéma est divisé en 8 segments. Les valeurs vont de 0 à 1, avec 0,9 ou 1 désignant une dispersion uniforme, et 0 ou 0,1 une dispersion irrégulière ([Katz](#), 1996)

Rubriques connexes : [schéma KeyWords](#), [schéma de dispersion Concord](#).

12.30 paramètres d'impression

Cliquez sur *Modifier les paramètres | Général* dans le [contrôleur](#) de la suite Oxford WordSmith Tools pour y accéder.





Démarrage

Restaurer votre travail vous permet de retrouver la dernière liste de mots, concordance ou liste de mots-clés lorsque vous démarrez WordSmith.

Barre d'outils et barre d'état

Chaque outil comporte une barre d'état (en bas) et une barre d'outil (en haut). Par défaut, la barre d'outils est masquée, pour éviter l'encombrement de l'écran.

Impression

Si vous réglez vos paramètres d'impression sur monochrome, votre imprimante mettra en italique ou en gras toutes les colonnes qui sont d'une [couleur](#) différente de la couleur actuelle du "texte brut". Une imprimante couleur imprimera les couleurs ou utilisera des niveaux de gris le cas échéant. Vous pouvez modifier les unités, l'orientation (portrait  ou paysage ) , les marges ainsi que les en-têtes et pieds-de-page par défaut. Vous pouvez utiliser la fonction

[Aperçu avant impression.](#)

Confirmation

Vous avez la possibilité de paramétrer Oxford WordSmith Tools dans le [fichier des paramètres par défaut \(wshell.ini\)](#). Si le fichier comporte la ligne "confirm printing=YES", vous serez informé (e) pour chaque ligne de la concordance actuelle ou de la liste des lignes imprimées.,

Rubriques connexes : [Impression](#)

12.31 mémoire RAM disponible

Plus votre ordinateur a de mémoire RAM, plus il est rapide et plus sa capacité de stockage est importante. Chaque programme doit pouvoir stocker des résultats lors de l'utilisation. Une liste de mots de plus de 80 000 entrées, qui correspond à plus de 4 millions de mots de texte, occupera environ 3 Mo de mémoire. Une liste en Finnois occuperait une place bien plus importante encore. Lorsque la mémoire est insuffisante, Windows tente de libérer de l'espace en mettant quelques résultats en mémoire temporaire sur votre disque dur. Si cela se produit, vous entendrez probablement de nombreux cliquetis lorsqu'il place des données sur le disque et les lit à nouveau. Dans tous les cas, vous entendrez probablement *quelques* cliquetis puisque la plupart des programmes d'**Oxford WordSmith Tools** accèdent à vos textes originaux depuis le disque dur, mais si les bruits émis sont *constants*, cela veut dire que la machine a atteint ses limites.

Vous pouvez vérifier l'espace disque disponible même en plein milieu d'un processus, en appuyant sur la touche F9 (option À Propos dans le menu principal d'*Aide* de chaque programme). La première ligne indique la mémoire RAM disponible. Les autres chiffres concernent les ressources système de Windows : elles ne devraient pas poser de problème, mais si elles sont inférieures à 20%, vous devez [enregistrer les résultats](#), quitter Windows et redémarrer.

Théoriquement, les listes de mots et les mots-clés peuvent contenir jusqu'à 2 147 483 647 entrée différentes. Chaque mot peut apparaître dans vos textes jusqu'à 2 147 483 647 fois. (Ce nombre étrange, 2 147 483 647, qui correspond à la moitié de 2 exposant 32, est le plus grand nombre entier relatif pouvant être stocké en 32 bits, on l'appelle également 2 Go.) Il est peu probable que vous atteigniez cette limite théorique : pour que l'élément *the* apparaisse 2 147 483 647 fois dans vos textes, il faudrait que vous ayez traité 30 000 millions de mots (un CD-ROM ne contenant que du texte brut peut contenir environ 100 millions de mots, ce nombre correspond donc à quelques 300 CD-ROM.) Vous manquerez de RAM bien avant d'avoir traité tant de mots.

Si vous disposez de 64 Mo de RAM ou plus, vous devriez être en mesure de garder en mémoire à la fois une copie d'une liste de mots basée sur des millions de mots, un logiciel de traitement de texte puissant ainsi qu'un fichier texte.

Rubriques connexes : [vitesse](#)

12.32 corpus de référence

Corpus de référence

Corpus de textes utilisé à titre comparatif. Vous pouvez par exemple comparer un texte avec le [British National Corpus](#), qui comporte quelques 100 millions de mots. Ceci peut s'avérer utile pour calculer les [mots-clés](#).

Dans le [contrôleur](#), vous pouvez [définir la liste de mots de votre corpus de référence](#) que KeyWords et Concord doivent utiliser.

12.33 restaurer le dernier fichier

Par défaut, la dernière liste de mots, concordance ou liste de mots-clés enregistrée ou extraite sera automatiquement restaurée au démarrage de **Oxford WordSmith Tools**. Si le dernier outil utilisé est **Concord**, la liste des dix termes de recherche les plus récents sera sauvegardée. Il est possible de désactiver cette fonctionnalité temporairement à partir d'une option de menu **Wshell** ou de la désactiver définitivement dans le fichier `wshell.ini` (qui se trouve dans le dossier `\wsmith4`).

12.34 mots uniques/clusters

Le but

Les clusters sont des mots qui apparaissent souvent ensemble l'un à côté de l'autre. Les clusters sont plus étroitement liés que les co-occurents, comme les unités multi-mots, les groupes ou les syntagmes. On les appelle *clusters* car les groupes et les syntagmes ont déjà leur utilisation en grammaire et le fait que le logiciel les trouve ensemble ne signifie pas forcément que ce soient de véritables *unités* multi-mots. [Biber](#) les appelle des "ensembles lexicaux".

Une langue est faite de mots et de locutions. Il ne faut pas la voir comme une sélection de mots destinée à remplir un espace grammatical, comme l'indiquent les théories structurelles. Les mots sont liés : les exemples les plus flagrants sont les locutions, dans lesquelles les mots sont intrinsèquement liés les uns aux autres, cependant, tous les mots ont tendance à former des clusters entre eux. Ces relations de clusters peuvent inclure des colligations (par exemple les relations entre **compter** et **sur**), des [collocations](#), et de la prosodie sémantique (**provoquer** a tendance à être lié à des effets négatifs comme **accident**, **problème**, etc.).

Oxford WordSmith Tools offre deux possibilités pour identifier les clusters de mots : [WordList](#) et [Concord](#). Les méthodes sont différentes. Concord ne traite que les lignes de concordance, alors que WordList traite des textes entiers.

Fonctionnement de Concord

Supposons que votre texte commence par :

*Once upon a time, there was a beautiful princess (Il était une fois une ravissante princesse).
She snored (Elle ronflait). But the prince didn't (Mais pas le prince).*

Si vous avez sélectionné un cluster de 2 mots, le texte sera segmenté comme suit :

Once upon

upon a

a time

(remarque : **pas** "time there" n'apparaît pas, car il y a une virgule entre les deux mots)

there was (etc)

Si vous paramétrez des clusters de trois mots, vous obtiendrez :

Once upon a

upon a time

there was a

was a beautiful

a beautiful princess

But the prince

the prince didn't

etc.

C'est-à-dire que chaque cluster de n mots sera enregistré s'il a une longueur minimum de n mots *jusqu'à un symbole de ponctuation* comme ;,!,? Logiquement, un cluster ne se trouve pas à cheval sur deux phrases séparées par une marque de ponctuation.

12.35 vitesse

Pour créer une liste de mots de 4,2 millions de mots, il fallait environ 20 minutes avec un ordinateur (486/33) de 1993 de 8 Mo de [RAM](#). Le tri à la fin de la procédure prenait une trentaine de secondes. Un Pentium de 200 MHz. avec 64 Mo de RAM peut traiter plus d'1,7 millions de mots par minute. Sur un Pentium de 100 MHz avec 32 Mo de RAM, le processus complet a pris 3 minutes 30, traitant plus d'un million de mots par minute.

Des tests de concordance ont été effectués avec un fichier de 55 Mo comportant 9,3 millions de mots sur le même Pentium 100 avec une lecteur CD-ROM vitesse 4x. En voici les résultats :

terme de recherche	source	vitesse
rapide	CD-ROM	6 millions de mots par minute (mpm)
rapide	disque dur	12 000 000 mpm
le	CD-ROM	900 000 mpm
le	disque dur	1 000 000 mpm
lez	CD-ROM	6 000 000 mpm
lez	disque dur	16 000 000 mpm

Avec un ensemble de textes de 4 à 20 ko et *rapide* comme terme de recherche, la vitesse était de 2 millions de mots par minute, pouvant aller jusqu'à 4 millions pour les fichiers plus longs. La création d'une liste de mots à partir du même ensemble de textes a atteint une vitesse de 800 000 mpm. Pour le fichier de 55 Mo, la vitesse s'élevait à environ 1 350 000 mpm.

Ces données montrent que les facteurs qui ralentissent la concordance sont, dans l'ordre, la rareté des mots (*le* est beaucoup plus rapide que *rapide* ou *lez*, qui n'existait pas), la taille du fichier texte (le traitement des très petits fichiers de 500 mots, soit environ 3 ko, est trois fois plus lent que dans le cas de gros fichiers) et la vitesse du disque. Lorsque Concord trouve un mot, il enregistre la ligne de concordance et les co-occurents et les affiche, pour que vous puissiez [arrêter](#) le traitement si les résultats ne vous conviennent pas ou que vous en avez déjà suffisamment. Ceci est un facteur majeur de ralentissement. D'autre part, lire un fichier fait appel au système de gestion des dossiers de votre ordinateur, qui met du temps à charger, si on le compare à Concord qui les parcourt. De plus, les vitesses des disques sont très différentes, les disquettes étant les plus lentes.

Si le traitement est trop lent, fermez des programmes et relancez **Oxford WordSmith Tools**.

Sinon, installez davantage de RAM. Vous pouvez obtenir de l'aide pour configurer correctement Windows (mémoire vive, cache de disque, etc.). Utilisez de préférence un disque dur performant.

Vous pouvez utiliser d'autres logiciels pendant que les programmes effectuent une tâche, mais cela aura pour effet de ralentir le processeur. Les jeux vidéo prennent trop de ressources, mais vous pouvez par exemple [imprimer](#) un document pendant que vous travaillez.

12.36 barre d'état

La barre située au bas d'une fenêtre permet de l'agrandir ou de la rétrécir et affiche également une série de panneaux avec des informations sur les données actuelles. La barre d'état peut généralement être affichée ou masquée via une option du menu principal. Un clic droit sur le panneau affiche un menu contextuel qui permet de choisir entre [Éditer](#), [Saisir et Enregistrer](#).

12.37 outils pour le repérage de schémas

Les outils informatiques sont utiles car ils facilitent certaines opérations et permettent d'effectuer des tâches plus compliquées. Cela permet d'avoir une vue d'ensemble car vous pouvez rapidement mettre en pratique une idée et faire des essais. Réorganiser un ensemble de données sous une nouvelle forme permet de repérer les schémas.

Mais c'est illusoire. L'ordinateur n'est pas un outil idéal de repérage de schémas. Il permet bien

d'additionner et de trier par exemple. L'ordinateur a une mémoire mais ne sait ni ne connaît rien et reconnaître des caractères imprimés n'est pas chose facile.

Cependant, l'ordinateur *nous* aide à repérer les schémas et les tendances. C'est pourquoi il ne faut pas surestimer des outils comme **Oxford WordSmith Tools**. Un outil reste une aide, et n'effectue pas le travail à votre place.

Outil et Produit

Certains logiciels sont conçus comme un produit. Un jeu ou un dictionnaire électronique ne dépend de rien. Un traitement de texte, une feuille de calcul ou une base de données sont des outils, car vous pouvez l'utiliser pour des actions que le développeur n'avait pas prévues. **Oxford WordSmith Tools**, comme son nom l'indique, n'est pas un produit mais une suite d'outils. Vous pouvez les utiliser pour analyser de nombreux types des schémas dans de nombreux textes dans différentes [langues](#).

Transformation

L'homme repère des schémas en modifiant la forme des données, en les réduisant et en les réorganisant. L'ordinateur ne se "rend pas compte" de ce que vous faites (si vous indiquez 2 à la calculatrice et doublez ce chiffre à répétition, l'ordinateur ne le fait pas automatiquement pour vous). L'être humain repère facilement des schémas visuels.

Si vous transformez un texte en liste ou schématisez les mots-clés en fonction de l'emplacement d'apparition dans le texte source, vous y verrez probablement un schéma. Le fait de voir un schéma est toujours subjectif.

Oxford WordSmith Tools permet de faciliter le repérage des schémas. Le but des outils de cette suite est donc de vous permettre d'avoir un bon aperçu des données de vos textes.

Types d'outils

Tous les outils agissent à deux niveaux : spécialisé et permanent.

utilisation générale/spécialisée

utilisation générale

Un logiciel de traitement de texte est d'utilisation générale, car il permet d'effectuer différentes tâches.

spécialisé

Le correcteur orthographique d'un logiciel de traitement de texte est spécialisé, car il ne peut exécuter d'autre tâche.


temporaire/constant

temporaire

Un outil employé une fois pour une utilisation donnée est un outil temporaire, car il est normalement destiné à une autre utilisation.

constant

Les outils de la suite **Oxford WordSmith Tools** sont des outils temporaires devenus

permanents. Ce sont des outils à utilisation générale. Ils représentent le couteau suisse 

Rubriques connexes : [Remerciements](#)

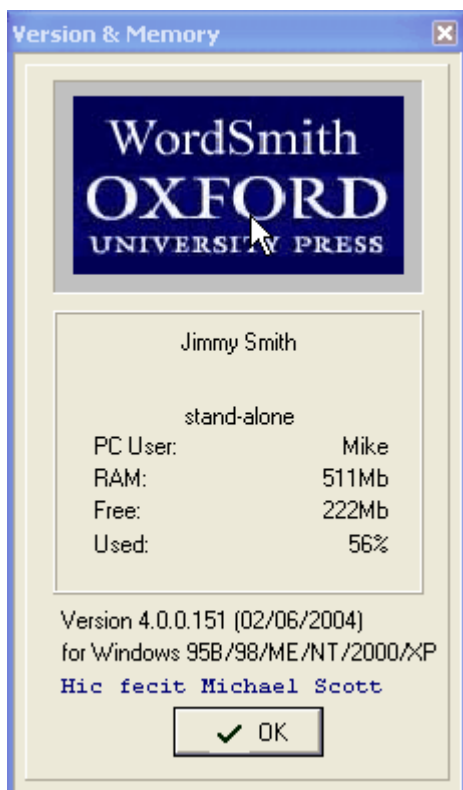
12.38 informations sur la version

Ce fichier d'aide concerne la version actuelle de **Oxford WordSmith Tools**.

La version de la suite **Oxford WordSmith Tools** est affichée en cliquant sur l'option de menu *À propos* (F9) qui indique également le nom sous lequel vous êtes enregistré ainsi que la [mémoire](#) disponible. Si vous êtes en version démonstration, cela s'affichera directement en dessous de votre nom.

La date dans cette fenêtre vous indique si votre version actuelle est à jour. Vous pouvez télécharger les versions plus récentes. Conservez votre code d'inscription pour les versions mises à jour.

Vous pouvez cliquer sur l'image WordSmith Oxford dans la fenêtre À Propos pour voir votre code.



Rubriques connexes : [Différences de la version 32 bits](#), [Version de démonstration](#), [Adresses de contact](#).

12.39 fichiers zip

Les fichiers zip sont des fichiers qui ont été compressés de manière standard. **WordSmith** est maintenant capable de lire et d'écrire des fichiers *.zip*.

Le but

Vos fichiers sont non seulement nettement plus petits que les fichiers de départ, mais vous gaspillez également moins d'espace disque en procédant ainsi. Tout fichier texte, même s'il ne contient que le mot "bonjour", occupera de 4 000 à 32 000 octets sur votre disque en fonction de votre système. Si vous avez 100 fichiers courts, vous perdez plusieurs milliers d'octets d'espace. Si vous "zippez" 100 fichiers courts, ils occuperont l'espace d'un seul d'entre eux. C'est en raison de ces avantages que l'on utilise souvent des fichiers zip pour des transferts sur Internet. Si vous voulez stocker de nombreuses listes de mots, il sera nettement plus efficace de les stocker dans un fichier zip.

Le "prix à payer" pour zipper est a) que cela nécessite un peu de temps (bien que très réduit), b) il faut un logiciel capable d'interpréter le format standard pour lire le fichier zip. On trouve de nombreux programmes de zip sur le marché, y compris *PKZip™* et *Winzip™*. Si vous zippez une liste de mots, ces programmes peuvent les dézipper mais ne pourront pas traiter la liste qui en résulte. **WordSmith** peut la dézipper et l'afficher.

Procédure

Vous pouvez cocher l'option "créer un fichier zip" lorsqu'on vous le propose. Les résultats sont

ainsi stockés à l'endroit indiqué avec l'extension *.zip*.

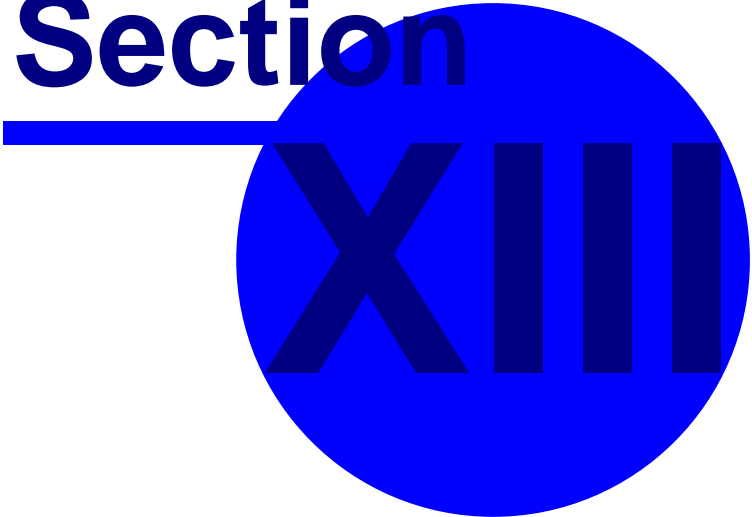
Lorsque vous ouvrez une liste de mots, une concordance, un fichier texte, etc. au format zippé comportant plusieurs fichiers, vous pouvez choisir lesquels ouvrir. Sinon, le processus aura lieu en arrière-plan et n'affectera pas votre travail normal avec **WordSmith**.

WordSmith Tools

Résolution des problèmes

Section

XIII



13 Résolution des problèmes

13.1 liste des FAQ

Rubriques connexes : [connexion](#).

Voici les questions fréquemment posées.

Vous trouverez une liste d'explications plus détaillées dans la section [Messages d'erreur](#).

[Impossible de traiter les apostrophes](#)

[Est-ce du russe, du grec ou de l'anglais ? Des symboles étranges s'affichent](#)

[C'est en panne](#)

[Le programme ne démarre pas.](#)

[Le processus est lent](#)

[Les touches ne répondent pas](#)

[La ligne se trouve au-delà de la limite autorisée par la version de démonstration](#)

[Non correspondance entre les résultats de Concord et de WordList](#)

[Aucune balise visible dans la concordance](#)

[Problème d'impression](#)

[Texte illisible à cause des couleurs](#)

[Espace entre les colonnes trop petit ou trop grand](#)

[La liste de mots ne fonctionne pas](#)

[Refuse de faire le café](#)

13.2 apostrophes introuvables

Apostrophes non traitées

Si vos textes originaux ont été enregistrés avec Microsoft Word™, il se peut que **Concord** ne puisse pas trouver les apostrophes ou les guillemets qu'ils contiennent. Ceci est dû au fait que Word génère des symboles spécifiques. L'apostrophe ordinaire (droite) ou les guillemets sont remplacés par une apostrophe arrondie (arrondie vers la droite ou la gauche en fonction de sa position). Ces symboles spécifiques ne sont pas les mêmes que les apostrophes droites ou les guillemets.

Solution : faites glisser le symbole de l'ensemble situé en dessous lorsque vous saisissez votre [terme de recherche](#), sinon, remplacez les termes posant problème avec Text Converter.

Rubriques connexes : [paramètres](#)

13.3 espace entre les colonnes

l'espace entre les colonnes n'est pas bon

Vous pouvez modifier ceci en cliquant sur le bouton [mise en page](#).

13.4 problème de balises Concord

aucune balise visible dans la concordance

Si vous ne pouvez voir aucune balise après avoir demandé la *Balise la plus proche* dans **Concord**, le problème vient sûrement du fait que les [Balises à ignorer](#) ont le même format. Par exemple, si *Texte à ignorer* comporte <*>, toutes les balises telles que <title>, <quote>, etc. seront supprimées de la concordance sauf si vous les spécifiez dans un [fichier de balises](#).

Solution : modifiez le fichier de balises et effectuez à nouveau la concordance.

13.5 Incohérence entre Concord et WordList

Incohérence entre Concord et WordList

Si **WordList** trouve un certain nombre d'occurrences d'un [cluster](#) (liste de mots) mais que **Concord** trouve un nombre différent, c'est parce que les procédures sont différentes. WordList procède mot par mot, en ignorant la ponctuation (sauf pour les traits d'unions et les apostrophes). Lorsque **Concord** recherche un [cluster \(concordance\)](#), il tient compte de la ponctuation.

13.6 en panne

C'est en panne !

Solution : quittez **Oxford WordSmith Tools** et redémarrez le programme. Si cela ne fonctionne pas, quittez Windows et ré-essayez.

13.7 limite de la version de démonstration

limite de la version de démonstration atteinte

Vous venez de télécharger le logiciel mais n'avez pas encore renseigné les champs nécessaires à l'enregistrement. Accédez à la fenêtre principale d'Oxford WordSmith Tools, puis cliquez sur *Paramètres | Enregistrement*.

Si vous ne disposez pas du code d'enregistrement à 20 caractères, contactez [Oxford University Press](#). La **seule** différence entre une [version de démonstration](#) et une version complète est la suivante : la dernière vous permet d'afficher ou d'imprimer toutes les données, alors que la première ne vous permet d'afficher que 25 lignes de résultats.

13.8 symboles inhabituels

symboles étranges

symboles étranges lors de l'utilisation de la suite Oxford WordSmith Tools

1. Vérifiez vos fichiers texte. Ouvrez-les dans **Bloc-notes**. Comportent-ils de nombreux symboles étranges ? Ces derniers peuvent être des codes cachés utilisés par votre logiciel de traitement de texte habituel. Solution : ouvrez-les dans votre logiciel de traitement de texte habituel et cliquez sur *Enregistrer sous*, donnez un nouveau nom et sélectionnez le format [texte brut](#), parfois également appelé "Texte seulement" ou ***.txt**.
2. *Choisissez les textes*, mettez le fichier texte en surbrillance, et avant de cliquer sur *OK*, cliquez sur *Affichage*. Y a-t-il des symboles étranges ? Solution : modifiez les *paramètres de texte*, puis essayez de basculer des jeux de caractères DOS vers des jeux Windows ou vice-versa. Le texte était en [ASCII](#) simple mais **Oxford WordSmith Tools** l'a pris pour du Windows ANSI.
3. Les symboles étranges que l'on peut trouver dans une liste de mots peuvent aussi provenir de fautes de frappe ou d'orthographe dans les fichiers texte originaux.

Grec, Russe, etc.

4. Si le texte est par exemple en russe ou en grec, une police adéquate est nécessaire. Vous pouvez en obtenir une via Microsoft Plus (pour les ordinateurs fonctionnant avec Windows 95 ou supérieur).
5. Si plusieurs listes utilisant *différents* jeux de caractères sont ouvertes et que vous souhaitez modifier la [police](#) ou [les caractéristiques de texte](#), les listes sont toutes mises à jour pour afficher la police et le jeu de caractères utilisés, sauf si vous réduisez toute fenêtre pouvant être affectée par ces modifications.

symboles étrange lors de la lecture de données WordSmith dans une autre application

Cliquez sur [Enregistrer](#) ou Enregistrer sous puis [Enregistrer au format texte](#)". L'élément sera [imprimé](#) dans un fichier. "Enregistrer" et "Enregistrer sous" permettent de stocker un fichier dans un format spécifique pour une réutilisation ultérieure par **Oxford WordSmith Tools**. Ce format ne permet pas de lire le fichier dans un logiciel de traitement de texte. Vous pouvez ainsi stocker votre travail et y revenir ultérieurement.

"Enregistrer au format texte", en revanche, revient à enregistrer en texte brut en "imprimant" dans un fichier. Cette fonction est utile si vous ne souhaitez pas imprimer le fichier sur du papier mais utiliser les données dans une feuille de calcul ou un logiciel de traitement de texte tel que **Microsoft Word**. Il est généralement plus rapide de copier le texte sélectionné vers le [presse-papiers](#).

13.9 couleurs illisibles

texte illisible à cause des couleurs

Solution : dans *Paramètres*, sélectionnez *Couleurs*. Vous pouvez à présent définir les couleurs qui correspondent le mieux à votre moniteur. Les paramètres "Monochrome" sont disponibles.

13.10 les touches ne répondent pas

Les touches ne répondent pas

Si rien ne se passe lorsque vous appuyez sur une touche, c'est sûrement parce que c'est la mauvaise fenêtre qui est active. Comme vous le savez, Windows est conçu pour permettre aux utilisateurs d'ouvrir un certain nombre de programmes dans le même écran. Ainsi chaque fenêtre répond à des combinaisons de touches différentes. Vous pouvez voir quelle fenêtre est active car sa barre de titre est d'une couleur différente de celle des autres. Il suffit de cliquer n'importe où dans la fenêtre que vous souhaitez utiliser, puis d'appuyer sur la touche que vous souhaitez.

13.11 pineapple-slicing

won't slice a pineapple

"Propose to any Englishman any principle, or any instrument, however admirable, and you will observe that the whole effort of the English mind is directed to find a difficulty, a defect, or an impossibility in it. If you speak to him of a machine for peeling a potato, he will pronounce it impossible: if you peel a potato with it before his eyes, he will declare it useless, because it will not slice a pineapple." Charles Babbage, 1852.

(Babbage was the father of computing, a 19th Century inventor who designed a mechanical computer, a mass of brass levers and cog-wheels. But in order to make it, he needed much greater accuracy than existing technology provided, and had all sorts of problems, technical and financial. He solved most of the former but not the latter, and died before he was able to see his Difference Engine working. The proof that his design was correct was shown later, when working versions were made. The difficulties he encountered in getting support from his government weren't exclusively English.)

13.12 l'imprimante n'a pas imprimé

problème d'impression

Si le document imprimé ressort de telle façon qu'une ou plusieurs colonnes sont blanches et les autres sont imprimées correctement, c'est que l'imprimante ne prend en charge que le noir et le blanc et pas les dégradés de gris. Dans le [contrôleur](#), modifiez le paramètre (*Modifier les paramètres | Général*) en monochrome.

13.13 trop lent

Le processus est lent

Si vous traitez un grand nombre de textes et que votre ordinateur est assez ancien, ne dispose que de peu de mémoire, et que le disque dur est de faible capacité, le processus peut prendre beaucoup de temps. Lorsque la [mémoire](#) est faible, vous entendez des cliquetis provenant du disque dur. Solution : vous pouvez rendre votre ordinateur plus rapide en installant plus de mémoire (ce qui fait une différence *considérable*), en défragmentant votre disque dur, en utilisant une mémoire cache disque ou en ajustant les paramètres de la mémoire vive. Si vous exécutez **Oxford WordSmith Tools** sur un réseau, consultez votre administrateur réseau, et demandez lui si les performances sont considérablement diminuées en raison de l'accès au réseau.

Solution 2 : quittez tous les programmes dont vous n'avez pas besoin. Cela peut restaurer une grande quantité de mémoire système.

Solution 3 : quittez Windows et redémarrez. Cela peut restaurer une grande quantité de mémoire système.

Solution 4 : enregistrez et lisez à partir du disque local, et non du réseau.

13.14 ne démarre pas

Le programme ne démarre pas.

13.15 la liste de mots ne fonctionne pas

WordList ne fonctionne pas

Les mots sont triés en fonction de routines Microsoft qui dépendent de la langue. Si vous traitez des fichiers en espagnol mais que vous laissez les paramètres de langue sur "anglais", vous obtiendrez des résultats qui ne respectent pas l'ordre de classement espagnol (par ex. **LL** viendra avant **LM**).

Solution : sélectionnez votre [langue](#) et re-calculez la liste de mots.

WordSmith Tools

Messages d'erreur

Section

XIV

14 Messages d'erreur

14.1 accès au fichier refusé

Accès au fichier refusé

Le fichier que vous souhaitez ouvrir est peut-être utilisé par un autre programme. Vous constaterez que la plupart des logiciels de traitement de texte marquent tous les fichiers texte ouverts comme étant en cours d'utilisation, et ne permettent pas à d'autres programmes d'y accéder, même pas pour une simple lecture. Fermez le fichier texte dans votre logiciel de traitement de texte.

14.2 associés introuvables

Associés introuvables

Modifiez les paramètres (*Paramètres | Fréquences min. et max.*) puis réessayez.

14.3 aucun co-occurent trouvé

Aucun co-occurent trouvé

Modifiez les paramètres via le [contrôleur](#) (*Modifier les paramètres | Concord | Fréquence min.*) puis réessayez. Il y avait trop peu de lignes de concordance pour trouver le nombre minimum nécessaire.

14.4 aucun élément activé

Aucun élément activé

Certaines formes comportent des sélections marquées "activé" que vous pouvez désactiver si vous le souhaitez. Si elles sont décochées, vous pouvez toujours les voir, mais **WordSmith** les ignorera.

14.5 aucun groupe identifié

Aucun groupe identifié

Modifiez les paramètres et réessayez.

14.6 aucun mot-clé à schématiser

Aucun mot-clé à traiter

Les avez-vous tous supprimés ?

14.7 aucun mot-clé trouvé

Aucun mot-clé trouvé

Modifiez les paramètres et réessayez. La fréquence minimum est définie sur une valeur trop basse et/ou la [p valeur](#) est trop faible pour qu'un mot-clé soit détecté. Pour des textes très courts, une fréquence minimum de 2 peut s'avérer nécessaire.

14.8 aucun terme de liste d'arrêt

Aucun terme de liste d'arrêt

WordSmith a échoué lors de la lecture de votre fichier de liste d'arrêt ou bien le fichier était vide.

14.9 aucun terme de liste d'arrêt de concordance

Aucun terme de liste d'arrêt de concordance

14.10 aucun terme de liste d'arrêt KeyWords

Aucun terme de liste d'arrêt KeyWords

WordSmith a échoué lors de la lecture de votre fichier de liste d'arrêt ou bien le fichier était vide.

14.11 aucun terme de liste de balises

Aucun terme de liste de balises

WordSmith a échoué lors de la lecture de votre fichier de balises ou bien le fichier était vide.

14.12 aucun terme de liste de correspondance

Aucun terme de liste de correspondance

WordSmith a échoué lors de la lecture de votre fichier de [liste de correspondance](#), ou bien le fichier était vide. Il se peut également que vous n'avez pas spécifié l'action à effectuer (une des options est *Aucune*). Une autre possibilité est que vous avez essayé d'effectuer une correspondance en utilisant une liste de mots ou un modèle alors que la colonne actuelle ne comporte que des nombres. Dernière solution : aucun des éléments spécifiés n'est présent.

14.13 aucun terme de liste de lemmes

Aucun terme de liste de correspondance de lemme

WordSmith a échoué lors de la lecture de votre fichier de liste de lemme ou bien le fichier était vide.

14.14 aucune entrée de concordance

Aucune entrée de concordance trouvée

Si vous n'obtenez aucune entrée de concordance, il peut y avoir plusieurs raisons : a) il n'y en avait aucune dans vos textes, b) vous avez mal spécifié ce que vous recherchez, c) un problème s'est produit avec la sélection du texte. Vérifiez que vous avez bien orthographié le terme de recherche et le terme de contexte. Si vous utilisez un [texte accentué](#), vérifiez le format de vos textes. Si vous utilisez un [fichier de termes de recherche](#), assurez-vous qu'il a été préparé avec un éditeur de texte Windows de base tel que Bloc-notes.

Avez-vous correctement spécifié les [caractères de remplacement](#) (* et ?) ? Si vous recherchez un point d'interrogation, vous avez peut-être saisi "?" correctement, mais n'oubliez pas que les points d'interrogation apparaissent généralement à la fin des phrases. Il faut donc écrire : *"?".

Astuce

Insérez un [astérisque](#) ou deux. Il y a plus de chances de trouver livre* que livre.

14.15 aucune entrée présente dans la base de données KeyWords

Aucune entrée présente dans la base de données KeyWords

Modifiez les paramètres et réessayez.

14.16 aucune ligne supprimée à ignorer

Aucune ligne supprimée à ignorer

Vous avez appuyé sur les touches Alt+Z mais n'avez supprimé aucune ligne à [ignorer](#). Cela n'a aucune conséquence.

14.17 aucune liste de mots sélectionnée

Aucune liste de mots sélectionnée

Afin que **WordSmith** sache quelles listes doivent être comparées, vous devez les sélectionner en cliquant sur une liste dans chaque dossier. Si vous changez d'avis, cliquez sur Annuler.

14.18 calcul KeyWords non valide

Calcul KeyWords non valide

Pour que **KeyWords** puisse calculer le nombre de mots-clé dans un fichier texte en le comparant à un corpus de référence, les deux doivent être dans la même langue et être classés de la même manière (ordre alphabétique, croissant). Si ce message apparaît c'est que vous calculez les mots-clés sans respecter ces critères. Solution : ouvrez chaque liste de mots puis vérifiez qu'elle est correcte et qu'elle est classée par ordre alphabétique de la même façon. Assurez-vous que toutes deux ont été créées avec les mêmes paramètres de langue et, si nécessaire, recalculez l'une d'entre elles ou les deux.

14.19 ce programme nécessite Windows 98 ou une version supérieure

Ce programme nécessite Windows 98 ou une version supérieure

À partir de la version 4.0, le programme est une version 32 bits (le fichier d'aide en ligne également).

14.20 c'est une version de démonstration

C'est une version de démonstration

Il faut [mettre à niveau](#) vers une version complète.

14.21 clusters introuvables

Clusters introuvables

Modifiez les paramètres (*Paramètres / Clusters*) puis réessayez. Il y a trop peu de lignes de concordance pour trouver le nombre minimum nécessaire, ou bien le cluster est trop long.

14.22 détails de la liste de correspondance non spécifiés

Détails de la liste de correspondance non spécifiés

Vous avez cliqué sur le bouton [liste de correspondance](#), mais avez sélectionné une liste de correspondance ou un modèle de filtre non valide. Réessayez.

14.23 deux fichiers sont nécessaires

Vous devez sélectionner 2 fichiers pour cette procédure. Sélectionnez deux noms de fichiers dans la liste de fichiers (cliquez en maintenant la touche Ctrl enfoncée).

14.24 dossier cible introuvable

Dossier cible introuvable

WordSmith n'a pas trouvé ce dossier, veuillez vérifiez que le nom saisi est correct.

14.25 échec lors de la lecture du fichier

Échec lors de la lecture

Il se peut que la gestion des dossiers ait eu un problème. Cela peut notamment se produire si vous utilisez de gros fichiers dans votre logiciel de traitement de texte. Il est conseillé de procéder à une analyse de disque en cliquant sur *Outils système* | *Scandisk*.

14.26 échec lors de l'enregistrement du fichier

Échec lors de l'enregistrement

Cela peut venir du fait que le dossier que vous enregistrez est un dossier en lecture seule sur un réseau, que le disque est plein, ou encore que votre système de fichiers sur le disque a rencontré un problème. En fait, ce dernier problème est assez courant et peut notamment se produire si vous utilisez de gros fichiers dans votre traitement de texte. Il est conseillé de procéder à une analyse de disque en cliquant sur *Outils système* | *Scandisk*. Si vous travaillez en réseau, vous pouvez [enregistrer](#) des fichiers sur certains lecteurs et dans certains dossiers mais pas dans d'autres ; réessayez alors sur le lecteur A: ou un disque dur sur lequel vous pouvez enregistrer.

14.27 erreur de liste de base

erreur de liste de base

WordSmith tente d'accéder à un mot ou une ligne de concordance située juste avant ou juste après les données calculées. C'est un bug.

14.28 espace insuffisant pour la variable calculée

Espace insuffisant pour la variable calculée

Il n'y a pas assez d'espace disque pour la variable que vous essayez de calculer.

14.29 fichier *.ini introuvable

fichier *.ini introuvable

Au démarrage, **WordSmith** recherche le fichier **wshell.ini** qui contient vos [paramètres par défaut](#) actuels. Si vous l'avez supprimé ou renommé, restaurez-le. Ce fichier doit se trouver dans le même dossier que celui où se trouvent les Outils.

14.30 fichier base de données KeyWords non valide

Fichier base de données KeyWords non valide

Chaque type de fichier créé par **Oxford WordSmith Tools** possède sa propre extension par défaut (par exemple *.KWS, *.KDB) ainsi que sa propre structure interne. Si vous possédez un autre fichier avec la même extension créé par un autre programme, il y aura incompatibilité. Il est déconseillé de renommer un fichier .KDB en un fichier .TXT ou vice-versa. **WordSmith** a détecté que le fichier demandé n'a pas été créé pour une base de données par la version actuelle de **KeyWords**.

14.31 fichier de balises introuvable

Fichier de balises introuvable

Vous avez saisi le nom d'un fichier inexistant. Lorsque vous saisissez un [nom de fichier](#), incluez le nom complet du lecteur, du dossier ainsi que du fichier lui-même.

14.32 fichier de balises non lu

Fichier de liste de balises non lu

Une erreur s'est produite lors de la lecture sur le disque. Le fichier que vous tentez de lire est peut-être corrompu. C'est une erreur courante si vous manipulez fréquemment de très gros fichiers, surtout si vous n'avez pas effectué d'*analyse de disque* récemment pour vérifier qu'aucun cluster du fichier n'a été perdu. Veuillez consulter votre manuel DOS ou Windows pour obtenir de plus amples informations sur la fragmentation.

14.33 fichier de comparaison WordList non valide

Fichier de comparaison Wordlist non valide

Chaque type de fichier créé par **Oxford WordSmith Tools** possède sa propre extension par défaut (par exemple *.LST, *.CNC) ainsi que sa propre structure interne. Si vous possédez un autre fichier avec la même extension créé par un autre programme, il y aura incompatibilité. Il n'est pas recommandé de renommer un fichier *.CNC en un fichier *.TXT ou vice-versa.

WordSmith a détecté que le fichier demandé n'est pas un fichier de comparaison créé par WordList.

14.34 fichier de concordance non valide

Fichier de concordance non valide

Chaque type de fichier créé par **Oxford WordSmith Tools** possède sa propre extension par défaut, (par exemple .CNC, .LST) ainsi que sa propre structure interne. Si vous possédez un autre fichier avec la même extension créé par un autre programme, il y aura incompatibilité. Il n'est pas recommandé de renommer un fichier *.CNC en un fichier *.TXT ou vice-versa.

WordSmith a détecté que le fichier demandé n'a pas été créé par la version actuelle de Concord.

14.35 fichier de conversion introuvable

Fichier de conversion introuvable

Vous avez saisi le nom d'un fichier inexistant. Lorsque vous saisissez un [nom de fichier](#), incluez le nom complet du lecteur, du dossier ainsi que du fichier lui-même.

14.36 fichier de liste d'arrêt de concordance introuvable

Fichier de liste d'arrêt de concordance introuvable

Vous avez saisi le nom d'un fichier inexistant. Lorsque vous saisissez un [nom de fichier](#), incluez le nom complet du lecteur, du dossier ainsi que du fichier lui-même.

14.37 fichier de liste d'arrêt introuvable

Fichier de liste d'arrêt introuvable

Vous avez saisi le nom d'un fichier inexistant. Lorsque vous saisissez un [nom de fichier](#), incluez le nom complet du lecteur, du dossier ainsi que du fichier lui-même.

14.38 fichier de liste d'arrêt non lu

Fichier de liste d'arrêt non lu

Une erreur s'est produite lors de la lecture sur le disque. Le fichier que vous tentez de lire est peut-être corrompu. C'est une erreur courante si vous manipulez fréquemment de très gros fichiers, surtout si vous n'avez pas effectué d'*analyse de disque* récemment pour vérifier qu'aucun cluster du fichier n'a été perdu. Veuillez consulter votre manuel DOS ou Windows pour obtenir de plus amples informations sur la fragmentation.


14.39 fichier de liste de mots introuvable

Vous avez saisi le nom d'un fichier inexistant. Lorsque vous saisissez un [nom de fichier](#), incluez le nom complet du lecteur, du dossier ainsi que du fichier lui-même.

14.40 fichier introuvable

Fichier introuvable

Ce message, tel que [Texte Original introuvable](#), peut apparaître lorsque **WordSmith** doit pouvoir accéder au texte source original utilisé lors de la création de la liste, mais ne le trouve pas.

L'avez-vous supprimé ou déplacé? Si le fichier est toujours disponible, vous pouvez [éditer les noms de fichiers](#) dans la fenêtre noms des fichiers () de cette liste.

Le message peut également apparaître une fois que vous avez fourni vous-même le nom du fichier. Vous avez peut-être fait une erreur de frappe. Est-ce un [nom de fichier long](#) Windows 95 ou NT? Lorsque vous saisissez un [nom de fichier](#), incluez le nom complet du lecteur, du dossier ainsi que du fichier lui-même.

14.41 fichier spécifié introuvable

Fichier spécifié introuvable

Vous avez saisi le nom d'un fichier inexistant. Lorsque vous saisissez un [nom de fichier](#), incluez le nom complet du lecteur, du dossier ainsi que du fichier lui-même.

14.42 fichier texte original nécessaire mais introuvable

Fichier texte original nécessaire mais introuvable

WordSmith doit trouver le [fichier](#) texte original sur lequel la liste est basée. Le programme n'a donc pas pu continuer. Ce fichier a été déplacé ou renommé.

Il se peut également, si vous utilisez un réseau, que la connexion réseau ne soit pas établie, que le réseau ne fonctionne pas ou que vous n'ayez pas inséré le bon disque ou CD-ROM dans le lecteur.

14.43 fichier WordList non valide

Fichier Wordlist non valide

Chaque type de fichier créé par **Oxford WordSmith Tools** possède sa propre extension par défaut (par exemple *.LST, *.CNC) ainsi que sa propre structure interne. Si vous possédez un autre fichier avec la même extension créé par un autre programme, il y aura incompatibilité. Il est déconseillé de renommer un fichier .LST en un fichier .TXT ou vice-versa. **WordSmith a détecté que le fichier demandé n'a pas été créé par la version actuelle de WordList.**

14.44 fichier(s) source introuvable(s)

Fichier(s) source(s) introuvable(s)

Vous avez saisi le nom d'un fichier inexistant. Lorsque vous saisissez un [nom de fichier](#), incluez le nom complet du lecteur, du dossier ainsi que du fichier lui-même.

14.45 format du code d'enregistrement erroné

Le code d'inscription doit avoir la forme suivante

2 lettres ou chiffres, point, suivi de 4 chiffres, point, 4 chiffres, etc.
Exemple : XX.1234.5678.9012.3456 (un point toutes les 4 lettres)

14.46 formulaire incomplet

Formulaire incomplet

Vous avez essayé de fermer un formulaire dans lequel un ou plusieurs champs devaient être remplis avant que **WordSmith** puisse continuer.

14.47 il doit y avoir un nombre

Il doit y avoir un nombre

Vous avez saisi autre chose qu'un nombre. Faites particulièrement attention aux minuscules **L** et **1**, et **O** (la lettre) au lieu de **0** (le nombre).

14.48 il n'est possible d'enregistrer les mots qu'au format ASCII

Seuls les mots peuvent être enregistrés au format texte brut

Oxford WordSmith Tools ne peut pas enregistrer les graphiques au format fichier texte. Si vous obtenez ce message d'erreur, vous ne pouvez enregistrer ce type de données qu'en les copiant dans le [presse-papiers](#) et en les collant dans un logiciel de traitement de texte.

14.49 impossible d'accéder à Internet

Impossible d'accéder à Internet

Cette fonction dépend des éléments suivants: a) si vous disposez d'un navigateur Internet sur votre ordinateur, b) si votre système "associe" une extension d'URL Internet **.htm** avec ce navigateur.

14.50 impossible d'accéder au fichier texte - fichier utilisé ailleurs ?

Impossible d'accéder au fichier texte : le fichier est-il utilisé ailleurs ?

Impossible d'accéder au fichier. Il est peut-être en cours d'utilisation par une autre application. Si c'est le cas, fermez le fichier dans l'application en question puis réessayez.

14.51 impossible d'appeler un autre outil

Impossible d'appeler un autre outil

La communication inter-outils a été interrompue. [Enregistrez](#) d'abord votre travail. Puis, si nécessaire, fermez complètement **Oxford WordSmith Tools** et redémarrez le programme principal **wshell.exe**.

14.52 impossible de calculer les mots-clés si les langues diffèrent

Impossible de calculer les mots-clés si les langues diffèrent

Les mots-clés ne peuvent être calculés que si le fichier texte et le corpus de référence sont dans la même langue. Il est possible de calculer les mots-clés lorsque les deux textes utilisent deux variantes d'anglais ou deux variantes d'espagnol, mais pas lorsqu'un texte est en anglais et l'autre en français.

14.53 impossible de créer le dossier car un autre dossier porte le même nom

Impossible de créer le dossier car un autre dossier porte le même nom

S'il existe déjà un *fichier* nommé C:\TEMP\FRED, vous ne pouvez pas créer de *sous-dossier* de C:\TEMP nommé FRED. Veuillez saisir un autre nom.

14.54 impossible de créer un nouveau nom de dossier

Échec lors de la création d'un nouveau dossier

Un dossier et un fichier ne peuvent pas porter le même nom. S'il existe déjà un *fichier* nommé C:\TEMP\FRED, vous ne pouvez pas créer de *sous-dossier* de C:\TEMP nommé FRED. Veuillez saisir un autre nom.

14.55 impossible de fusionner la liste avec elle-même

Impossible de fusionner la liste avec elle-même

Vous ne pouvez fusionner une liste de mots ou base de données de mots-clés qu'avec une seule autre à la fois. Sélectionnez deux noms de fichiers dans la liste de fichiers (cliquez en maintenant la touche Ctrl enfoncée).

14.56 impossible de fusionner les bases de données KeyWords

La quantité de mémoire [RAM](#) est peut-être insuffisante.

14.57 impossible de lire le fichier

Can't read file

Si cette erreur se produit au démarrage de **Oxford WordSmith Tools**, elle est probablement due à un fichier manquant. Prenons l'exemple du fichier **sayings.txt**, qui contient les citations apparaissant dans la fenêtre principale du [contrôleur](#). Si vous l'avez supprimé, la meilleure solution est d'utiliser le **Bloc-notes** pour créer un nouveau fichier **sayings.txt**, et de ne mettre qu'une ligne vierge dedans.

Si ce message d'erreur apparaît à un autre moment, c'est qu'un problème s'est produit lors de la lecture sur disque. Le fichier que vous tentez de lire est peut-être corrompu. C'est une erreur courante si vous manipulez fréquemment de très gros fichiers, surtout si vous n'avez pas effectué d'*analyse de disque* récemment pour vérifier qu'aucun cluster du fichier n'a été perdu. Veuillez consulter votre manuel DOS ou Windows pour obtenir de plus amples informations sur la fragmentation.

14.58 imprimante nécessaire

Un pilote d'imprimante doit être installé sur votre ordinateur, même si vous n'imprimez aucun document. Vous n'avez pas besoin d'acheter ou d'allumer une imprimante, mais la fonction [Aperçu avant impression](#) de Concord, WordList et KeyWords doit connaître la taille du papier sur lequel vous imprimerez. Si un message s'affiche indiquant qu'aucune imprimante n'a été installée, cliquez sur Démarrer | Paramètres | Imprimantes et Fax et installez une imprimante par défaut (n'importe laquelle) dans Windows.

14.59 informations mutuelles incompatibles

La liste des informations mutuelles est incompatible

Une liste d'informations mutuelles provient d'un fichier d'index et sait de quel fichier d'index elle provient lorsqu'elle est calculée. Normalement, le fichier d'index correspondant s'ouvre en même temps que la liste d'informations mutuelles. Ce message s'affiche si le fichier d'index n'a pas été trouvé sur votre ordinateur ou qu'il a été renommé. Vous pouvez toujours accéder aux informations mutuelles, mais a) les informations sur la fréquence et les listes alphabétiques se rapportent à un autre fichier d'index et b) il est impossible d'obtenir des concordances directement à partir de la liste.

14.60 inscription non valide

Inscription non valide

Elle ne répond pas aux exigences d'une version complètement mise à jour. Le code d'inscription des versions antérieures n'est plus utilisé. **WordSmith** fonctionnera toujours, mais en [version de démonstration](#).

14.61 inscription réseau utilisée ailleurs

Inscription réseau effective à un autre emplacement

L'inscription pour une utilisation sur réseau n'est pas valide pour une utilisation sur un ordinateur seul, et vice-versa. Si ce message s'affiche, recommencez une installation adéquate.

14.62 jeu de caractères réinitialisé sur <x> pour convenir à : <language>

Jeu de caractères réinitialisé sur <x> pour convenir à la balise <language>

Avant la version 2.00.07, **Oxford WordSmith Tools** ne prenait pas en charge autant de [jeux de caractères](#) et de [langues](#) qu'actuellement. Il en résulte que les données enregistrées au format utilisé précédemment peuvent ne pas "reconnaître" la langue utilisée. Si ce message s'affiche lorsque vous ouvrez un ancien fichier de données **WordSmith**, c'est parce que **WordSmith** ne reconnaît pas la langue utilisée. La langue par défaut est l'anglais.

Si les données sont correctes, cliquez simplement sur le bouton Enregistrer afin que les paramètres soient gardés en mémoire la prochaine fois que vous utilisez le programme. Sinon, réinitialisez la langue dans le [contrôleur](#), *Modifier les paramètres | Texte*, puis enregistrez à nouveau la liste.

14.63 la fonction n'est pas encore disponible

la fonction n'est pas encore disponible

Cette fonction est en cours de développement, elle n'est pas encore totalement implémentée.

14.64 la fonction n'est pas encore prête

Cette fonction n'est pas encore prête

Il s'agit d'un message temporaire pour les fonctions encore en cours de test.

14.65 le fichier comporte des "trous"

Le fichier comporte des "trous"

Votre fichier texte est défectueux. Il est probable qu'il contienne du texte utile, mais également un ou plusieurs caractères non reconnus tels que (0). Ce problème se produit probablement suite à un transfert du fichier d'un système à un autre. Il se peut aussi que le disque soit corrompu, que le fichier comporte des caractères non reconnus ou encore qu'il ne s'agisse pas d'un fichier en texte brut mais par exemple d'un [document Word](#).

Vous pourrez voir le contexte dans lequel le problème s'est produit, et une indication vous sera donnée quant à l'emplacement approximatif de l'erreur dans le texte.

WordSmith peut continuer si vous le souhaitez, mais vous avez la possibilité d'ignorer le texte.

Il est possible de résoudre le problème (de toute façon, il reviendra à chaque fois que vous ouvrirez ce fichier) en ouvrant le fichier dans un traitement de texte, et en l'enregistrant en texte brut, au format *.txt. De plus, dans [File Utilities](#), il existe un outil permettant de détecter ce type de fichier.

14.66 le fichier de base de données KeyWords est défectueux

Le fichier de base de données KeyWords est défectueux

Chaque type de fichier créé par **Oxford WordSmith Tools** possède sa propre extension par défaut, (par exemple *.KDB, *.KWS) ainsi que sa propre structure interne. Si vous possédez un autre fichier avec la même extension créé par un autre programme, il y aura incompatibilité. Il est déconseillé de renommer un fichier .KDB en un fichier .TXT ou vice-versa. **WordSmith** a détecté que le fichier demandé n'a pas été créé pour une base de données de mots-clés par la version actuelle de **KeyWords**.

14.67 le fichier de comparaison WordList est défectueux

Chaque type de fichier créé par **Oxford WordSmith Tools** possède sa propre extension par défaut, (par exemple *.LST, *.KWS) ainsi que sa propre structure interne. Si vous possédez un autre fichier avec la même extension créé par un autre programme, il y aura incompatibilité. Il n'est pas recommandé de renommer un fichier *.CNC en un fichier *.TXT ou vice-versa.

WordSmith a détecté que le fichier demandé n'est pas un fichier de comparaison créé par **WordList**.

14.68 le fichier de concordance n'est pas valide

Le fichier de concordance n'est pas valide

Chaque type de fichier créé par **Oxford WordSmith Tools** possède sa propre extension par défaut, (par exemple *.CNC, *.LST) ainsi que sa propre structure interne. Si vous possédez un autre fichier avec la même extension créé par un autre programme, il y aura incompatibilité. Il n'est pas recommandé de renommer un fichier *.CNC en un fichier *.TXT ou vice-versa.

WordSmith a détecté que le fichier demandé n'a pas été créé par la version actuelle de **Concord**

14.69 le fichier de liste de mots est défectueux

Chaque type de fichier créé par **Oxford WordSmith Tools** possède sa propre extension par défaut, (par exemple *.LST, *.KWS) ainsi que sa propre structure interne. Si vous possédez un autre fichier avec la même extension créé par un autre programme, il y aura incompatibilité. Il n'est pas recommandé de renommer un fichier *.CNC en un fichier *.TXT ou vice-versa.

WordSmith a détecté que le fichier demandé n'a pas été créé par la version actuelle de **WordList**.

14.70 le fichier KeyWords est défectueux

Le fichier KeyWords est défectueux

Chaque type de fichier créé par **Oxford WordSmith Tools** possède sa propre extension par défaut (par exemple *.KWS, *.KDB) ainsi que sa propre structure interne. Si vous possédez un autre fichier avec la même extension créé par un autre programme, il y aura incompatibilité. Il est déconseillé de renommer un fichier .KWS en un fichier .TXT ou vice-versa. **WordSmith** a

détecté que le fichier demandé n'a pas été créé par la version actuelle de **KeyWords**.

14.71 le fichier ne contient aucune des balises spécifiées

Le fichier ne contient aucune des balises spécifiées

Aucune des balises que vous avez spécifiées n'a été trouvée.

14.72 lecteur non valide

Lecteur non valide

WordSmith ne peut pas accéder à ce lecteur. Cela peut se produire si vous tentez d'accéder à un lecteur qui n'existe pas. Par exemple, vous spécifiez le lecteur P:, alors que vous n'avez que les lecteurs A:, C:, D: et E:.

14.73 les dispersions vont de paire avec les concordances

Les dispersions vont de paire avec les concordances

Elles ne peuvent être [enregistrées](#) séparément.

14.74 les liens entre les outils sont défailants

Les liens entre les outils sont défailants

Le [contrôleur](#) d'Oxford WordSmith Tools ou un des outils de la suite a appelé un autre outil et l'opération a échoué. Un autre programme en cours d'exécution a peut-être rencontré un problème ou bien la mémoire est insuffisante. Les [liens](#) sont essentiels dans cette suite car ils assurent la communication entre les outils. Quittez WordSmith et redémarrez l'application.

14.75 les noms de fichiers doivent être différents

Les noms de fichiers doivent être différents

Vous ne pouvez pas comparer un fichier à lui-même.

14.76 limite de regroupement atteinte

Limite de regroupement atteinte : regroupez à nouveau et réessayez

Le nombre de mots pouvant être [lemmatisés](#) au cours d'une même opération est limité. Si la limite est atteinte et que ce message apparaît :

1. lemmatisez en appuyant sur F4
2. mettez à nouveau l'entrée principale en surbrillance
3. appuyez sur F5 et poursuivez la lemmatisation en appuyant sur F5 pour chaque entrée que vous souhaitez attacher à l'entrée principale
4. lorsque vous avez terminé, appuyez sur F4 pour les regrouper.

14.77 limite de termes de recherche des fichiers atteinte

Limite de termes de recherche atteinte

Vous ne pouvez pas traiter plus de 15 termes de recherche à la fois, sauf si vous utilisez un [fichier de termes de recherche](#) pour indiquer à **Concord** de les mettre dans un lot, dans une limite de 500.

14.78 liste des messages d'erreur

Liste des messages d'erreur

Rubriques connexes : [Résolution des problèmes](#).

[Il n'est possible d'enregistrer les mots qu'au format ASCII](#)

[Impossible d'appeler un autre outil](#)

[Impossible de créer le dossier car un autre dossier porte le même nom](#)

[Impossible de fusionner la liste](#)

[Impossible de lire le fichier](#)

[Jeu de caractères réinitialisé sur <x> pour convenir à la balise <language>](#)

[Le fichier de concordance n'est pas valide](#)

[Fichier de liste d'arrêt de concordance introuvable](#)

[Fichier de conversion introuvable](#)

[Dossier cible introuvable](#)

[Problème de disque : le fichier n'a pas été enregistré](#)

[Les dispersions vont de paire avec les concordances](#)

[Lecteur non valide](#)

[Impossible d'accéder à Internet](#)
[Impossible de créer un nouveau nom de dossier](#)
[Accès au fichier refusé](#)
[Le fichier ne contient aucune des balises spécifiées](#)
[Fichier introuvable](#)
[Les noms de fichiers doivent être différents](#)
[nom complet lecteur:\dossier nécessaire](#)
[la fonction n'est pas encore disponible](#)
[fichier INI introuvable](#)
[Fichier de concordance non valide](#)
[Nom de fichier non valide](#)
[Fichier base de données KeyWords non valide](#)
[Fichier KeyWords non valide](#)
[Fichier de comparaison Wordlist non valide](#)
[Fichier Wordlist non valide](#)
[Limite de regroupement atteinte : regroupez à nouveau et réessayez](#)
[Le fichier KeyWords est défectueux](#)
[Le fichier de base de données KeyWords est défectueux](#)
[Limite de 500 termes de recherche des fichiers atteinte](#)
[Les liens entre les outils sont défaillants](#)
[Détails de la liste de correspondance non spécifiés](#)
[Il doit y avoir un nombre](#)
[Inscription réseau effective à un autre emplacement](#)
[Impossible d'accéder au fichier texte : le fichier est-il utilisé ailleurs ?](#)
[Associés introuvables](#)
[Aucun groupe identifié](#)
[Clusters introuvables](#)
[Aucun co-occurent trouvé](#)
[Aucune entrée de concordance trouvée](#)
[Aucun terme de liste d'arrêt de concordance](#)
[Aucune ligne supprimée à ignorer](#)
[Aucune entrée présente dans la base de données KeyWords](#)
[Aucun mot-clé trouvé](#)
[Aucun mot-clé à traiter](#)
[Aucun terme de liste d'arrêt KeyWords](#)
[Aucun terme de liste de lemmes](#)
[Aucun terme de liste de correspondance](#)
[Espace insuffisant pour la variable calculée](#)
[Statistiques non disponibles](#)
[Aucun terme de liste d'arrêt](#)
[Fichier spécifié introuvable](#)
[Aucun terme de liste de balises](#)
[Nombre non valide](#)
[Aucune liste de mots sélectionnée](#)
[Fichier texte original nécessaire mais introuvable](#)
[Chaîne d'inscription non valide](#)
[La chaîne d'inscription doit comporter 20 lettres](#)
[Mémoire insuffisante](#)
[Fichier\(s\) source\(s\) introuvable\(s\)](#)
[Fichier de liste d'arrêt introuvable](#)
[Fichier de liste d'arrêt non lu](#)
[Fichier de balises introuvable](#)
[Fichier de liste de balises non lu](#)
[Cette fonction n'est pas encore prête](#)
[C'est une version de démonstration](#)
[Ce programme nécessite Windows 95 ou une version supérieure](#)
[Pour ne plus voir ce message s'afficher, mettez à niveau à partir de la version de démonstration](#)

[dans setup.exe](#)
[Nombre de fichiers ignorés trop important \(limite : 50\)](#)
[Nombre de phrases trop important \(limite : 8 000\)](#)
[Deux fichiers sont nécessaires](#)
[Troncature à xx mots – le fichier de liste de balises en comporte davantage](#)
[Impossible de fusionner les bases de données KeyWords](#)
[Pourquoi la recherche a-t-elle échoué ?](#)
[Fichier de liste de mots introuvable](#)
[Le fichier de comparaison WordList est défectueux](#)
[Le fichier de liste de mots est défectueux](#)
[Oxford WordSmith Tools a expiré : obtenir une autre version](#)
[Oxford WordSmith Tools est déjà en cours d'exécution](#)
[Non correspondance de version de WordSmith](#)
[xx jours restants](#)

14.79 mémoire insuffisante

Mémoire insuffisante

L'opération a été interrompue en raison d'un manque de mémoire [RAM](#)

14.80 messages de confirmation : acceptez-vous une relecture ?

Acceptez-vous une relecture ?

C'est un message de confirmation. Afin de continuer, **Viewer** va maintenant relire le fichier disque. Cela aura un impact sur toutes les modifications que vous avez apportées à l'affichage. Vous avez la possibilité d'enregistrer le fichier d'abord puis de réessayer ultérieurement. De plus, Viewer essaiera de lire le fichier texte en entier (dans une limite de 16 368 phrases). Si vous lisez un très gros fichier sur un lecteur CD-ROM lent, le processus prendra du temps.

14.81 nom de fichier non valide

Nom de fichier non valide

[Les noms de fichiers](#) ne peuvent pas contenir d'espaces ou certains symboles comme ? et *. Dans les versions de Windows antérieures à Windows 95, ils étaient limités à 8 lettres, un point et trois lettres supplémentaires aussi. Réessayez.

14.82 nombre de phrases trop important

La limite est de 8 000. Fragmentez la tâche.

14.83 nombre non valide

Nombre non valide

Soit vous venez de saisir une entrée, ou alors **Oxford WordSmith Tools** vient d'essayer de lire (par exemple à partir de **wshell.ini**, le fichier des [paramètres par défaut](#)) un élément qui devrait être un nombre mais n'en est pas un. Les ordinateurs ne reconnaissent pas la majuscule **O** comme équivalent du nombre **0**. Il peut aussi y avoir un nombre accompagné d'autres lettres ou symboles, par exemple **£30**. Si cela se produit lorsque **WordSmith** démarre, vérifiez si le fichier **wshell.ini** comporte des erreurs.

14.84 noms complets du lecteur et du dossier nécessaires

nom complet lecteur:\dossier nécessaire

Lorsque vous saisissez un [nom de fichier](#), incluez le nom complet du lecteur, du dossier ainsi que du fichier lui-même.

14.85 Non correspondance de version de WordSmith

Comme les différents outils sont [liés](#) les uns aux autres, il est important de s'assurer que les fichiers composants sont compatibles entre eux. Si vous voyez ce message s'afficher, c'est qu'un des composants est daté différemment des autres.

Solution : téléchargez ceux dont vous avez besoin à partir d'un des [sites Internet](#) de contact.

14.86 pas un fichier WordSmith

Le fichier que vous essayez d'ouvrir n'est pas un fichier Oxford WordSmith Tools. WordSmith crée des fichiers contenant vos résultats, des fichiers dont l'extension est *.LST, *.CNC, *.KWS, etc. Ces fichiers sont au format natif WordSmith et ne peuvent être ouverts par Microsoft Word. De même, un [fichier texte brut](#) ou un fichier Word *.doc ne peut pas être lu par WordSmith en tant que fichier de données, mais seulement comme fichier texte à traiter.

Rubriques connexes : [Conversion de données provenant de versions antérieures](#)

14.87 pas un fichier WordSmith actuel

Pas un fichier WordSmith actuel

Le fichier que vous essayez d'ouvrir a été créé avec WordSmith, mais :

il a été créé avec une version antérieure (versions 1 à 3)

ou


c'est un fichier créé avec la version beta de WordSmith 4, et le format a dû être changé.

Dans le premier cas, vous pouvez convertir votre fichier avec [Convertir](#).

14.88 pour que ce message ne s'affiche plus...

Effectuez une mise à niveau. Ce message s'affiche à l'ouverture de la [version de démonstration](#).

14.89 pourquoi la recherche a-t-elle échoué ?

La recherche standard (F12 ou ) dans le cas d'une liste de données s'effectue sur la colonne en surbrillance. Si vous souhaitez effectuer une recherche au sein de données dans une autre colonne, cliquez d'abord sur cette colonne.

Par défaut, la recherche est une "recherche de mot entier". Utilisez * au début ou à la fin du mot ou nombre que vous recherchez pour le trouver, par exemple pour des données comportant plusieurs mots. L'avantage que présente le système avec astérisque est que cela vous permet de spécifier un préfixe, un suffixe, ou les deux, contrairement à l'option "recherche de mot entier" standard de Windows.

14.90 problème de disque – le fichier n'a pas été enregistré

Problème de disque : le fichier n'a pas été enregistré

Une erreur s'est produite lors de l'écriture sur le disque. L'espace disque est peut-être insuffisant. Si c'est le cas, veuillez supprimer des fichiers de ce disque.

14.91 statistiques non disponibles

Statistiques non disponibles

Certains types de listes de mots créées par **Oxford WordSmith Tools**, par exemple une liste de mots d'une base de données de mots-clés, contiennent les mots classés par ordre de fréquence et par ordre alphabétique, mais il n'y a aucune statistique sur les fichiers texte originaux. Il n'est donc pas possible d'appeler les statistiques dans **WordList**. Ce message d'erreur apparaîtra également si le fichier de statistiques que vous tentez d'appeler est corrompu.

14.92 troncature à xx mots – le fichier de liste de balises en comporte davantage

Le fichier de liste de balises comporte plus d'entrées que la limite actuelle. Il se peut également que ce ne soit pas du tout un fichier de liste de balises.

14.93 trop de requêtes pour ignorer les groupes correspondants

La limite est de 50. Regroupez manuellement les entrées restantes.

14.94 utilisation sur l'ordinateur X uniquement

Utilisation avec un ordinateur nommé XXX exclusivement

Le logiciel a été enregistré pour fonctionner sur un autre ordinateur. Si ce message s'affiche, recommencez une installation adéquate.

14.95 WordSmith Tools a expiré

Ce message ne s'affiche que si la période d'utilisation est limitée. Cette version de **Oxford WordSmith Tools** a dépassé sa période de validité et tourne actuellement en mode [démonstration](#). Téléchargez une autre version à partir d'[Internet](#).

14.96 WordSmith Tools est déjà en cours d'utilisation

N'essayez pas de redémarrer **Oxford WordSmith Tools** s'il est déjà en cours d'utilisation. Appuyez simplement sur Alt+Tab pour retourner à l'instance en cours d'utilisation. En revanche, plusieurs copies de chaque outil peuvent être en cours d'utilisation en même temps.

14.97 XX jours restants

Ce message ne s'affiche que si la période d'utilisation est limitée. À la fin de cette période, **WordSmith** repassera en mode [démonstration](#).

Index

- 2 -

25 lignes 199

- A -

accents 174
 accents et symboles 174
 accès aux résultats précédents 45
 adresses de contact 178
 affichage de comparaison 103
 affichage décimal 131
 affichage des concordances 68
 affichage des schémas 94
 affichage hexadécimal 131
 afficher l'aide au démarrage 48
 afficher le fichier texte original 68
 afficher le nom des fichiers 51
 ajouter au texte 87
 ajouter des notes aux données 18
 ajustement avec la souris 130
 alignement 40, 130
 alignement du texte avec Viewer 130
 Alt+Tab 55
 analyse de cohérence (détaillée) 105
 analyse de cohérence (simple) 106
 Analyse de deux listes de mots 86
 analyse de plusieurs fichiers 91
 ansi 175
 aperçu avant impression 44
 apostrophes et tri 79
 appeler d'autres outils 184
 appeler une concordance 89
 arrêt 52
 arrêter le traitement 52
 ascii 175
 associé - définition 86
 associés 87
 associés de collocation 87
 astérisque 81
 autres termes de recherche 81

- B -

Babbage 200
 balise la plus proche 76
 balises de fichiers son 162
 balises en marron 161
 balises HTML et SGML 161
 balises multimédia 162
 balises utilisées comme sélecteurs 163
 balises : présentation générale 160
 Baltiques 23
 barre d'état 185, 193
 barre d'outils 185
 basculer du mode édition au mode saisie 181
 base de données KeyWords 91
 batch processing: file-names 21
 batch processing: folders 21
 bibliographie 172
 BNC : fichier de balises 166
 BNC : sélection au sein de textes 165
 BNC : sélection parmi plusieurs textes 164
 BNC: format texte 182
 booléen et/non 74
 booléen ou 81
 boutons 185
 bugs 173
 bureau 183

- C -

Calcul des mots-clés 92
 calcul des schémas 93
 calculer la fréquence des données 30
 calculer une nouvelle colonne de données 29
 caractères accentués 18
 caractères dans l'enregistrement au format texte 82
 caractères dans un mot 53
 caractéristiques de texte 53
 catégories définies par l'utilisateur 59
 CD-ROM : stockage 191
 CD-ROM : vitesse 193
 cela ne fonctionne pas 198
 charger automatiquement le fichier de balises 161
 Charles Babbage 200

- Cherokee 176
chi-deux 92
Chinois 176
choix de copie 32
clavier 188
clés de recherche 114
clusters 192
clusters dans KeyWords 88
clusters : définition 179
codes 174
codes des jeux de caractères Windows 175
codes DOS 175
Codes Unicode 176
collocation : paramètres 60
collocation : spécifications 60
Colonne Classe 59
combinaisons de touches pour les raccourcis clavier 188
comment créer une base de données 89
comparaison de listes de mots 102
Concord : affichage des options 68
Concord : astuces pour bien commencer 11
Concord : balise la plus proche 76
Concord : caractères de remplacement 81
Concord : catégories 59
Concord : ce que vous pouvez voir et faire 68
Concord : clusters 64
Concord : collocation 63
Concord : création d'exercices 58
Concord : enregistrement et impression 67
Concord : étendre l'affichage 68
Concord : fichier texte source 68
Concord : ignorer les lignes indésirables 75
Concord : limites 183
Concord : présentation générale 4, 58
Concord : schéma uniforme 66
Concord : schémas 79
Concord : segments de texte 80
Concord : son et vidéo 68
Concord : termes de recherche multiples 75
Concord : tri 79
concordance - traitement par lots 70
concordance automatisée des fichiers 75
concordance des balises 73
concordance multimédia 68
configuration minimale requise 189
connexion 36
contexte de balises 70
contrôleur (wshell.exe) 4
contrôleur WordSmith : Concord : paramètres 82
contrôleur WordSmith : Paramètres de KeyWords 97
contrôleur WordSmith : Paramètres de WordList 127
convertir 152
convertir les données provenant de versions antérieures 136
co-occurents 63
co-occurents : affichage 61
co-occurents : fréquence minimum 60
co-occurents : tri 80
co-occurents : voisinage 60
copie : sélective 32
copier les données dans Word 176
copier : spécifier 32
copier : tous 32
corpus de référence 191
correspondance de lemmes : WordList 101
couleurs 26
couleurs dans les balises 166
couper les débuts de ligne 165
création de la base de données 89
création d'un fichier de balises 166
création d'un index WordList 110
Création d'une base de données 89
créer une liste de mots à partir de données de mots-clés 93
Cyrillique 23
- D -**
- date de la version 194
date d'expiration 218
début et fin de phrase 168
décimales 40
définition de la pertinence 90
définition des balises multimédia 162
Définition des options par défaut 48
définition d'un associé 86
définition d'un mot-clé clé 90
définitions 179
départager 79
déplacement de phrases 130

dernier fichier 183
 détails sur la commande 180
 deux fichiers sont nécessaires 206
 différences de la version 4 171
 dispersion 66
 dissociation 132
 dissocier 107
 données sous forme de fichier texte 94
 dossier d'instructions 15
 dossiers 182
 dossiers de Text Converter 153

- E -

écran des informations mutuelles 117
 écrasement 153
 éditer des entrées de WordList 38
 effacer la sélection précédente 23
 effectuer un regroupement de synonymes 95
 effectuer une concordance sur des balises 73
 English 200
 enregistrement des résultats 49
 enregistrement partiel 49
 enregistrer au format HTML 46
 enregistrer au format texte 46
 enregistrer au format XML 46
 enregistrer la mise en page 40
 enregistrer les valeurs par défaut 48
 Enregistrer un ensemble de fichiers texte favoris 22
 enregistrer une partie des données 49
 en-têtes (spécifier) 124
 en-têtes HTML : suppression 163
 en-têtes : début et fin 168
 en-têtes : définition 179
 entrées associées 107
 étendre et réduire 68
 Excel 46
 exemple 90
 exemple de mots-clés 90
 exercices 58
 exporter vers une feuille de calcul, etc. 46

- F -

fermer 45

fichier *.dll personnalisé 33
 fichier de balises 166
 fichier de liste de mots introuvable 208
 fichier journal retraçant les problèmes 36
 fichier texte : utiliser pour créer une liste de mots 120
 fichiers d'index 48
 fichiers son et vidéo balisés 68
 fichiers zip 195
 file associations 170
 File Utilities : comparer deux fichiers 150
 File Utilities : découpeur de fichiers 151
 File Utilities : index 147
 File Utilities : présentation générale 6
 File Utilities : rechercher des doublons 151
 File Utilities : renommer 152
 file-types 170
 filtrer 42
 fonction de Splitter 147
 fonction de Text Converter 152
 fonction de Viewer 130
 format 40
 format de date 179
 formats de texte 53
 former les étudiants 40
 fréquence de happi* 30
 fusionner des concordances 103
 fusionner des listes de mots 103

- G -

gestion de la fenêtre 55
 gestion de multiples fenêtres 55
 gestion des types de balises 161
 Grec 23
 groupes 87
 groupes de synonymes : regrouper 95
 guillemets 198

- H -

HTML
 XML 182

- I -

ignorer 56
 illisible 200
 importer du texte dans une liste de mots 120
 impossible d'afficher les balises Concord 198
 impossible de comparer des listes de mots dans différentes langues 205
 impossible de fusionner des bases de données
 KeyWords 210
 impression 44
 impression des colonnes 190
 informations mutuelles : calcul 117
 informations sur la version de WordSmith 194
 installation de WordSmith Tools 15
 instructions de cours 40
 instructions du professeur 40
 interface 170
 international versions 170
 Internet Explorer 189
 introduction à WordSmith Tools 2
 irrégularité 189

- J -

Japonais 176
 jeux de caractères 174

- K -

KeyWords : affichage 96
 KeyWords : astuces 91
 KeyWords : astuces pour bien commencer 12
 KeyWords : but 86
 KeyWords : calcul 92
 KeyWords : clusters 88
 KeyWords : conseils 91
 KeyWords : liens 92
 KeyWords : limites 183
 KeyWords : présentation générale 5
 KeyWords : tri 95

- L -

la liste de mots est défectueuse 213

langue 23
 Langues d'Europe centrale 23
 largeur de colonne 40
 Le fichier de comparaison WordList est défectueux 213
 lemmatisation à partir d'un fichier 107
 lemmatiser avec un fichier *.dll personnalisé 33
 lemmes 107
 lent 201
 levée d'ambiguïté 87
 liens entre les outils 184
 limites 183
 limites spécifiques 183
 liste de correspondance 42
 liste des boutons 185
 liste d'historique 45, 81
 listes d'arrêt 51
 listes de cohérence : tri 119
 listes d'index WordList : affichage 109
 listes d'index : utilisation 108
 listes multiples 18
 listes précédentes 45
 localisation 170
 locutions 192
 log de vraisemblance 92
 l'opérateur ~ 74

- M -

manuel de WordSmith Tools 189
 marques en violet 68
 marqueur de fin de paragraphe 53
 marqueur de fin de phrase 53
 Marqueur de fin d'en-tête 53
 marqueur de paragraphe 53
 marqueur de phrase 53
 marqueur d'en-tête 53
 masquer des entrées 58
 mémoire RAM disponible 191
 menus contextuels 36
 message d'avertissement 217
 messages de confirmation : acceptez-vous une relecture ? 216
 messages d'erreur 214
 messages d'erreur : accès au fichier refusé 203
 messages d'erreur : associés introuvables 203

- messages d'erreur : aucun co-occurent trouvé 203
- messages d'erreur : aucun élément activé 203
- messages d'erreur : aucun groupe identifié 203
- messages d'erreur : aucun mot-clé à schématiser 203
- messages d'erreur : aucun mot-clé trouvé 203
- messages d'erreur : aucun terme de liste d'arrêt 204
- messages d'erreur : aucun terme de liste d'arrêt de concordance 204
- messages d'erreur : aucun terme de liste d'arrêt KeyWords 204
- messages d'erreur : aucun terme de liste de balises 204
- messages d'erreur : aucun terme de liste de correspondance 204
- messages d'erreur : aucun terme de liste de lemmes 204
- messages d'erreur : aucune entrée de concordance trouvée 204
- messages d'erreur : aucune entrée présente dans la base de données KeyWords 205
- messages d'erreur : aucune ligne supprimée à ignorer 205
- messages d'erreur : aucune liste de mots sélectionnée 205
- messages d'erreur : ce programme nécessite Windows 98 ou une version supérieure 205
- messages d'erreur : c'est une version de démonstration 205
- messages d'erreur : chaîne d'inscription non valide 211
- messages d'erreur : clusters introuvables 205
- messages d'erreur : dossier cible introuvable 206
- messages d'erreur : échec lors de la création d'un nouveau dossier 210
- messages d'erreur : échec lors de la lecture du fichier 206
- messages d'erreur : échec lors de l'enregistrement 206
- messages d'erreur : erreur de liste de base 206
- messages d'erreur : espace insuffisant pour la variable calculée 206
- messages d'erreur : fichier *.ini introuvable 206
- messages d'erreur : fichier base de données KeyWords non valide 207
- messages d'erreur : fichier de balises introuvable 207
- messages d'erreur : fichier de balises non lu 207
- messages d'erreur : fichier de comparaison WordList non valide 207
- messages d'erreur : fichier de concordance non valide 207
- messages d'erreur : fichier de conversion introuvable 207
- messages d'erreur : fichier de liste d'arrêt de concordance introuvable 208
- messages d'erreur : fichier de liste d'arrêt introuvable 208
- messages d'erreur : fichier de liste d'arrêt non lu 208
- messages d'erreur : fichier introuvable 208
- messages d'erreur : fichier KeyWords non valide 205
- messages d'erreur : fichier spécifié introuvable 208
- messages d'erreur : fichier texte original nécessaire mais introuvable 208
- messages d'erreur : fichier WordList non valide 209
- messages d'erreur : fichier(s) source introuvable(s) 209
- messages d'erreur : formulaire incomplet 209
- messages d'erreur : il doit y avoir un nombre 209
- messages d'erreur : il n'est possible d'enregistrer les mots qu'au format ASCII 209
- messages d'erreur : impossible d'accéder à Internet 209
- messages d'erreur : impossible d'accéder au fichier texte - fichier utilisé ailleurs ? 210
- messages d'erreur : impossible d'appeler un autre outil 210
- messages d'erreur : impossible de créer le dossier car un autre dossier porte le même nom 210
- messages d'erreur : impossible de fusionner la liste avec elle-même 210
- messages d'erreur : impossible de lire le fichier 211
- messages d'erreur : impossible de trouver l'imprimante 211
- messages d'erreur : inscription réseau utilisée ailleurs 211
- messages d'erreur : jeu de caractères réinitialisé sur <x> pour convenir à : <language> 212
- messages d'erreur : la chaîne d'inscription doit comporter 20 lettres 209
- messages d'erreur : la fonction n'est pas encore disponible 212
- messages d'erreur : la fonction n'est pas encore prête 212

- messages d'erreur : le fichier comporte des "trous" 212
- messages d'erreur : le fichier de base de données KeyWords est défectueux 212
- messages d'erreur : le fichier de concordance n'est pas valide 213
- messages d'erreur : le fichier KeyWords est défectueux 213
- messages d'erreur : le fichier ne contient aucune des balises spécifiées 213
- messages d'erreur : lecteur non valide 213
- messages d'erreur : les dispersions vont de paire avec les concordances 214
- messages d'erreur : les liens entre les outils sont défaillants 214
- messages d'erreur : les noms de fichiers doivent être différents 214
- messages d'erreur : limite de regroupement atteinte 214
- messages d'erreur : limite de termes de recherche des fichiers atteinte 214
- messages d'erreur : liste de correspondance 205
- messages d'erreur : mémoire insuffisante 216
- messages d'erreur : nom de fichier non valide 216
- messages d'erreur : nombre non valide 216
- messages d'erreur : noms complets du lecteur et du dossier nécessaires 217
- messages d'erreur : pas un fichier WordSmith 217
- messages d'erreur : problème de disque – le fichier n'a pas été enregistré 218
- messages d'erreur : statistiques non disponibles 218
- mettre la première ligne entre parenthèses 147
- Microsoft Word 176
- Minimal Pairs 144
- Minimal Pairs : affichage 145
- Minimal Pairs : critères 144
- Minimal Pairs : exécution du programme 146
- Minimal Pairs : fonction 144
- Minimal Pairs : présentation générale 6
- Minimal Pairs : règles et paramètres 145
- Minimal Pairs : sélection des fichiers 145
- mise à jour de votre version 180
- mise en page 40
- mise en place d'une session de formation 40
- mode édition 181
- mode saisie 181
- modification de la police 21
- modification des couleurs 26
- modifier des concordances 75
- modifier les paramètres 18
- modifier vos données 33
- mot : définition 179
- mot-clé clé 91
- mot-clé clé : définition 90
- mots uniques 192
- ## - N -
- nombre de mots 183
- nombre de phrases trop important 216
- nombre d'entrées de concordance 82
- nombres 53
- nombres : leur gestion 189
- noms de fichiers 189
- noms de fichiers longs 189
- noms de fichiers Windows 95 189
- note Z 117
- notes 18
- nouveautés 4
- nouveautés de la version 4 4
- nouvel utilisateur 45
- numérotation des paragraphes 131
- numérotation des phrases 131
- numérotation : paragraphes 131
- numérotation : phrases 131
- ## - O -
- option À propos 183
- options de menus 185
- outils pour le repérage de schémas 193
- Oxford University Press. 180
- ## - P -
- p valeur 93
- pages de code 174
- paragraphe : début et fin 168
- paragraphes (spécifier) 124
- paragraphes : définition 179
- paramètres avancés 36
- paramètres avancés de concordance 70
- paramètres de concordance 70
- paramètres de Text Converter 153

paramètres d'impression 190
 paramètres par défaut 48
 paramètres personnalisés 48
 paramètres personnalisés pour les balises du BNC 163
 paramètres réseau 15
 paramètres réseau par défaut 48
 parcourir 131
 parcourir l'original 68
 pas un fichier WordSmith actuel 217
 personnalisation des menus contextuels 36
 pertinence négative 90
 pertinence : définition 90
 phrase : début et fin 168
 phrases (spécifier) 124
 phrases : définition 179
 police grecque 21
 police russe 21
 polices 21
 Portugais 23
 potato-peeling machine 200
 pourquoi cela ne fonctionne pas 198
 pourquoi la recherche a-t-elle échoué ? 217
 première utilisation de WordSmith 45
 premiers pas 2
 premiers pas avec Concord 11
 premiers pas avec KeyWords 12
 premiers pas avec WordList 13
 Présentation générale de WordList 5
 presse-papiers 176
 prix 180
 processus défini par l'utilisateur 33

- Q -

qu'est-ce qu'une concordance ? 58
 quitter 45

- R -

raccourcis 188
 rang 105, 106
 ratio type/token standard ou moyen 122
 ratios type/token 122
 recherche d'un mot 50
 recherche d'un mot ou d'une partie d'un mot 50

recherche en mode saisie 50
 recherche limitée 74
 recherche par mot entier 81
 recherche par saisie 50, 181
 rechercher avec le menu 114
 rechercher des entrées 114
 rechercher des fichiers pertinents 39
 rechercher des textes source 181
 rechercher et remplacer 50
 récupérer une sélection de textes favoris 22
 réduire les données à N entrées 46
 références d'entité 160
 registry 170
 règle 94
 regroupement automatique de lemmes 100
 regrouper des entrées 107
 regrouper des fichiers texte 149
 regroupeur 149
 remerciements 171
 remplacement 153
 renommer plusieurs fichiers 152
 repérer les types d'entrées 114
 répertoires 182
 réseaux : paramètres par défaut 48
 résolution des problèmes 198
 résolution des problèmes : apostrophes introuvables 198
 résolution des problèmes : couleurs illisibles 200
 résolution des problèmes : en panne 199
 résolution des problèmes : espace entre les colonnes 198
 résolution des problèmes : guillemets arrondis 198
 résolution des problèmes : guillemets spécifiques 198
 résolution des problèmes : Incohérence entre Concord et WordList 199
 résolution des problèmes : le processus est lent 201
 résolution des problèmes : les touches ne répondent pas 200
 résolution des problèmes : limite de la version de démonstration 199
 résolution des problèmes : l'imprimante n'imprime pas 200
 résolution des problèmes : ne démarre pas 201
 résolution des problèmes : points d'interrogation introuvables 198

résolution des problèmes : problème de balises
 Concord 198
 résolution des problèmes : symboles accentués
 199
 résolution des problèmes : symboles étranges 199
 résolution des problèmes : Viewer 134
 résolution des problèmes : WordList ne fonctionne
 pas 201
 restaurer le dernier fichier 183, 192
 re-trier 56
 re-trier des listes de mots 38
 re-trier une liste de mots 124
 re-trier : Concord 79
 re-trier : co-occurents 80
 re-trier : KeyWords 95
 re-trier : listes de cohérence 119
 re-trier : schéma de dispersion 80
 Russe 23

- S -

schéma de dispersion : tri 80
 schéma global 189
 schémas de collocation 79
 schémas des mots 79
 schémas et liens 92
 schématisation des mots-clés 93
 scores d'informations mutuelles 115
 section : début et fin 168
 segments de texte dans Concord 80
 Sélecteur de langue : autres langues 143
 Sélecteur de langue : enregistrement les paramètres
 144
 Sélecteur de langue : langue 141
 Sélecteur de langue : police 142
 Sélecteur de langue : présentation générale 140
 Sélecteur de langue : tri 142
 sélection au sein de textes 165
 sélection de textes 23
 sélection des fichiers à partir d'une boîte de dialogue
 standard 26
 Sélection des langues : présentation générale 5
 sélection du lot 87
 sélection parmi plusieurs textes 164
 sélectionner vos fichiers 87
 sensibilité à la casse 81
 séparateur de fin de texte 147

séparateurs de mots 180
 slash 81
 Splitter 147
 Splitter : caractères de remplacement 148
 Splitter : noms de fichiers 148
 Splitter : présentation générale 6
 Splitter : symboles 148
 statistiques 120
 statistiques de la base de données 91
 statistiques de résumé 30
 statistiques d'une base de données 91
 statistiques IM 117
 statistiques MI3 117
 stockage 191
 stocker les fichiers texte 23
 stoplist.cod 157
 suppression aléatoire d'entrées 46
 supprimer des entrées 56
 symboles 174
 symboles de fin de texte 148
 syntagmes 192
 syntaxe du terme de recherche 81

- T -

taille du texte 183
 télécharger une nouvelle version 16
 terme de contexte 74
 termes de recherche ou syntagmes issus d'un fichier
 75
 Text Converter : astérisque 155
 Text Converter : caractères de remplacement 155
 Text Converter : déplacer si... 156
 Text Converter : exemple de fichier de conversion
 158
 Text Converter : fichier de conversion 157
 Text Converter : limites 183
 Text Converter : présentation générale 6
 Text Converter : syntaxe 155
 texte balisé 160
 texte de Dickens 23, 45
 textes favoris 22
 textes sources 181
 textes : plus de textes 23
 textes : sélection 23
 totaux des colonnes 28

traitement des traits d'union 183
 traitement par lots 18
 traitement personnalisé 33
 traits d'union 53
 travail scolaire ou universitaire 40
 tri des balises 76
 tri des nombres 79
 tri précis dans WordList 124
 tri : Concord 79
 tri : KeyWords 95
 tri : WordList 124
 troncature à xx mots 218
 trop de requêtes pour ignorer les groupes correspondants 218
 troubleshooting: pineapple-slicing 200
 trous dans le fichier 212
 type de caractères 40
 types de balises 160

- U -

updater.exe 180
 utilisation avec un ordinateur dont le nom est XXX 218
 utilisation de la mémoire 191
 utilisation des phrases et en-têtes dans le BNC 168
 utilisation étudiante 58
 utilisation pour la recherche 58
 utilisation pour l'enseignement 58

- V -

valeur de dispersion du schéma 189
 valeur du log de vraisemblance 117
 vérification de la version 16
 vérifier la version actuelle 16
 version 32 bits 171
 version CD-ROM : paramètres par défaut 48
 Version Checker : présentation générale 7
 version de démonstration 180
 version de WordSmith 194
 version échantillon du BNC 180
 version française 170
 version non compatible 217
 version réseau 15

Viewer 130
 Viewer : affichage des options 131
 Viewer : alignement de phrases 130
 Viewer : aspects techniques 133
 Viewer : balises 131
 Viewer : édition 131
 Viewer : limites 183
 Viewer : mots en surbrillance 131
 Viewer : non correspondances de traduction 133
 Viewer : paramètres 132
 Viewer : parcourir 131
 Viewer : phrases inhabituelles 134
 Viewer : présentation générale 8
 Viewer : regroupement des phrases 132
 vitesse 193
 voisinage 60

- W -

WebGetter : affichage 138
 WebGetter : limites 139
 WebGetter : paramètres 137
 WebGetter : présentation générale 9, 136
 Windows 2000 189
 Windows 98 189
 Windows NT 189
 Windows XP 189
 WordList : affichage de base 125
 WordList : astuces pour bien commencer 13
 WordList : but 100
 WordList : clusters 111
 WordList : créer à partir d'un fichier texte 120
 WordList : limites 183
 WordList : modifier des entrées 38
 WordList : paramètres minimum et maximum 124
 WordList : sensibilité à la casse 123
 WordList : tri 124, 201
 WordSmith est déjà en cours d'utilisation 218
 WordSmith Tools : installation 15
 WordSmith Tools : manuel 189
 wshell.exe (contrôleur) 4
 wshell.ini et réseaux 15

- X -

XX jours restants 218