

Oxford WordSmith Tools

Version 4.0

© 2004-2006 Mike Scott

Oxford WordSmith Tools

version 4.0

by Mike Scott

© 2004-2006 Mike Scott

WordSmith Tools

© 2004-2006 Mike Scott

All rights reserved. No parts of this work may be reproduced in any form or by any means - graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems - without the written permission of the publisher.

Products that are referred to in this document may be either trademarks and/or registered trademarks of the respective owners. The publisher and the author make no claim to these trademarks.

While every precaution has been taken in the preparation of this document, the publisher and the author assume no responsibility for errors or omissions, or for damages resulting from the use of information contained in this document or from the use of programs and source code that may accompany it. In no event shall the publisher and the author be liable for any loss of profit or any other commercial damage caused or alleged to have been caused directly or indirectly by this document.

Printed: March 2006

Publisher

Oxford University Press

Special thanks to:

All the people who contributed to this document by testing WordSmith Tools in its various incarnations. Especially those who reported problems and sent me suggestions.

Table of Contents

Foreword	1
Part I WordSmith Tools	3
Part II Übersicht	5
1 Was neu ist in Version 4	5
2 Controller	5
3 Concord	5
4 KeyWords	6
5 WordList	6
6 Utilities	6
Sprachen auswählen	6
Datei-Utilities	7
Minimal Pairs	7
Splitter	7
Text Converter	7
Version Checker	8
Viewer	9
WebGetter	10
Part III Erste Schritte	12
1 erste Schritte mit Concord	12
2 erste Schritte mit KeyWords	13
3 erste Schritte mit WordList	14
Part IV Installation und Aktualisierung	16
1 Oxford WordSmith Tools installieren	16
2 Netzwerk-StandardEinstellungen	16
3 Version überprüfen	17
Part V Controller	19
1 Einstellungen anpassen	19
2 Akzentzeichen	19
3 Notizen hinzufügen	19
4 Batch-Verarbeitung	19
5 Batch-Verzeichnisse	22
6 Schriftart verändern	22
7 bevorzugte Texte (Favouriten) auswählen	23
8 Sprache auswählen	24
9 Texte auswählen	24

10	Dateien aus Standard-Dialogfeld auswählen	27
11	Farben	27
12	Spaltensummen	29
13	neue Datenspalte berechnen	30
14	Datenfrequenzen zählen	31
15	Ergebnisse kopieren	33
16	benutzerdefinierte Verarbeitung	33
17	Popupmenüs anpassen	37
18	Datenliste bearbeiten	38
19	relevante Dateien finden	40
20	Unterrichtsanweisungen	40
21	Layout & Format	41
22	Wörter in Liste übereinstimmen	42
23	Drucken und Druckvorschau	44
24	erstmalige Verwendung von WordSmith	45
25	vorherige Listen	45
26	WordSmith beenden	45
27	Daten auf N Einträge reduzieren	46
28	als Text speichern	46
29	Standardeinstellungen speichern	48
30	Ergebnisse speichern	49
31	suchen und ersetzen	50
32	suchen durch Eingabe	50
33	nach einem Wort bzw. Wortteil suchen	50
34	Dateinamen anzeigen	51
35	Stopplisten	51
36	Prozess unterbrechen	52
37	Text und Sprachen	53
38	Window Management	55
39	nicht gewollte Zeilen entfernen	56

Part VI Concord 58

1	Zweck	58
2	was ist eine Konkordanz	58
3	ausblenden	58
4	Kategorien	59
5	Kollokationsbereich	60
6	Kollokationseinstellungen	60
7	Kollokatoren-Anzeige	61
8	Verbindung zwischen Kollokatoren	62

9	Kollokation	63
10	Concord: Cluster	64
11	Concord: Verteilung	65
12	Concord: speichern und drucken	67
13	Concord: Anzeigoptionen	67
14	Concord: Audio und Video wiedergeben	67
15	Concord: Anzeige und Funktionen	68
16	Konkordanzeinstellungen	70
17	Konkordanz mit Tags	72
18	Kontextwort	73
19	Konkordanzen bearbeiten	74
20	dateibasierte Suchwörter	74
21	nächstes Tag	75
22	Muster	78
23	neu sortieren	78
24	neu sortieren: Kollokatoren	79
25	neu sortieren: Verteilungsplot	79
26	Textsegmente in Concord	79
27	Suchwort-Syntax	80
28	WordSmith Controller: Concord: Einstellungen	81
Part VII	KeyWords	85
1	Zweck	85
2	Analyse von zwei Wortlisten	85
3	Definition assoziierte Wörter	85
4	assoziierte Wörter	86
5	Dateien auswählen	86
6	assoziierte Gruppen	86
7	KeyWords: Cluster	87
8	Konkordanz	88
9	Erstellung einer Datenbank	88
10	Beispiel Schlüsselwörter	89
11	Definition eines gewichteten Schlüsselworts	89
12	Definition der Gewichtung	89
13	KeyWords-Datenbank	90
14	KeyWords: Tipps	90
15	KeyWords: Berechnung	91
16	KeyWords: Links	91
17	Wortliste aus KeyWords-Daten erstellen	92
18	P-Wert	92

19 Plot-Berechnung	92
20 Plotanzeige	93
21 assoziierte Schlüsselwortgruppen neu gruppieren	94
22 neu sortieren: KeyWords	95
23 Anzeige der Schlüsselwörter	95
24 WordSmith Controller: KeyWords-Einstellungen	96

Part VIII WordList 99

1 Zweck	99
2 Lemmata automatisch zusammenführen	99
3 Lemmadatei auswählen	100
4 Wortlisten vergleichen	101
5 Wortlisten zusammenführen	102
6 Vergleichsanzeige	102
7 Konsistenzanalyse (detailliert)	104
8 Konsistenzanalyse (einfach)	105
9 Lemmata	106
10 Indexlisten: Verwendungen	107
11 Indexlisten: anzeigen	108
12 einen WordList-Index erstellen	109
13 Indexcluster	110
14 Menüsuche	113
15 Transinformationswerte	114
16 Transinformation: Berechnen	116
17 Transinformation: Anzeige	117
18 neu sortieren: Konsistenzliste	119
19 Statistik	119
20 Wörter aus Textliste importieren	120
21 Type-Token-Relationen	122
22 Groß- und Kleinschreibung beachten	123
23 Einstellungen für Minimum und Maximum	124
24 Sortierreihenfolge	124
25 WordList-Anzeige	125
26 WordSmith Controller: WordList-Einstellungen	127

Part IX Viewer und Aligner 130

1 Zweck	130
2 Alignment mit Viewer	130
3 Alignment und Verschieben	131
4 bearbeiten	131

5 Sätze und Absätze nummerieren	131
6 Optionen	132
7 Sätze verbinden und trennen	132
8 Einstellungen	132
9 technische Aspekte	133
10 falsch übereingestimmte Übersetzungen	133
11 Fehlerbehebung	134
12 ungewöhnliche Sätze	134

Part X Utility-Programme 136

1 Daten aus früheren Versionen konvertieren	136
Daten aus früheren Versionen konvertieren	136
2 WebGetter	136
Übersicht	136
Einstellungen	137
Anzeige	138
Einschränkungen	139
3 Languages Chooser	140
Übersicht	140
Sprache	141
Schriftart	142
Sortierreihenfolge	142
Andere Sprachen	143
Auswahl speichern	144
4 Minimal Pairs	144
Zweck	144
Anforderungen	144
Dateien auswählen	144
Output	145
Regeln und Einstellungen	145
Programm verwenden	146
5 Datei-Utilities	147
Index	147
Splitter	147
Zweck von Splitter.....	147
Splitter: Dateinamen.....	148
Splitter: Wildcards.....	148
Textdateien zusammenführen	149
zwei Dateien vergleichen	150
Datei-Chunker	151
Duplikate finden	151
umbenennen	152
6 Text Converter	152
Zweck	152
Text Converter: Einstellungen	153
Text Converter: Syntax	154
Text Converter: verschieben, wenn	156
Konvertierungsdatei für Text Converter	156
Text Converter: Vorlage Konvertierungsdatei	157

Part XI Tags und Markup 160

1 Übersicht	160
2 Tag-Arten	160
3 Umgang mit Tags	161
4 Multimedia-Tags	162
5 Tags als Selektoren	163
6 Texte nach Inhalt filtern	164
7 innerhalb von Texten auswählen	165
8 eine Tagdatei erstellen	166
9 Anfang und Ende von Textsegmenten	168

Part XII Referenz 170

1 32-Bit-Version	170
2 Bibliografie	170
3 Bindestriche	171
4 Bugs	171
5 Danksagung	172
6 Datumsformat	173
7 Definitionen	173
Definitionen	173
Worttrennzeichen	173
8 Demoversion	174
9 Druckereinstellungen	174
10 Einschränkungen	175
11 einzelne Wörter vs. Cluster	175
12 file types	176
13 Geschwindigkeit	176
14 Handbuch für WordSmith Tools	177
15 HTML, SGML und XML	177
16 international versions	178
17 Internet	179
18 Internet	179
19 Internet	179
20 Kontaktadressen	179
21 lange Dateinamen	180
22 letzte Datei wiederherstellen	180
23 Menü- und Schaltflächenoptionen	180
24 Modus "Bearbeiten" vs. Modus "Eingabe"	183
25 Plot-Verteilungswert	183
26 Quelltexte finden	184

27	RAM-Auslastung	184
28	Referenzkorpus	185
29	Schnittstellen zwischen Tools	185
30	Statusleiste	186
31	Systemanforderungen	186
32	Tastenkombinationen	186
33	Tools zum Erkennen von Mustern	187
34	Tool-spezifische Einschränkungen	188
35	Verlauf	189
36	Versionsinformation	189
37	Verzeichnisse\Laufwerke	190
38	Zahlen	191
39	Zeichensätze	191
	Übersicht	191
	Akzentzeichen und Symbole	192
	ANSI und ASCII	192
	DOS	192
	Windows	192
	Unicode	193
40	Zip-Dateien	193
41	Zwischenablage	194

Part XIII Fehlerbehebung 198

1	abgestürzt	198
2	Apostrophe nicht gefunden	198
3	Demolimit	198
4	Drucker druckt nicht	198
5	eigenartige Symbole	198
6	Ergebnisse für Concord und WordList unterschiedlich	199
7	Liste häufig gestellter Fragen	199
8	pineapple-slicing	200
9	Spaltenabstand	200
10	startet nicht	200
11	Tagproblem in Concord	200
12	Tasten reagieren nicht	200
13	unleserliche Farben	201
14	Wortliste nicht sortiert	201
15	zu langsam	201

Part XIV Fehlermeldungen 203

1	.ini-Datei nicht gefunden	203
2	anderes Tool kann nicht aufgerufen werden	203

3 Arbeitsspeicher ist knapp	203
4 auf die Datei konnte nicht zugegriffen werden	203
5 auf Textdatei kann nicht zugegriffen werden – wird sie anderweitig verwendet?	203
6 Basislistenfehler	203
7 Bestätigungsmeldungen: Datei neu einlesen?	203
8 bricht bei xx Wörtern ab -- Tagdatei hat mehr!	204
9 can't merge list with itself!	204
10 Datei enthält keines der festgelegten Tags	204
11 Datei hat "Löcher"	204
12 Datei konnte nicht gelesen werden	204
13 Datei konnte nicht gelesen werden	204
14 Datei konnte nicht gespeichert werden	205
15 Datei nicht gefunden	205
16 Dateinamen dürfen nicht identisch sein!	205
17 Datenträgerproblem -- Datei nicht gespeichert	205
18 Details für Matchliste nicht angegeben.	205
19 dies ist eine Demoversion	205
20 diese Funktion ist noch nicht bereit	206
21 diese Meldung nicht mehr anzeigen... ..	206
22 Drucker notwendig	206
23 Format der Lizenznummer nicht korrekt	206
24 Formular unvollständig	206
25 Funktion kann noch nicht vollständig ausgeführt werden	206
26 Grenze bei Zusammenführung erreicht	206
27 Internetverbindung konnte nicht hergestellt werden	207
28 kein Platz für berechnete Variable	207
29 keine aktuelle WordSmith-Datei	207
30 Keine assoziierten Gruppen identifiziert	207
31 keine assoziierten Wörter gefunden	207
32 keine Cluster gefunden	207
33 keine Einträge in der KeyWords-Datenbank	207
34 keine gelöschten Zeilen zu entfernen	208
35 keine Kollokatoren gefunden	208
36 keine Konkordanzeinträge	208
37 keine Schlüsselwörter für Plot	208
38 keine Schlüsselwörter gefunden	208
39 keine solchen Dateien gefunden	208
40 keine Statistik verfügbar	209
41 keine WordSmith-Datei	209

42	keine Wörter aus KeyWords-Stoppliste	209
43	keine Wörter aus Konkordanz-Stoppliste	209
44	keine Wörter aus Lemmaliste	209
45	keine Wörter aus Matchliste	209
46	keine Wörter aus Stoppliste	209
47	keine Wörter aus Tagliste	210
48	keine Wortlisten ausgewählt	210
49	KeyWords-Datei ist fehlerhaft	210
50	KeyWords-Datenbank ist fehlerhaft	210
51	Konkordanzdatei ist fehlerhaft	210
52	Konkordanz-Stopplistendatei nicht gefunden	210
53	Konvertierungsdatei nicht gefunden	211
54	Laufwerk ungültig	211
55	Limit für dateibasierte Suchwörter erreicht	211
56	Liste der Fehlermeldungen	211
57	Lizenznummer für Netzwerk wird bereits verwendet	213
58	Lizenznummer ist nicht korrekt	213
59	muss eine Zahl sein	213
60	neues Verzeichnis mit diesem Namen konnte nicht erstellt werden	213
61	nichts aktiviert	213
62	nur für die Nutzung am PC mit dem Namen XXX	213
63	Programm funktioniert nur unter Windows 98 oder höher	213
64	Quelldatei(en) nicht gefunden	213
65	Schlüsselwortdatenbanken konnten nicht zusammengeführt werden.	214
66	Schlüsselwörter können wegen unterschiedlicher Sprachen nicht berechnet werden	214
67	Schnittstelle zwischen den Tools unterbrochen	214
68	Stopplistendatei nicht eingelesen	214
69	Stopplistendatei nicht gefunden	214
70	Tagdatei nicht eingelesen	214
71	Tagdatei nicht gefunden	215
72	Transinformation inkompatibel	215
73	ungültige KeyWords-Berechnung	215
74	ungültige KeyWords-Datenbank	215
75	ungültige Konkordanzdatei	215
76	ungültige Nummer	216
77	ungültige WordList-Datei	216
78	ungültige WordList-Vergleichsdatei	216
79	ungültiger Dateiname	216
80	Verteilung wird mit Konkordanz gespeichert	216

81	Verzeichnis kann nicht erstellt werden, da bereits ein Verzeichnis mit diesem Namen vorhanden ist	
82	vollständiger Pfadname (Laufwerk:\Verzeichnis) wird benötigt	217
83	Wieso ist die Suche fehlgeschlagen?	217
84	WordSmith Tools abgelaufen	217
85	WordSmith wird bereits ausgeführt	217
86	WordSmith-Versionen stimmen nicht überein	217
87	Wörter können nur als ASCII gespeichert werden	217
88	Wortlistendatei ist fehlerhaft	217
89	Wortlistendatei nicht gefunden	218
90	Wortlisten-Vergleichsdatei ist fehlerhaft	218
91	XX Tage verbleiben	218
92	Zeichensatz an Sprache (<language>) anpassen und auf <x> zurücksetzen.	218
93	Zielverzeichnis nicht gefunden	218
94	zu viele Sätze	218
95	zu viele übereinstimmende Gruppen zu ignorieren	218
96	Zugriff auf die originale Textdatei nicht möglich, da nicht vorhanden	219
97	zwei Dateien notwendig	219
	Index	220

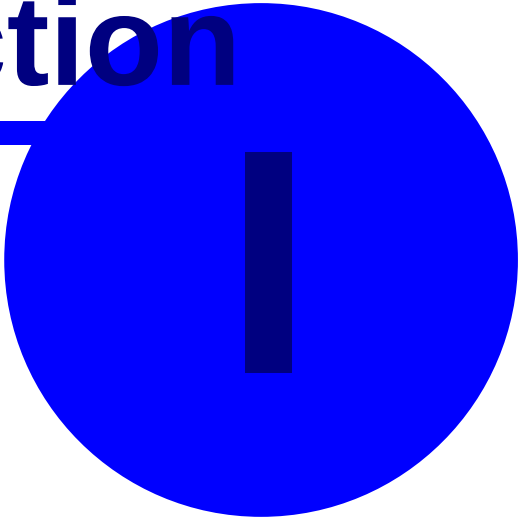
Foreword

This is just another title page
placed between table of contents
and topics

WordSmith Tools

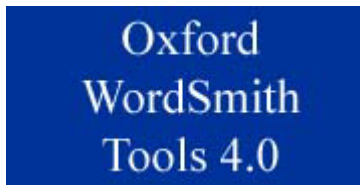
WordSmith Tools

Section



I

1 WordSmith Tools



Oxford WordSmith Tools ist eine Gruppe integrierter Programme, die zur Untersuchung von Texten und den darin vorkommenden Wörtern dienen. Sie können mit Oxford WordSmith Tools herausfinden, wie Wörter in Texten verwendet werden, sei es in Ihren eigenen Texten oder in denen anderer.

Das Tool **WordList** erstellt eine Liste aller Wörter oder Wortgruppen (Cluster), die in einem Text vorkommen. Diese Liste wird alphabetisch und entsprechend der Wortfrequenz geordnet. Das Konkordanzprogramm **Concord** erstellt Konkordanzen, in denen jedes Wort bzw. jede Phrase mit seinem/ihrer Kontext angezeigt wird. So finden Sie heraus, welche Wörter häufig gemeinsam auftreten. Das Tool **KeyWords** erstellt eine Liste der Schlüsselwörter eines Texts. Diese Tools werden von der Oxford University Press für eigene lexikografische Arbeiten bei der Erstellung von Wörterbüchern verwendet. Sprachlehrer und -schüler und Sprachforscher in vielen Ländern weltweit benutzen dieses Programm für die Untersuchung von Sprachmustern vieler verschiedener Sprachen.

Wie bekomme ich Hilfe?

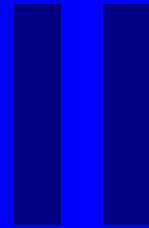
[Wenn Sie eine Internetverbindung haben, klicken Sie hier für eine schrittweise WordSmith-Einführung mit Screenshots.](#)

Die meisten Menüs und Dialogfelder verfügen über eine Hilfoption. Hilfe erhalten Sie oft, indem Sie F1 drücken, auf  klicken, oder indem Sie das Menü Hilfe aufrufen (in den meisten Menüs auf der rechten Seite). Innerhalb einer Hilfedatei (wie dieser) ist es manchmal am einfachsten, auf die Schaltfläche Suchen zu klicken und den angezeigten Index zu studieren, oder Sie "blättern" in den Hilfethemen.

Klicken Sie hier, wenn Sie gleich mit [WordList](#), [Concord](#) oder [KeyWords](#) beginnen möchten.

Übersicht

Section



2 Übersicht

2.1 Was neu ist in Version 4

Version 4 wurde vollkommen neu geschrieben.

Neue Funktionen:

- [schrittweise Online-Einführung mit Screenshots, in der die Erstellung von Wortlisten, Konkordanzen, etc. erklärt wird \(nur auf Englisch verfügbar\)](#)
- [komplette Online-Version dieser Hilfedatei](#)
- [praktisch unbegrenzte](#) Konkordanzen und Wortlisten
- verbesserter Umgang mit [Clustern](#) in Wortlisten
- variable Größe bei Concord-[Clustern](#)
- Concord zeigt, an welcher Stelle in [Satz, Absatz, Überschrift und Abschnitt](#) das gefundene Wort vorkommt
- [Unicode](#)-Bearbeitung von Text; ermöglicht eine viel höhere Anzahl an [Sprachen](#)
- eine Konkordanz erstellen und Wiedergabe der zugehörigen [Audiodatei](#)
- verbessertes [Tag-Handling](#)
- Aufbau Ihres eigenen [Textkorpus aus dem Internet](#)
- verbesserte Statistikfunktionen für [Kollokation](#)
- verbesserte Lemmatisierung durch die Verwendung [eigener .dll-Dateien](#)
- Verwendung von [.zip-Dateien](#)
- mehr Utility-Tools
- [Version](#)sprüfung
- mehr Sprachen in [Aligner](#)

2.2 Controller



Dieses Programm steuert die Tools. Es zeigt aktuelle Standardeinstellungen an, ermöglicht das Verändern dieser Einstellungen, das Auswählen von Textdateien und von dort werden die verschiedenen Tools aufgerufen.

Der Controller erscheint oben links auf dem Bildschirm.

Sie können ihn minimieren, wenn der Bildschirm [unübersichtlich](#) wird.

Wenn Sie eine schrittweise Einführung mit Screenshots machen möchten, klicken Sie [hier um die WordSmith-Website zu besuchen](#).

2.3 Concord



Concord ist ein Programm, das aus [DOS](#)-, [Nur-Text](#)-, [ASCII](#)- oder [ANSI](#)-Textdateien eine [Konkordanz](#) erstellt.

Um dieses Programm zu verwenden, müssen Sie ein Suchwort angeben, nach dem Concord dann alle ausgewählte Textdateien durchsucht. Die Ergebnisse dieser Suche werden in einer Konkordanzanzeige präsentiert und bieten Informationen über die Kollokatoren des Suchworts. Listen können für den späteren Gebrauch [gespeichert](#), bearbeitet, ausgedruckt, in Ihr Textverarbeitungsprogramm kopiert oder als Textdateien gespeichert werden.

Siehe auch: Inhalt der Concord-Hilfe, [Schaltflächen](#)

2.4 KeyWords



Dieses Programm dient dazu, Schlüsselwörter in einem bestimmten Text zu finden und zu identifizieren. Dazu werden die Wörter in einem Text mit einer Referenzgruppe von Wörtern verglichen, die üblicherweise aus einem großen Textkorpus stammen. Jedes Word, das in Bezug auf seine Frequenz in einem Text ungewöhnlich ist, ist ein "Schlüsselwort". Die Schlüsselwörter danach geordnet, wie ungewöhnlich ihre Frequenz ist.

Die Verteilung der Schlüsselwörter kann in einem [Plot](#) grafisch dargestellt werden.

Listen können für den späteren Gebrauch [gespeichert](#), bearbeitet, ausgedruckt, in Ihr Textverarbeitungsprogramm kopiert oder als Textdateien gespeichert werden.

Dieses Programm benötigt zwei oder mehr Wortlisten, die zuerst mit dem [WordList-Tool](#) erstellt werden müssen.

Siehe auch: Inhalt der KeyWords-Hilfe, [Schaltflächen](#)

2.5 WordList



Dieses Programm generiert Wortlisten basierend auf einer oder mehreren [ANSI](#)- oder [ASCII](#)-Textdateien. Die Anzeige der Wortlisten ist sowohl alphabetisch als auch nach Frequenz sortiert. Wortlisten können für den späteren Gebrauch [gespeichert](#), bearbeitet, ausgedruckt, in Ihr Textverarbeitungsprogramm kopiert oder als Textdateien gespeichert werden.

Siehe auch: Inhalt der WordList-Hilfe, [Schaltflächen](#)

2.6 Utilities

2.6.1 Sprachen auswählen



A tool for selecting Languages which you want to process.
You will probably only need to do this once, when you first use Oxford WordSmith Tools.

See also: [Choose Language Tool](#)

2.6.2 Datei-Utilities



Programme, die

- [zwei Dateien vergleichen](#)
- [große Dateien in Chunks aufteilen](#)
- [Duplikate finden](#)
- mehrere Dateien [umbenennen](#)
- "[Löcher](#)" in Texten finden
- große Dateien in ihre Komponenten splitten
- viele kleine Textdateien zu größeren Dateien [zusammenführen](#)

2.6.3 Minimal Pairs



Dieses Programm dient dazu, Tippfehler und Wortpaare mit minimalen Unterschieden zu finden.

Siehe auch: [Zweck](#), [Anforderungen](#), [Dateien auswählen](#), [Output](#), [Regeln und Einstellungen](#), [Programm verwenden](#).

2.6.4 Splitter



Splitter ist ein Utility das große Dateien in kleinere unterteilt, um eine Textanalyse durchführen zu können. Sie können ein Symbol angeben, mit dem das Textende markiert wird (z.B. **</Text>**); Splitter wird eine große Datei durchsuchen, den Text kopieren und jedes Mal, wenn ein solches Symbol gefunden wird, eine neue Textdatei beginnen.

Siehe auch: Inhalt der Splitter-Hilfe, [Schaltflächen](#)

2.6.5 Text Converter



Text Converter ist ein universelles Utility, das Sie hauptsächlich für die folgenden Aufgaben verwenden können: Texte bearbeiten, Textdateien umbenennen, Dateierweiterungen ändern und Dateien in ein neues Verzeichnis verschieben, wenn sie bestimmte Wörter oder Phrasen enthalten.

Die wichtigste Verwendung ist das Ersetzen von Strings in Textdateien. Die Funktion "Suchen

und Ersetzen" ist ähnlich der eines Textverarbeitungsprogramms, nur dass mit Text Converter dieser Vorgang für viele Dateien ausgeführt werden kann, eine Datei nach der anderen. Es kann auch jede beliebige Anzahl von Strings ersetzt werden, nicht nur eine.

Das ist sehr nützlich, wenn Sie eine große Anzahl von Texten neu formatieren möchten, z. B. unnötige Leerstellen löschen, Absatzmarken in der Mitte eines Absatzes entfernen, Akzentzeichen ändern.

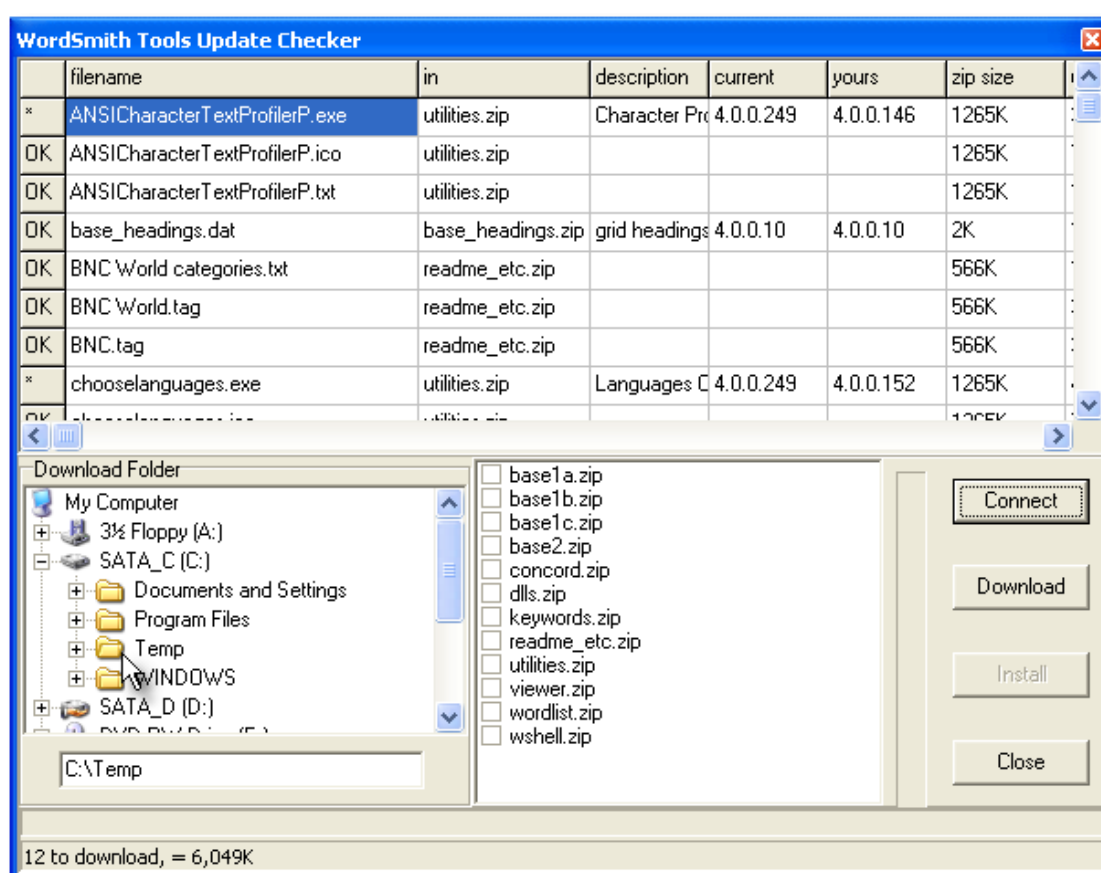
Siehe auch: Inhalt der Text-Converter-Hilfe, [Menü- und Schaltflächenoptionen](#)

2.6.6 Version Checker

4.0

A tool to check whether any components of your current version need updating and if so, download them for you. Accessed via the main Controller menu, *File | Web version check*.

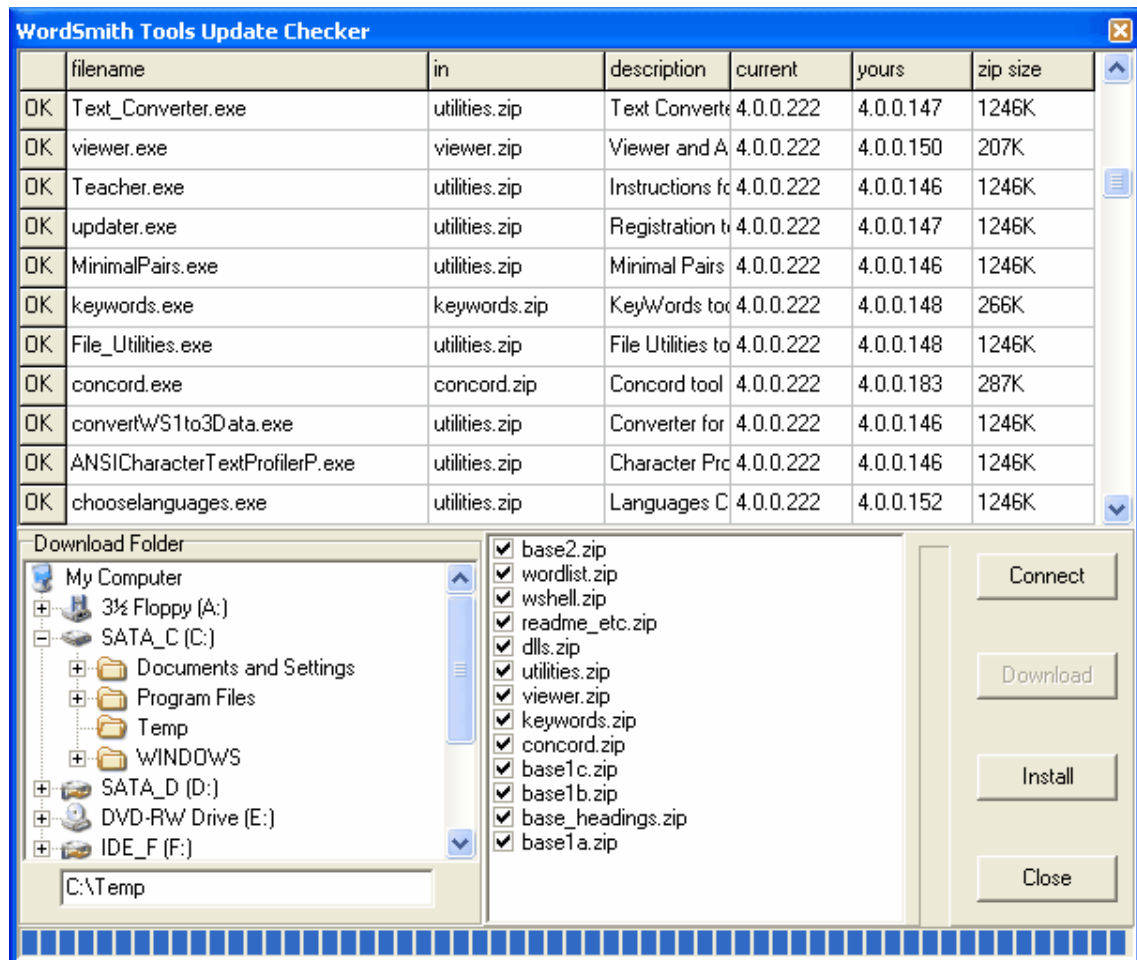
When you run the program, after pressing *Connect*, you see something like this:



The various components of WordSmith are listed in the top window and the current version is compared with your present situation. If they are different, all the files in the relevant zip file will be starred (*) in the left margin.

By default you will download to wherever WordSmith is already but you're free to choose

somewhere else as in the screenshot where **c:\temp** has been chosen. Press *Download* if you wish to get the updated files.



After the download, the various .zip files are checked (bottom right window) if downloaded successfully, and the Install button is now available for use. Install unzips all those which are checked.

2.6.7 Viewer



Viewer ist ein Utility, mit dem Sie Dateien in vielen Formaten untersuchen können. Es wird von anderen Tools aufgerufen, sobald Sie einen Quelltext ansehen möchten.

Viewer kann auch dazu verwendet werden, eine Kopie einer Textdatei zu erstellen, in der die [Sätze oder Absätze nummeriert](#) sind, oder um zwei Versionen eines Texts übereinzustimmen ([Alignment](#)), sodass abwechselnd entsprechende Absätze oder Sätze der beiden Dateien angezeigt werden.

Siehe auch: Inhalt der Viewer-Hilfe, [Schaltflächen](#)

2.6.8 WebGetter



Mit diesem Tool können Sie Texte im Internet suchen.

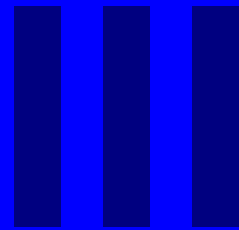
Wozu dient das?

Das dient dazu, Ihren eigenen Textkorporus aufzubauen, indem Sie Webseiten mithilfe einer Suchmaschine herunterladen.

Siehe auch: [Eine genauere Übersicht](#), [Einstellungen](#), [Anzeige](#), [Einschränkungen](#)

Erste Schritte

Section



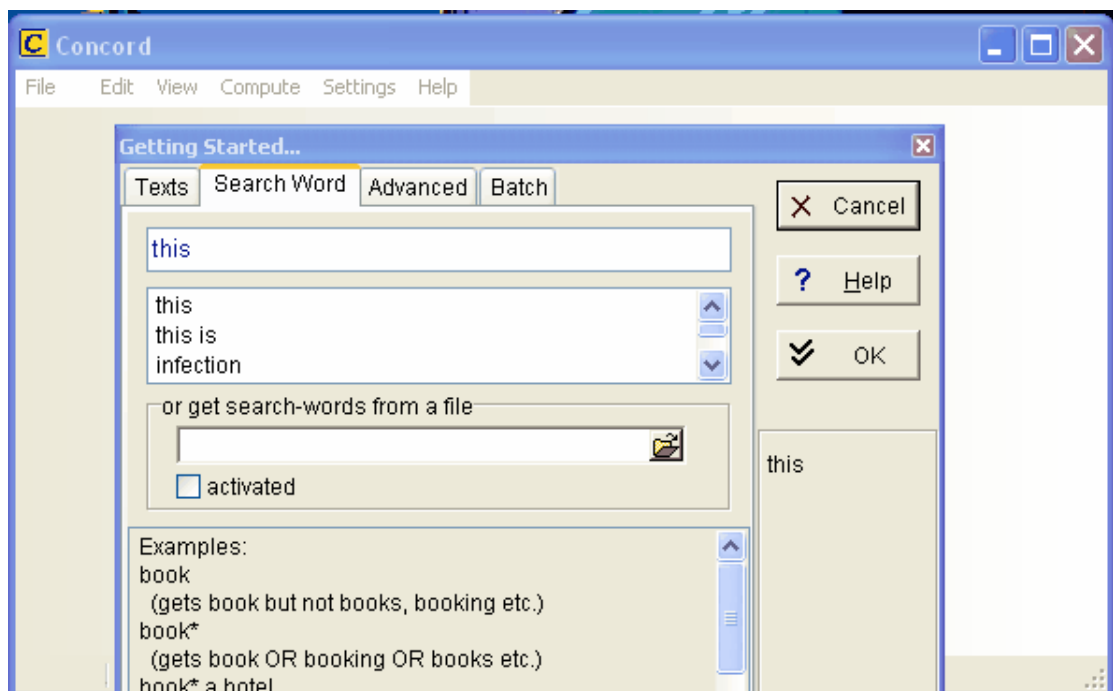
3 Erste Schritte

3.1 erste Schritte mit Concord

Wenn Sie eine schrittweise Einführung mit Screenshots machen möchten, klicken Sie [hier um die WordSmith-Website zu besuchen](#) (nur auf Englisch verfügbar).

Im Hauptfenster von Oxford WordSmith Tools (in der Titelleiste wird Oxford WordSmith Tools [Controller](#) angezeigt) klicken Sie auf die Schaltfläche "Concord". Auf die Schaltfläche klicken und das Tool Concord wird gestartet.

Nun sehen Sie ein Dialogfenster, in dem Sie [Ihre Texte auswählen](#) (oder diese Auswahl verändern) und dann eine neue Konkordanz erstellen können, die dann etwa folgendermaßen aussieht:

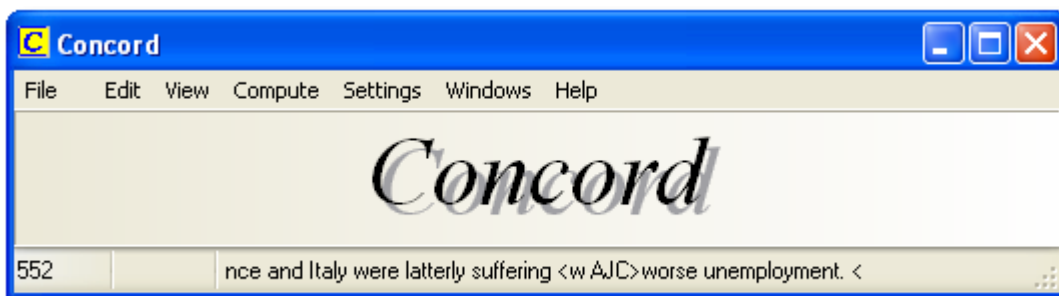


(Wenn Sie nur das Concord-Hauptfenster sehen, wählen Sie Datei | Neu () und das Fenster Erste Schritte wird geöffnet.)

Wenn Sie [WordSmith noch nie zuvor verwendet](#) haben, ist bereits automatisch ein Text ausgewählt, um Ihnen den Einstieg zu erleichtern.

Sie müssen jetzt ein [Suchwort oder eine Suchphrase](#) eingeben und dann auf OK klicken ().

Während Concord die Daten verarbeitet, sehen Sie eine Statusanzeige wie diese:



In diesem Fall wurden bereits 552 Einträge vorgenommen, und für den letzten Eintrag wird der Kontext des Suchworts angezeigt (das Suchwort ist in diesem Fall **worse**).

Wenn Sie andere Einstellungen ändern möchten, klicken Sie auf [Erweitert](#), aber wahrscheinlich können Sie die Standardeinstellungen beibehalten.

Concord durchsucht Ihren Text (bzw. Ihre Texte) nach dem Suchwort oder dem [Tag](#).

Vergessen Sie nicht, die Ergebnisse mit F2 oder  zu [speichern](#), wenn Sie die Konkordanz zu einem anderen Zeitpunkt wieder verwenden möchten.

Siehe auch: Inhalt der Concord-Hilfe.

3.2 erste Schritte mit KeyWords

Wenn Sie eine schrittweise Einführung mit Screenshots machen möchten, klicken Sie [hier um die WordSmith-Website zu besuchen](#) (nur auf Englisch verfügbar).

Im Hauptfenster von Oxford WordSmith Tools (in der Titelleiste wird Oxford WordSmith Tools [Controller](#) angezeigt) klicken Sie auf die Schaltfläche "KeyWords". Das Tool KeyWords wird gestartet.

Anfang

Sie sehen ein Dialogfeld, in dem Sie [Ihre Wortlisten auswählen](#) können. Sie müssen zwei Wortlisten auswählen, aus denen Sie eine Schlüsselwortliste erstellen möchten: eine davon basierend auf einem einzigen Text und die andere basierend auf mehreren Texten (genügend für einen guten Referenzkorpus zum Vergleich).

Sie sehen zwei Listen der Wortlistendateien in Ihrem aktuellen Wortlistenverzeichnis. Wenn dort keine sind, müssen Sie zu WordList zurückkehren und Wortlisten erstellen. Wählen Sie oben eine kleine Wortliste und unten eine Referenzkorpusliste, mit der Sie die kleinere Liste vergleichen werden. Haben Sie Ihre Texte ausgewählt, können Sie eine Schlüsselwortanalyse durchführen. Klicken Sie auf *Schlüsselwortliste jetzt erstellen*.

KeyWords beginnt nun mit der Verarbeitung Ihrer Datei und im Controller sehen Sie eine [Fortschrittsanzeige](#) mit einer Leiste, in der angezeigt wird, wie weit der Vorgang bereits fortgeschritten ist. Nachdem KeyWords diesen Vorgang abgeschlossen hat, wird eine Liste mit den Schlüsselwörtern angezeigt. Die Wörter am Anfang der Liste sind "wichtiger" als die weiter unten.

Vergessen Sie nicht, die Ergebnisse mit F2 oder  zu [speichern](#), wenn Sie die Schlüsselwortliste zu einem anderen Zeitpunkt wieder verwenden möchten.

Siehe auch: Inhalt der KeyWords-Hilfe, [Wozu dient das?](#)

3.3 erste Schritte mit WordList

Wenn Sie eine schrittweise Einführung mit Screenshots machen möchten, klicken Sie [hier um die WordSmith-Website zu besuchen](#) (nur auf Englisch verfügbar).

Es ist empfehlenswert damit zu beginnen, das Programm WordList auszuprobieren. Im Hauptfenster von Oxford WordSmith Tools (in der Titelleiste wird Oxford WordSmith Tools [Controller](#) angezeigt) wählen Sie die Option Tools, in der die Schaltfläche "WordList" angezeigt wird. Klicken Sie die Schaltfläche an, und WordList wird auf der rechten Seite Ihres Bildschirms geöffnet.

Anfang

Nun sehen Sie ein Dialogfenster, in dem Sie [Ihre Texte auswählen](#) (oder diese Auswahl verändern) und dann eine neue Wortliste erstellen können.

Wenn Sie [WordSmith noch nie zuvor verwendet](#) haben, ist bereits automatisch ein Text ausgewählt, um Ihnen den Einstieg zu erleichtern.

Es gibt weitere Einstellungen, die in diesem Menü verändert werden können, aber normalerweise können Sie diese Einstellungen ignorieren und sofort eine neue Wortliste erstellen, individuell oder als [Batch](#).

WordList beginnt nun mit der Verarbeitung Ihrer Datei(en) und im Controller sehen Sie eine [Fortschrittsanzeige](#) mit einer Leiste, in der angezeigt wird, wie weit der Vorgang bereits fortgeschritten ist. Nachdem WordList die Liste erstellt hat, sehen Sie zwei Fenster, in denen die Wörter Ihrer Textdatei in alphabetischer Reihenfolge und nach Frequenz geordnet angezeigt werden, und ein Fenster mit Statistik.

Vergessen Sie nicht, die Ergebnisse mit F2 oder  zu [speichern](#), wenn Sie die Wortliste zu einem anderen Zeitpunkt wieder verwenden möchten.

Siehe auch: Inhalt der WordList-Hilfe.

Installation und Aktualisierung

Section

IV

4 Installation und Aktualisierung

4.1 Oxford WordSmith Tools installieren

Dazu benötigen Sie knapp 20MB Speicherplatz auf Ihrer Festplatte, aber während des Installationsvorgangs benötigen Sie den doppelten Speicherplatz. (Generell wird es unter Windows bei *weniger* als 20MB freiem Speicherplatz zu Problemen kommen.)

1. Sie haben eine oder mehrere **.exe**-Dateien erhalten oder heruntergeladen. Speichern Sie die Dateien in einem temporären Verzeichnis, z. B. **c:\temp**. Es ist am einfachsten, wenn dieses Verzeichnis leer ist.
2. Öffnen Sie die Dateien. Dadurch werden alle Dateien, die für **Oxford WordSmith Tools** notwendig sind, im Verzeichnis Ihrer Wahl erstellt (standardmäßig **c:\wsmith4**).
3. Öffnen Sie die Datei **c:\wsmith4\wshell.exe** um zu beginnen. Sie werden gefragt, ob Sie die "Demoversion registrieren" möchten, und wenn Sie bereits eine Lizenznummer besitzen, können Sie diese jetzt eingeben. Ansonsten wird **WordSmith** als a [Demoversion](#) ausgeführt.
4. Wenn der Speicherplatz knapp ist, können Sie die **.exe**-Dateien in dem Verzeichnis **c:\temp** jetzt löschen.

Sie können eine aktuellere Version auf der [WordSmith Homepage](#) (nur auf Englisch verfügbar) erhalten.

Um das Programm zu deinstallieren, löschen Sie einfach alle Dateien in dem Verzeichnis **\wsmith4**.

Siehe auch: [Standardoptionen einstellen](#), [Kontaktadressen](#).

4.2 Netzwerk-StandardEinstellungen

Wenn Sie eine Site-Lizenz erworben haben, ist es viel einfacher eine Kopie von WordSmith auf einem Server zu installieren, auf den alle Benutzer zugreifen können. Normalerweise möchten Sie nicht, dass andere Benutzer Ergebnisse speichern oder die Originalkopie von WordSmith am Server verändern. Öffnen Sie **wshell.ini**. Darin sehen Sie einen Abschnitt, in dem Sie festlegen können, an welchem Ort die persönlichen Einstellungen jedes Benutzers gespeichert werden.

Die folgenden Bezeichnungen werden verwendet:

prohibited drives (nicht zulässige Laufwerke)

limited folder (begrenztes Verzeichnis)

instructions folder (Verzeichnis mit Anweisungen)

network-read/write folder (Netzwerkverzeichnis mit Lese- und Schreibzugriff)

ein Beispiel dafür wäre

[NETWORK]

network-read/write folder=m:\wsmith4

(Laufwerk M: wird deshalb als Netzwerkverzeichnis verwendet, da jeder Benutzer Schreibrechte auf M: hat.)

prohibited drives=xyz

(Auf den Laufwerken X: Y: und Z: sollen Benutzer nicht zugreifen, wenn Sie Texte auswählen.)

limited folder=v:\texts

(V:\TEXTS - und seine Unterverzeichnisse -- ist für alle Benutzer als Standardverzeichnis für

den Korpus auf dem Netzwerk festgelegt; jeder Benutzer kann aber auch auf alle anderen Verzeichnisse, zu denen er Schreibrechte hat, zugreifen.)

instructions folder=L:\English\WSmith instructions

(Wenn Sie die Software im Unterricht verwenden, legen Sie die Anweisungen in dieses Verzeichnis.)

Wenn ein neuer Benutzer WordSmith zum ersten Mal verwendet, erkennt WordSmith, dass eine Netzwerkversion verwendet wird und liest die oben erwähnten Informationen zu "net-work-read/write folder" ein. Danach wird WordSmith automatisch versuchen, das oben festgelegte Verzeichnis zu erstellen (theoretisch sollten Sie das nicht selbst machen müssen) und kopiert die verschiedenen anderen .ini-Dateien sowie andere Dateien mit Einstellungen von dem Netzwerkverzeichnis, in dem sich WordSmith befindet, in dieses Verzeichnis. Es ist empfehlenswert, Netzwerkverzeichnisse so einzurichten, dass die Benutzer dort auch tatsächlich Schreibrechte besitzen!

Siehe auch: [Unterrichtsanweisungen](#)

4.3 Version überprüfen

WordSmith beinhaltet eine Datei mit dem Namen **wordsmith_version_check.exe**, mit der Sie überprüfen können, ob Ihre Version aktuell ist, und wenn nicht, die notwendigen Updates und Patches herunterladen können. Um diese zu installieren, wird WordSmith beendet.

Siehe auch: [Versionsinformation](#), [Versionsupdate](#).

Controller

Section



V

5 Controller

5.1 Einstellungen anpassen

Das Hauptfenster "Einstellungen anpassen" im [Controller](#). Um dorthin zu kommen, wählen Sie im Controller im Menü *Einstellungen | Einstellungen anpassen...*

Dort können Sie folgende Einstellungen auswählen und [speichern](#):

- [Schriftart](#)
- [Farben](#)
- [Verzeichnisse](#)
- [Tags](#)
- [allgemeine Einstellungen](#)
- [Matchlisten](#)
- [Stopplisten](#)
- [Lemmalisten](#)
- [Text- und Spracheinstellungen](#)
- [Concord-Einstellungen](#)
- [KeyWords-Einstellungen](#)
- [WordList-Einstellungen](#)
- [erweiterte benutzerdefinierte Einstellungen](#)
- [Indexdateieinstellungen](#)

5.2 Akzentzeichen

This window shows the accented characters available for your currently-selected [language](#).

5.3 Notizen hinzufügen

So können Sie Notizen mit Ihren Daten [speichern](#).

Wenn Sie z. B. eine Konkordanz erstellt haben und diese nach eigenen [benutzerdefinierten Kategorien](#) sortiert haben, möchten Sie diese sicher auflisten und die Informationen für den späteren Gebrauch speichern.

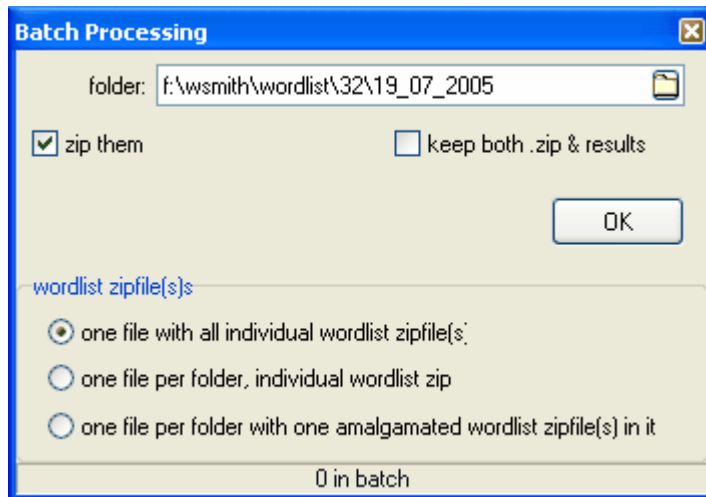
Wenn Sie diese Notizen außerhalb von **Oxford WordSmith Tools** aufrufen wollen, wählen Sie den Text aus, indem Sie die Umschalttaste und die Cursortasten drücken, kopieren Sie ihn mit *Strg+Einf* in die [Zwischenablage](#) und fügen Sie den Text in ein Textverarbeitungsprogramm ein (z. B. den **Windows Texteditor**).

5.4 Batch-Verarbeitung

Wozu dient das?

Verarbeitung in Gruppen wird dann verwendet, wenn Sie separate Listen erstellen möchten, aber dafür nicht jede Textdatei einzeln für die Wortliste oder Konkordanz auswählen, speichern, etc. möchten.

Wenn Sie mehr als eine Textdatei ausgewählt haben, können Sie diese Dateien in WordList, Concord und KeyWords als Batch verarbeiten.



Verzeichnis, in das die Dateien gespeichert werden

Als Name wird das aktuelle Datum vorgeschlagen. Sie können diesen Namen auch bearbeiten. Ein Verzeichnis mit dem Namen, den Sie ausgewählt haben, wird erstellt, wenn die Batch-Verarbeitung beginnt.

Die Ergebnisse werden in Verzeichnissen basierend auf diesem Namen gespeichert. Wenn Sie also beginnen, Wortlisten in dem Verzeichnis

`c:\wsmith\wordlist\05_07_19_12_00`, zu erstellen, sehen die Pfade folgendermaßen aus:

`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`

dann

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

etc.

Die Dateinamen bestehen aus dem Originaldateinamen mit der standardmäßigen Dateierweiterung (`.lst`, `.cnc`, `.kws`).

Dateien Zippen

Ist dieses Kontrollkästchen aktiviert, werden die Ergebnisse physisch in einer Standard-**Zip**-Datei gespeichert. Sie können die Dateien mit einem Standard-Zip-Tool wie Winzip extrahieren, oder WordSmith extrahiert die Dateien für Sie. Die in der Zip-Datei enthaltenen Dateien sind genau gleich wie die nicht komprimierten Dateien, sparen jedoch Speicherplatz auf der Festplatte. Außerdem befinden sich nicht mehr hunderte von Dateien in einem einzigen Verzeichnis, was günstig für das Betriebssystem ist.

Wenn Sie die Dateien zippen, erhalten Sie

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

und alle Unterdateien werden gelöscht, es sei denn Sie aktivieren das Kontrollkästchen ".zip & Ergebnisse behalten".

Eine Datei / Eine Datei pro Verzeichnis?

Die erste Einstellung (die Standardeinstellung) erstellt eine Zip-Datei, die alle einzelnen Wortlisten enthält. Pro Quelltest gibt es je eine Wortliste oder Konkordanz oder Schlüsselwortliste.

Wenn nun jedoch Ihre Textdateien wie folgt strukturiert sind:

`\..\BNC`

`\..\BNC\written`

`\..\BNC\written\humanities`

`\..\BNC\written\medicine`

\\.\BNC\written\science

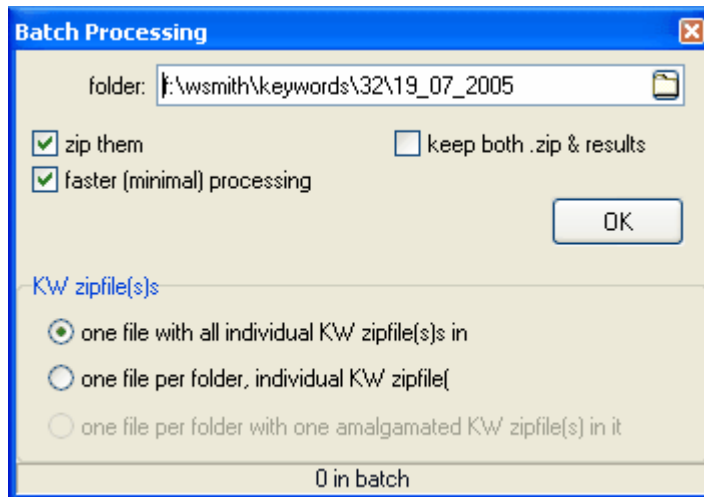
\\.\BNC\spoken

etc.

Die Option *Eine Datei pro Verzeichnis, einzelne Zip-Dateien* erstellt eine separate Zip-Datei für jedes Verzeichnis mit Textdateien (z. B. eine Datei für Geisteswissenschaften, ein weiteres für Medizin, etc.) mit einer Liste für jeden Quelltext.

Die Option *Eine Datei pro Verzeichnis, zusammengeführte Zip-Dateien* erstellt eine separate Zip-Datei für jedes Verzeichnis, aber nur eine Wortliste oder Konkordanz für das gesamte Verzeichnis mit Texten.

Schnelle (minimale) Verarbeitung



Dieses Kontrollkästchen ist nur dann aktiviert, wenn Sie gerade einen Vorgang begonnen haben, in dem mehr als ein Ergebnis gleichzeitig berechnet werden kann. Zum Beispiel werden nach jeder berechneten Konkordanz automatisch [Kollokatoren](#), [Muster](#) und [Verteilungsplots](#) berechnet. Auch in KeyWords gibt es [Verteilungsplots](#), [Linkberechnungen](#), etc., die berechnet werden, nachdem die Schlüsselwörter berechnet wurden.

Ist das Kontrollkästchen aktiviert, werden nur minimale Berechnung durchgeführt (Schlüsselwörter in KeyWords, Konkordanz in *Concord*). Das geht schneller, und Sie können die Plots auch später noch erstellen, solange die [Quelltexte](#) nicht verschoben oder gelöscht wurden.

Beispiel: Sie erstellen eine Wortliste und haben dazu 1.200 Textdateien aus einem Journal mit dem Namen "The Elephant" ausgewählt.

Sie legen

C:\WSMITH\WORDLIST\ELEPHANT als Verzeichnisname fest.

Wenn bereits ein Verzeichnis mit dem Namen C:\WSMITH\WORDLIST\ELEPHANT vorhanden ist, werden Sie gefragt, ob dieses Verzeichnis mit allen Unterverzeichnissen gelöscht werden kann.

Wenn Sie OK anklicken,

werden 1.200 neue Wortlisten mit den Namen ruesse1.LST, schwanz.LST, darm.LST, etc. erstellt. Sie befinden sich alle in [nummerierten Unterverzeichnissen](#) des Verzeichnisses

C:\WSMITH\WORDLIST\ELEPHANT.

Wenn Sie die Funktion "zu einer Zip-Datei zippen" nicht aktiviert haben, finden Sie die Dateien unter C:\WSMITH\WORDLIST\ELEPHANT\0.

Wenn Sie die Funktion "zu einer Zip-Datei zippen" aktiviert haben, wurde die Datei C:\WSMITH\WORDLIST\ELEPHANT.ZIP erstellt, die alle Ihre Ergebnisse enthält. (Die 1.200 .LST-Dateien werden gelöscht, sind aber in der zip-Datei noch vorhanden.)

Der Vorteil einer .zip-Datei ist, dass sie viel weniger Speicherplatz benötigt und per E-mail einfacher verschickt werden kann. WordSmith kann auf die Ergebnisse in einer .zip-Datei zugreifen, und Sie können die zu öffnende Wortliste, Konkordanz, etc. auswählen.

Ergebnisse mit WordSmith öffnen

Wählen Sie die Datei aus und öffnen Sie die Datei wie gewöhnlich, und ändern Sie dann den Dateityp zu "Batchdatei *.zip". Wenn Sie eine Zip-Datei auswählen, werden in einem Fenster deren Inhalte angezeigt. Durch Doppelklicken können Sie jede dieser Dateien öffnen.

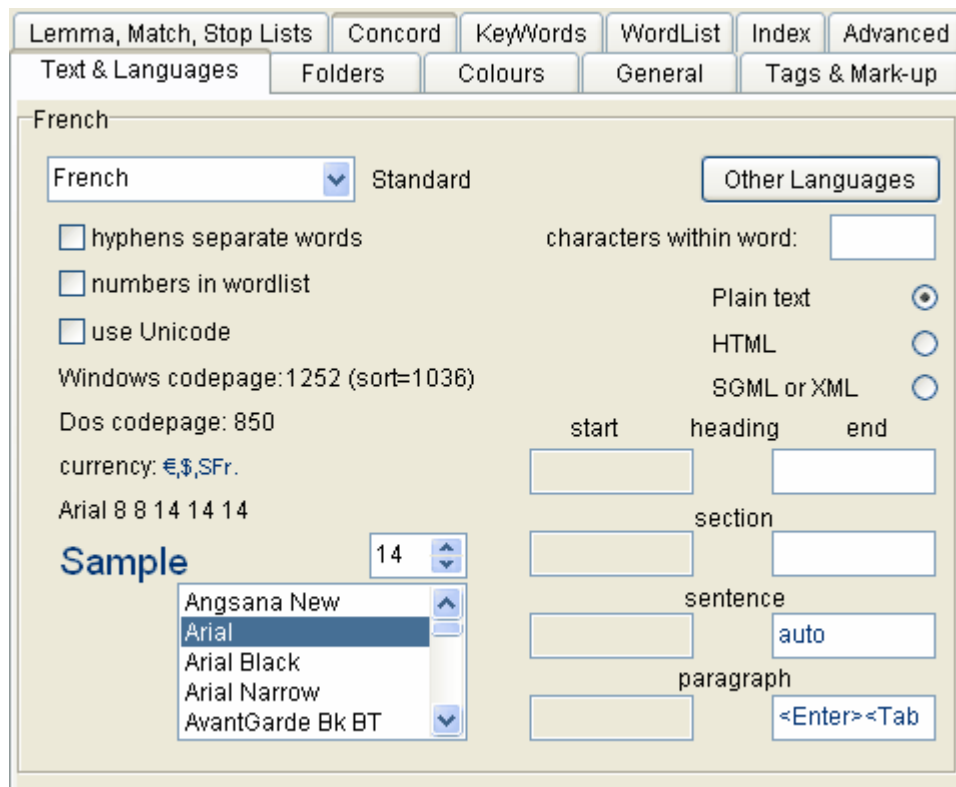
Anmerkung: Natürlich kann mit Concord nur eine Konkordanz bzw. mit KeyWords nur eine Schlüsselwortliste geöffnet werden. Wenn Sie eine Zip-Datei auswählen, die andere Dateien enthält, erscheint eine Fehlermeldung.

5.5 Batch-Verzeichnisse

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 Schriftart verändern

Schriftarteinstellungen können Sie im Menü Einstellungen in allen Tools sowie im Menü *Einstellungen | Einstellungen anpassen | Sprache* im [Controller](#) finden. Hier können Sie eine bevorzugte Windows-Schriftart und -Schriftgröße für das Anzeigefenster und für das [Drucken](#) in allen Oxford WordSmith Tools festlegen. Beachten Sie bitte, dass jede [Sprache](#) eine andere Standardschriftart haben kann.



Wenn Sie gerade in einem Tool Daten ansehen, wird die Schriftart automatisch geändert. Wenn Sie die Schriftart für ein spezielles Datenfenster nicht ändern wollen, weil Sie in verschiedenen Fenstern verschiedene Schriftarten oder Zeichensätze verwenden möchten, minimieren Sie zuerst diese Fenster.

Wenn Sie eine Datenspalte fett, kursiv, unterstrichen, etc. anzeigen wollen, verwenden Sie dazu die Option [Layout](#) (.

Oxford WordSmith Tools bietet Schriftarten an, die sich für die im Feld oben links ausgewählte [Sprache](#) eignen. Jede Sprache kann eine eigene Standardschriftart haben. Einstellungen zur Sprachauswahl, die gespeichert wurden, können unter `\wsmith4\language_choices.ini` angezeigt (und mit Sorgfalt auch verändert) werden.

5.7 bevorzugte Texte (Favouriten) auswählen

Favouriten speichern

Diese Funktion wird dazu verwendet, die aktuelle Textauswahl zu speichern. Sie ist besonders nützlich, wenn sich die Dateien z. B. in vielen verschiedenen Verzeichnissen befinden.

Damit wird eine Liste von Textdateien gespeichert, deren Status entweder nicht bekannt ist oder die Ihren Anforderungen entsprechen, wenn die [Dateien nach ihrem Inhalt ausgewählt](#) und alle Dateien, die den Anforderungen nicht entsprechen, ignoriert werden.

Favouriten laden

Mit dieser Funktion wird eine zuvor gespeicherte Auswahl vom Datenträger eingelesen.

Als Dateiname wird standardmäßig der [Name](#) des Tools angegeben, für das Sie die Texte auswählen, plus `recent_chosen_text_files.dat`. Die Datei wird in Ihrem WordSmith-

Hauptverzeichnis angelegt.

Sie können eine Standardtext-Datei verwenden, um eine Auswahl an Dateien zu laden () , die Sie mit dem Texteditor bearbeitet haben, aber beachten Sie bitte, dass für jede Datei der vollständige Pfadname angegeben werden muss. Wildcards können nicht verwendet werden.

Siehe auch: [Texte auswählen](#)

5.8 Sprache auswählen

Diese Einstellungen müssen Sie wahrscheinlich nur einmal vornehmen, und zwar dann, wenn Sie Oxford WordSmith Tools zum ersten Mal verwenden.

Wählen Sie die Sprache des Texts, den Sie analysieren möchten, im [Controller](#) unter Einstellungen | Einstellungen anpassen | Sprache aus. Sprache und [Zeichensatz](#) müssen kompatibel sein, so ist z.B. Deutsch kompatibel mit Windows Western (1252). Oxford WordSmith Tools kann mit vielen verschiedenen Sprachen, von Albanisch bis Ukrainisch, umgehen. Chinesisch, Japanisch, Arabisch, etc. werden in [Unicode](#) bearbeitet. Sie können die Wortlisten, Konkordanzen, etc. gleichzeitig in verschiedenen Sprachen anzeigen.

Wozu dient das?

Sprachen können sich in Bezug auf ihre Sortierreihenfolge deutlich unterscheiden. Spanisch beispielsweise wird folgendermaßen sortiert: 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. Akzentzeichen werden in manchen Sprachen standardmäßig gleichwertig wie ihre nicht akzentuierten Entsprechungen behandelt (z. B. auf Französisch: *donne*, *donné*, *données*, *donner*, *donnez*, etc.). In anderen Sprachen wird diese Verbindung zwischen Akzentzeichen und ihren nicht akzentuierten Entsprechungen nicht hergestellt (z. B. auf Tschechisch: *cesta* .. *cas* .. *hre* .. *chodník* ..).

Bei der Sortierung wird auf Microsoft-Routinen zurückgegriffen. Bearbeiten Sie Texte in einer Sprache, mit der Microsoft nicht richtig umgehen kann, sollten die Wortlisten dennoch eine konsistente Sortierung aufweisen.

Beachten Sie bitte, dass Groß- und Kleinschreibung berücksichtigt wird, z. B. wird *würde* nach *Würde* erscheinen.

Es ist wichtig zu wissen, dass für zufriedenstellende Ergebnisse beim Vergleich von zwei Wortlisten (z. B. in KeyWords) die Sortierung von entscheidender Bedeutung ist. Wenn Sie also zwei Wortlisten, die auf unterschiedliche Sprachen eingestellt wurden, miteinander vergleichen, werden Sie eigenartige Ergebnisse erhalten.

Wie werden Sprachen ausgewählt?

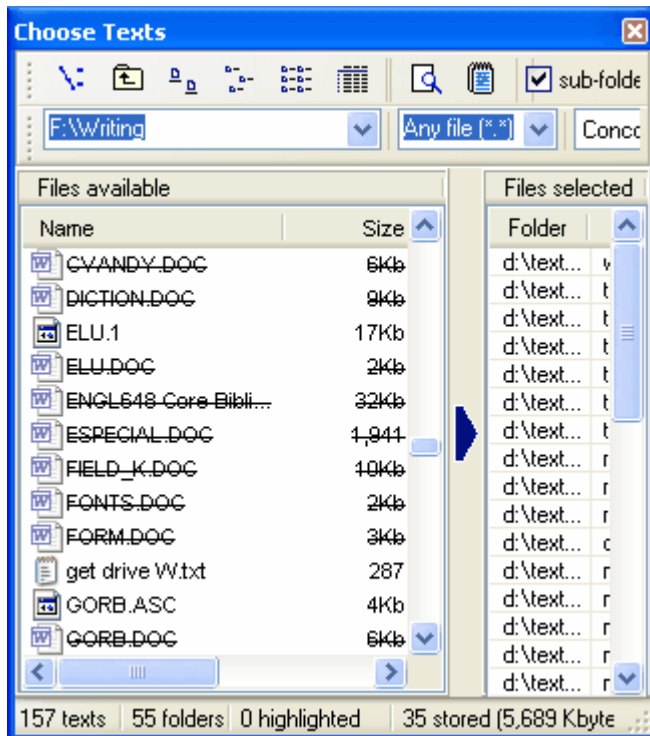
Rufen Sie den Language Chooser über das Menü Utilities im WordSmith Controller auf.

Siehe auch: [Akzente und Symbole auswählen](#), [Akzentzeichen](#), [Chinesische Texte verarbeiten](#) etc.

5.9 Texte auswählen

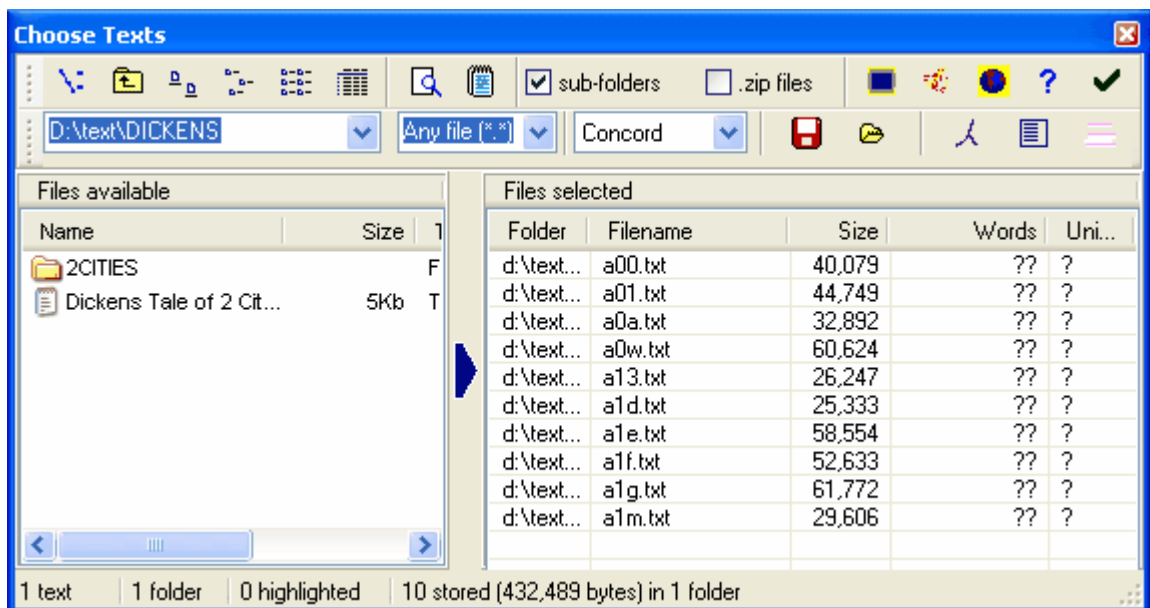
Arten von Textdateien

In WordSmith benötigen Sie [Standardtext-Dateien](#), die Sie erhalten, wenn Sie ein Word-Dokument als Standardtext (.txt) speichern. Alle Word-Dokumente in der Liste sind durchgestrichen.



Wie finde ich das?

Diese Funktion rufen Sie im **Controller**-Menü *Datei* auf, oder in dem Menü *Einstellungen* bzw. im Menüelement *Neu* (🟢) in den verschiedenen Tools.



Die zwei Hauptbereiche links und rechts sind

- Dateien zum Auswählen (links)
- bereits ausgewählte Dateien (rechts)

Um links markierte Dateien rechts hinzuzufügen, klicken Sie auf den dicken blauen Pfeil oder

ziehen Sie die Dateien einfach von links nach rechts.

Die Liste auf der rechten Seite zeigt alle Details der Dateien an (Name, Datum, Größe, Anzahl der Wörter und ob der Text in [Unicode](#) kodiert ist (die Informationen bezüglich Wortanzahl und Unicode wurden noch nicht berechnet und werden daher wie oben mit ?? angezeigt; diese Informationen angezeigt, sobald die Konkordanz bzw. Wortliste erstellt wurde). Rechts der Unicode-Spalte befindet sich die Spalte 'Filter', in der angegeben wird, ob die Textdatei Ihren [Filterkriterien entspricht](#).

Mit den Schaltflächen oben links können Sie die Ansicht der Dateien auf "Symbole", "Liste" oder "Details" (Standardeinstellung) festlegen.

Wenn Sie WordSmith noch nie zuvor verwendet haben (genauer gesagt, wenn Sie noch keine Konkordanzen, Wortlisten, etc. gespeichert haben), wird ein Kapitel aus Charles Dickens *A Tale of Two Cities* für Sie ausgewählt. Um dies zu verhindern, speichern Sie mindestens eine Wortliste oder Konkordanz. Siehe auch -- [vorherige Listen](#).

Dateitypen

Die Standardeinstellung für Dateien ist *.* (d. h. jede beliebige Datei), aber diese Einstellung kann im Fenster über dem dicken blauen Pfeil temporär oder in [wshell.ini](#) permanent verändert werden.

Tool

Im Screenshot oben sehen Sie "Concord" -- es werden also gerade Texte für Concord ausgewählt. Es sind auch Alternativen verfügbar (WordList, KeyWords, etc.).

Sortieren

Wenn Sie auf *Name*, *Größe*, *Typ*, *Wörter*, *Unicode* oder *Filter* klicken, wird die Liste entsprechend neu sortiert. Mit der rot-gelben Schaltfläche () werden die Dateien (auf beiden Seiten) zufällig neu sortiert.

Alle auswählen



Alle Dateien im aktuellen Verzeichnis werden ausgewählt.

Laufwerke und Verzeichnisse

Durch Doppelklicken auf das Verzeichnis wird dieses geöffnet. Sie können ein Verzeichnis einfach erneut öffnen, wenn der Name des Verzeichnisses in der Verlaufsliste des

Verzeichnisfensters erscheint, und mit der Windows-Schaltfläche "Zurück" () zum vorherigen Verzeichnis zurückzukehren. Oder klicken Sie auf die Schaltfläche () um ein neues Laufwerk oder Verzeichnis auszuwählen.

Unterverzeichnisse

Wenn dieses Kontrollkästchen aktiviert ist, während Sie ein gesamtes Laufwerk oder Verzeichnis auswählen, werden auch alle Dateien in den Unterverzeichnissen in diesem Laufwerk oder Verzeichnis ausgewählt.

Anzeigen



Damit können Sie alle aktuell ausgewählten Dateien durchsuchen und prüfen, ob Sie eine Datei auswählen möchten. Alle Akzentzeichen (z. B. æ, é) oder Währungssymbole wie etwa £, ¥, ¢ und [Tags](#) werden entsprechend Ihrer aktuellen [Einstellungen bezüglich Texteigenschaften](#) angezeigt. Sie können diese Einstellungen verändern, während Sie die Datei [anzeigen](#).

Im Editor anzeigen



Mit dieser Funktion können Sie die Textinhalte im *Editor* anzeigen, dem einfachen Windows Textverarbeitungsprogramm für Textdateien.

Aus dem Internet aufrufen

Mit dieser Funktion wird [WebGetter](#) aufgerufen, mit dem Sie Texte aus dem Internet herunterladen können.

Zip-Dateien

Wenn Sie auf eine [Zip-Datei](#) doppelklicken, wird diese Datei wie ein normales Verzeichnis geöffnet und der Inhalt angezeigt. Sie können die beinhalteten Dateien auch anzeigen.

Favouriten

Die zwei Schaltflächen rechts ( und ) ermöglichen es Ihnen, eine [Dateiauswahl zu speichern bzw. eine bereits gespeicherte Auswahl zu laden](#), was Ihnen die Arbeit erspart, sich an eine komplexe Auswahl zu erinnern und diese erneut vorzunehmen.

Unicode-Test

Mit dieser Schaltfläche werden alle ausgewählten Dateien getestet. Im Screenshot oben wurde noch kein Unicode-Test durchgeführt und deshalb ? für jede Datei angezeigt. Wenn die Textdatei in [Unicode](#) ist, wird U angezeigt, A für ASCII- oder [ANSI](#) -Text, D für ein Word-Dokument.

Entfernen

Wie der Name bereits vermuten lässt, können Sie hier die ausgewählten Dateien entfernen, um eine neue Auswahl vorzunehmen. Wenn ausgewählte Dateien markiert sind, werden nur die markierten Dateien entfernt.

OK

Hiermit wird die aktuelle Dateiauswahl gespeichert. Wenn das Kontrollkästchen "auch Unterverzeichnisse" aktiviert ist, werden auch alle festgelegten Dateitypen in Unterverzeichnissen gespeichert.

Sie können unter *alle aktuellen Einstellungen* überprüfen, welche Dateien bereits ausgewählt sind.

Siehe auch: [Quelltexte finden](#).

5.10 Dateien aus Standard-Dialogfeld auswählen

The dialogue box here is very similar to the one used for [choosing text files](#); it also allows you to choose from a [zip file](#).

You can use [Viewer](#) to examine a file: this makes no sense in the case of a word list, key word list, or concordance, but may be useful if you need to examine a related text file, e.g. a *readme.txt* in the same zip file as your concordance or word lists.

To choose more than one file, hold the Control key down as you click with your mouse, to select as many separate files as you want. Or hold down the Shift key to select a whole range of them.

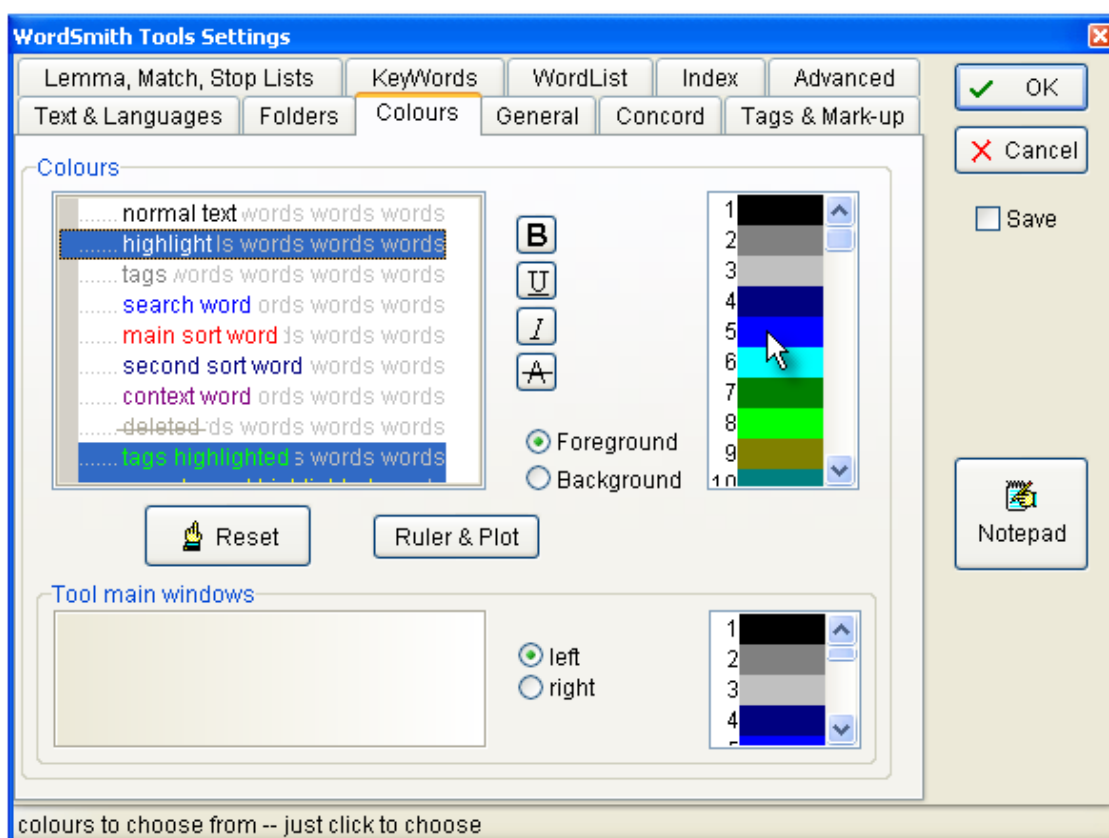
5.11 Farben

Farbeinstellungen können Sie im Menü "Einstellungen" in allen Tools sowie im Menü "Einstellungen | Einstellungen anpassen" im [Controller](#) festlegen. Hier können Sie die Standardfarben für alle Tools auswählen. Die verfügbaren Farben können für folgende Elemente festgelegt werden:

Standardtext

die Standardfarbe

markierter Text	wie oben, wenn ausgewählt
Tags	Mark-up
Suchwort	Suchwörter in Konkordanz; Wörter in (Schlüssel-)Wortlisten
Hauptsuchwort	zeigt die Primärsortierung an; für %-Daten in (Schlüssel-)Wortlisten verwendet
zweites Suchwort	zeigt die Farbe der Sekundärsortierung an
Kontextwort	Kontextwort
gelöschte Wörter	jede Zeile mit gelöschten Daten
nicht nummerierte Zeile	jede Zeile, die nicht vom Benutzer sortiert wurde
markiertes Suchwort	Konkordanzsuchwort, wenn es ausgewählt ist
markiertes Hauptsuchwort	Primärsortierung, wenn ausgewählt
markiertes zweites Suchwort	Sekundärsortierung, wenn ausgewählt
markiertes Kontextwort	Kontextwort, wenn ausgewählt
häufigster Kollokator	der häufigste Kollokator oder detailliertes Konsistenzwort, P-Wert bei Schlüsselwörtern
Texte anzeigen	im Text- Viewer

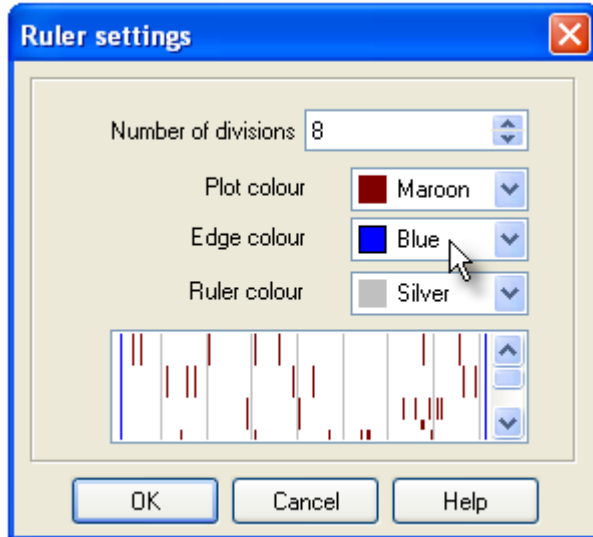


Um die Farben zu ändern, klicken Sie zuerst auf das Element, das Sie ändern möchten (Sie sehen den Unterschied am linken Rand: hier wurde Markierter Text ausgewählt), danach klicken Sie auf die Farbe in dem Farbenfeld (dort, wo sich im Screenshot der Cursor befindet). Mit den Optionsfeldern unten legen Sie fest, ob Schriftfarbe oder Hintergrund verändert werden soll. Sie können die Standardeinstellungen wiederherstellen, indem Sie auf die Schaltfläche "Zurücksetzen" klicken. Dieselben Farben bzw. Graustufen erscheinen auf Ihrem Ausdruck. Sie können auch den

[Drucker auf Schwarz/Weiß einstellen](#), dann werden alle Daten, die nicht Standardtext enthalten, in Kursivschrift (oder fett bzw. unterstrichen, wenn Sie die Spalten bereits auf Kursivschrift eingestellt haben).

Lineal und Plot

Durch Klicken auf diese Schaltfläche wird ein neues Dialogfenster geöffnet, in dem Sie die Farben und Plotabschnitte für das Lineal festlegen können:



Siehe auch: [Layout](#) um die einzelnen Farben für jede Datenspalte zu verändern.

5.12 Spaltensummen

The point of it...

This function allows you to see a total and basic statistics on each column of data, if the data are numerical.

How to do it

With a word-list, concordance or key-words list visible, choose the menu item *View | Column Totals* to switch column totals on or off.

- RelC(2)** für jede Zeile in Ihren Daten wird die neue Spalte auf Daten der Spalte 2 in derselben Zeile zugreifen, diese zur laufenden Summe addieren, und das Ergebnis in Ihre neue Spalte schreiben.
- Rel(3) + (Rel(2) ÷ 5)** für jede Zeile in Ihren Daten wird die neue Spalte auf Daten der Spalte 2 in derselben Zeile zugreifen, diese durch fünf dividieren, zu den Daten derselben Zeile in Spalte 3 addieren, und das Ergebnis in Ihre neue Spalte schreiben.
- Abs(2;1) ÷ 5** für jede Zeile in Ihren Daten wird die neue Spalte auf Daten in Zeile 1 der Spalte 2 zugreifen, diese durch fünf dividieren, und das Ergebnis in Ihre neue Spalte schreiben. Dieses Beispiel dient nur zur Veranschaulichung, da diese Formeln in jeder Zeile zum exakt gleichen Ergebnis führen.
- Rel(2) ÷ Abs(2;1) × 100** für jede Zeile in Ihren Daten wird die neue Spalte auf Daten der Spalte 2 in derselben Zeile zugreifen, diese durch den Wert in Zeile 1 der Spalte 2 dividieren, mit 100 multiplizieren, und das Ergebnis in Ihre neue Spalte schreiben. Somit würde in Spalte 3 ein Prozentwert des obersten Ergebnisses in Spalte 2 angezeigt. In der ersten Zeile wären das 100%, aber mit sinkender Frequenz sinkt auch der Prozentsatz, der auf dem häufigsten Element basiert.

Jede beliebige Variable, die so berechnet wurde, können Sie formatieren (oder auch löschen): siehe [Layout](#).

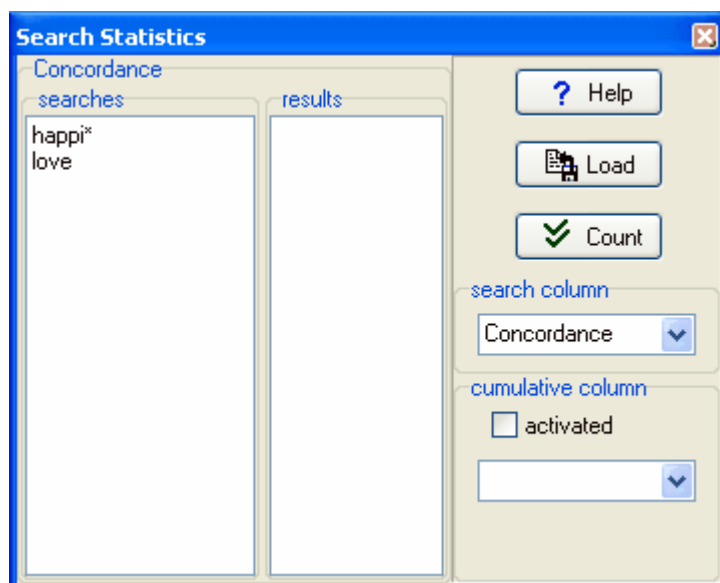
Siehe auch: [Datenfrequenzen zählen](#), Spaltensummen

5.14 Datenfrequenzen zählen

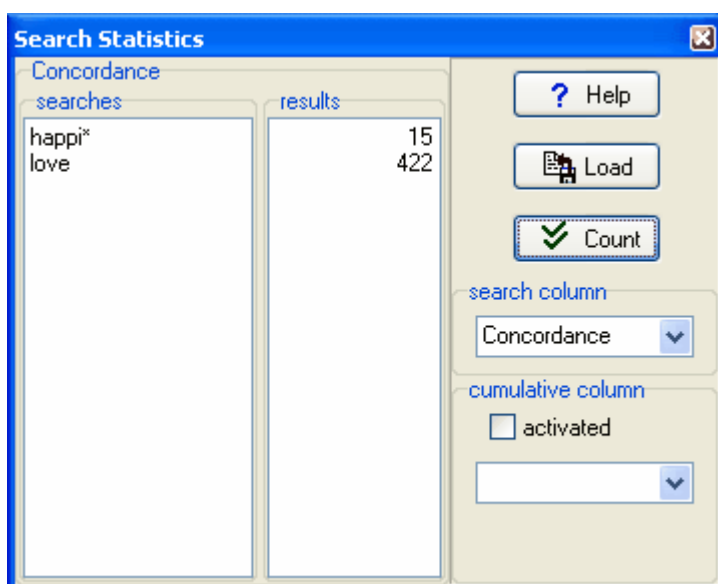
You may want to know how many of your concordance lines contain "**happi***" or how many items in a word-list end in ***1y**. To do this, choose *Summary Statistics* in the *Compute* menu.

An Example

You have a concordance already computed. Select anywhere in the concordance lines and choose *Compute : Summary Statistics*. Type **happi*** and **love** in the searches window.



Press Count -- you should see something like this:



The procedure has processed all your concordance lines and found out that 15 contain **happi*** and 422 contain the whole word **love** (not **loved**, **loves**).

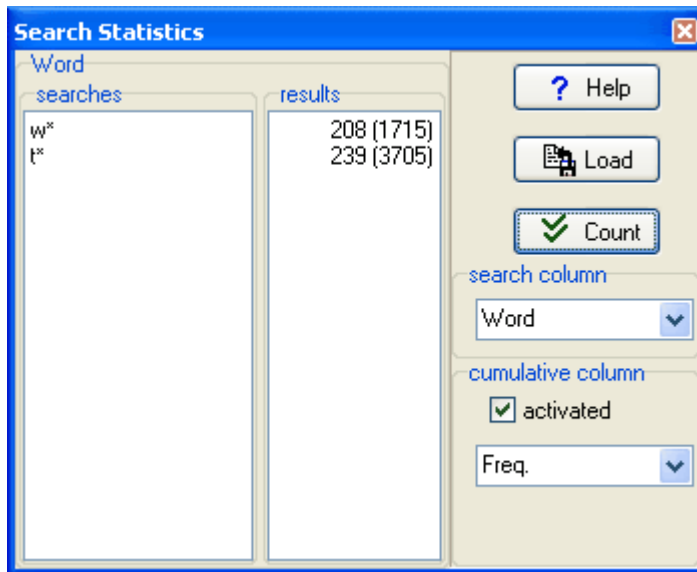
Search Column

This combobox lets you choose which column of data to count in.

Cumulative Column

A cumulative count adds up scores on another column of data apart from the one you are processing for your search. The columns in this combobox are of the numerical data only. Select one and ensure *activated* is ticked.

In this example, a word-list was computed and a search was made of words beginning with **W** or **T**. There are 208 of those beginning with **W**. In brackets you can see 1715 -- this means that the cumulative count of words beginning with **W** in terms of their frequency (*Freq.column*) is 1715, in other words an average frequency of about 8 ($1715 / 208$). But for the words beginning in **T**, although the absolute number is similar (239), the average cumulative frequency is about 15. That is because there are lots of high-frequency words beginning in **T** in English.



Load

This allows you to load into the searches window any [plain text](#) file which you have prepared previously.

See also: [compute new column of data](#).

5.15 Ergebnisse kopieren

Am schnellsten und einfachsten kopieren Sie Daten (z. B. in Ihr Textverarbeitungsprogramm), indem Sie mit den Pfeiltasten die Stelle auswählen und dann Strg+Einfg drücken. Die Auswahl wird so in die Zwischenablage kopiert.

Wenn Sie Kopieren anklicken (Aktualisierung erforderlich), haben Sie die folgenden Möglichkeiten:

[Zwischenablage](#)
[als Textdatei speichern](#)
[drucken](#)

als Daten [speichern](#) (nicht dasselbe wie als Text speichern, sondern Ihre Daten werden so gespeichert, dass Sie sie an einem anderen Tag wieder aufrufen können.)

alle: alle Zeilen und Spalten mit Daten kopieren

markierte: nur die Zeilen und Spalten kopieren, die bereits markiert sind

festlegen: entsprechend der Kriterien kopieren, die vom Tool abhängig sind:

In **Concord** können Sie alle Konkordanzzeilen [speichern](#), die Sie [entsprechend Ihrer eigenen Kategorien](#) klassifiziert haben. Sie können eine Kategorie (zwischen **a** und **z** oder zwischen **A** und **Z**) festlegen und diese speichern.

5.16 benutzerdefinierte Verarbeitung

This feature -- which is not for those without a tame programmer to help -- is found under Adjust Settings | Advanced.

The point of it...

I cannot know which criteria you have in processing your texts, other than the criteria already set up (the choice of texts, of search-word, etc.) You might need to do some specialised checks or alteration of data before it enters the **WordSmith** formats. For example, you might need to lemmatise a word according to the special requirements of your language.

This function makes that possible. If for example you have chosen to filter concordances, as **Concord** processes your text files, every time it finds a match for your search-word, it will call your .dll file. It'll tell your own .dll what it has found, and give it a chance to alter the result or tell **Concord** to ignore this one.

How to do it...

Choose your .dll file (it can have any filename you've chosen for it) and check one or more of the options in the Advanced page. You will need to call standard functions and need to know their names and formats. It is up to you to write your own .dll program which can do the job you want. This can be written in any programming language (C++, Java, Pascal, etc.).

An example for lemmatising a word in WordList

The following DLL is supplied with your installation, compiled & ready to run.

Your .dll needs to contain a function with the following specifications

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

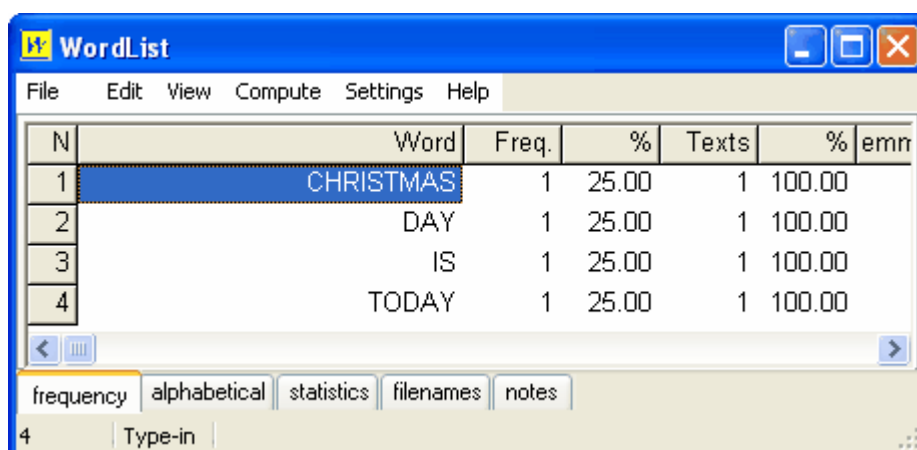
The language_identifier is a number corresponding to the language you're working with. See [List of Locale ID \(LCID\) Values as Assigned by Microsoft](#).

So the "original" (sent by WordSmith) can be a PCHAR (7 or 8-bit) or a PWIDECHAR (16-bit Unicode) and the result which your .dll supplies can point to

- a) nil (if you simply do not want the original word in your list)
- b) the same PCHAR/PWIDECHAR if it is not to be changed at all
- c) a replacement form

Here's an example where the source text was

Today is Easter Day.



The source code for the .dll in Delphi is this

```
*****
```

```
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;
    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 Popupmenüs anpassen

Protokollierung

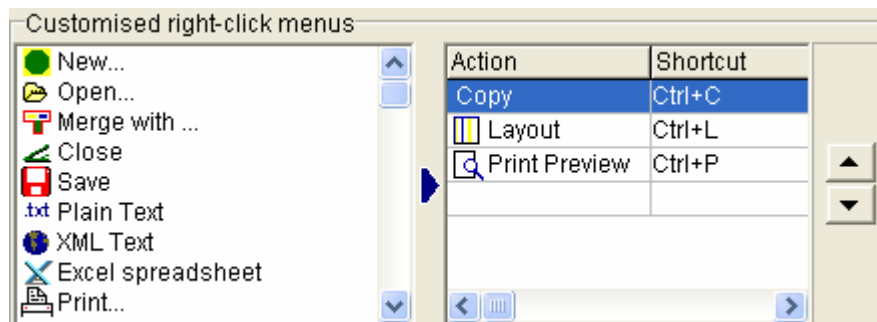
Protokollierung ist sinnvoll, wenn Sie eigenartige Ergebnisse erhalten und wissen möchten, wie diese Ergebnisse zustande gekommen sind. Wenn diese Funktion aktiviert ist, speichert WordSmith Informationen über den Verlauf der Ergebnisse in der Protokolldatei.

Wählen Sie dazu im WordSmith Controller *Einstellungen anpassen* | *Erweitert*. Hier können Sie die Protokollierung aktivieren bzw. deaktivieren und einen entsprechenden Dateinamen für die Protokolldatei (auch: Log-Datei) eingeben. Wenn Sie diese Funktion aktivieren, haben Sie die Möglichkeit die frühere Protokolldatei zu löschen. Siehe auch: [Fehlerberichte per e-mail senden](#).

Popupmenüs anpassen

Alle Datenraster verfügen über ein "Popupmenü", das erscheint, wenn Sie mit der rechten Maustaste auf die Daten klicken.

Um diese Popupmenüs anzupassen, wählen Sie im WordSmith Controller *Einstellungen anpassen* | *Erweitert*.



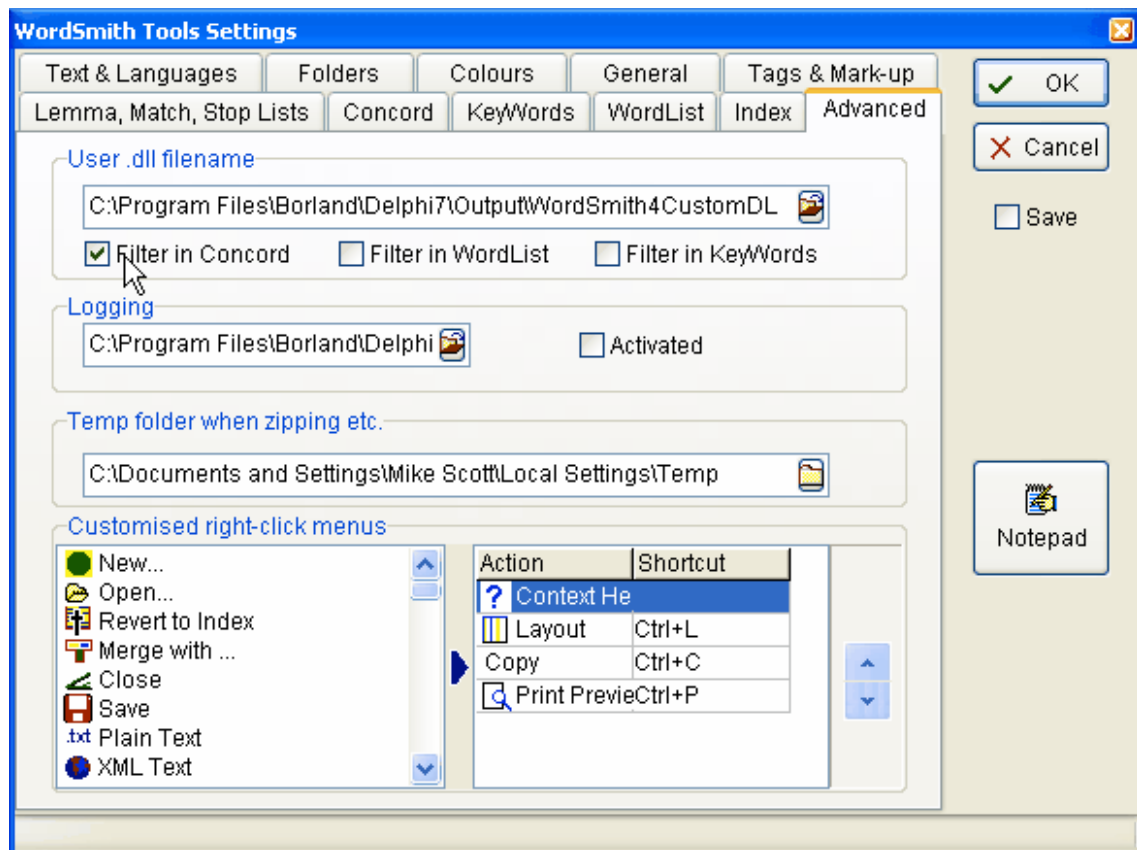
Sie werden auf der linken Seite eine Liste mit Menüoptionen sehen, und können sie zu der Liste auf der rechten Seite hinzufügen, indem Sie auf eine Option klicken und die Schaltfläche in der Mitte anklicken. Um eine ausgewählte Option von der rechten Seite zu entfernen, wählen Sie die Option aus und drücken *Entf*. Um Ihre Auswahl neu zu ordnen, drücken Sie auf den Pfeil nach oben oder den Pfeil nach unten.

Funktionen, die in Ihrem Popupmenü sind, scheinen auch in der [Symbolleiste](#) auf.

Um die Auswahl permanent zu speichern, siehe [Standardeinstellungen speichern](#).

Benutzer-.dll

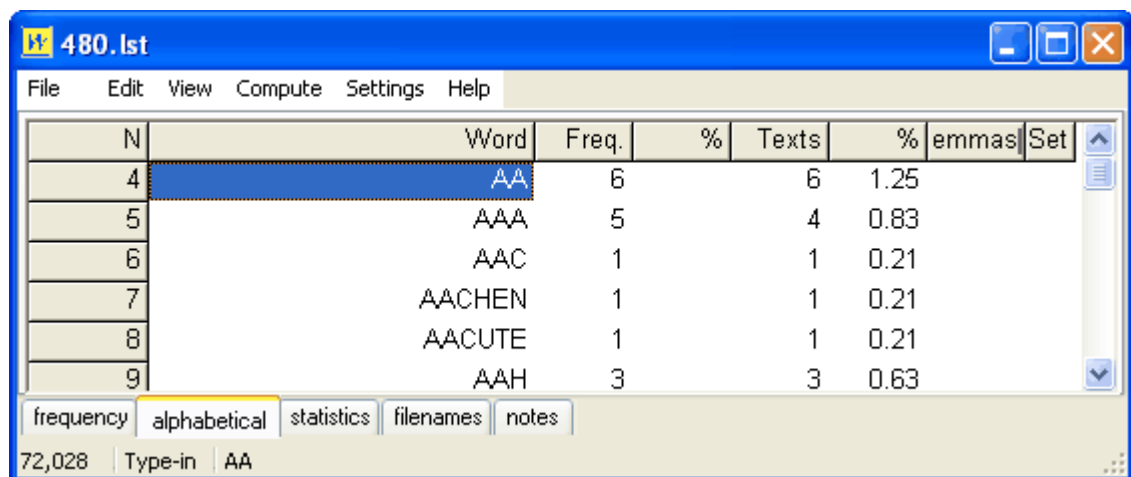
Wenn Sie eine DLL haben, mit der Sie die WordSmith-Ergebnisse aufnehmen möchten, können Sie diese hier auswählen. Dieser Benutzer wählt die Datei **wordSmithCustomDLL.dll** aus, die bei der Installation mitgeliefert wird und bei Bedarf verwendet werden kann. Wenn "In Concord filtern" aktiviert ist, speichert diese .dll alle gefundenen Konkordanzlinien hintereinander in eine Standardtext-Datei mit dem Namen **Concord_user_dll_concordance_lines.txt** in dem Verzeichnis **\wsmith4**, wenn auf der Festplatte genügend Platz vorhanden ist.



Siehe auch: [Menü- und Schaltflächenoptionen](#).

5.18 Datenliste bearbeiten

Wenn Sie eine Wortliste geöffnet haben, sieht Ihr Bildschirm ungefähr folgendermaßen aus:



In der Statusleiste unten,



die Zahl in der ersten Zelle entspricht der Anzahl der Wörter in der aktuellen Wortliste und **AA** in der dritten Zelle ist das ausgewählte Wort.

Wenn der Benutzer jetzt etwas eingibt, sucht WordList die eingegebene Zeichenfolge in der Liste.

Wenn Sie mit der rechten Maustaste auf die zweite Zelle klicken, sehen Sie

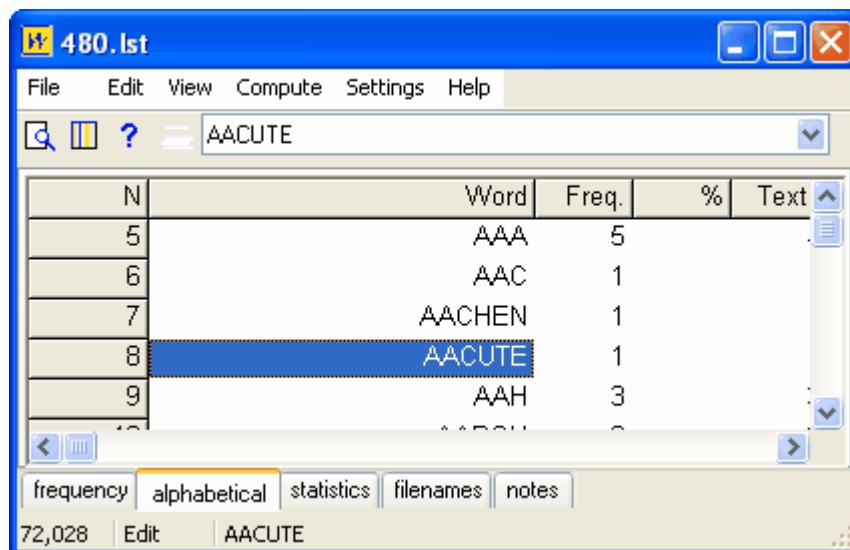


und können die Optionen für diese Liste auf *Klasse* (um die Wörter zu klassifizieren, z. B. Adjektive vs. Substantive) oder *Bearbeiten* einstellen, um die Wörter zu ändern. Beachten Sie bitte, dass einige Daten basierend auf anderen Daten berechnet wurden und deshalb nicht bearbeitet werden können. Frequenz-Prozentsätze etwa basieren auf der Wortfrequenz und der Gesamtanzahl der Wörter. Sie können die Wortfrequenz bearbeiten, aber nicht den Prozentsatz der Wortfrequenz.

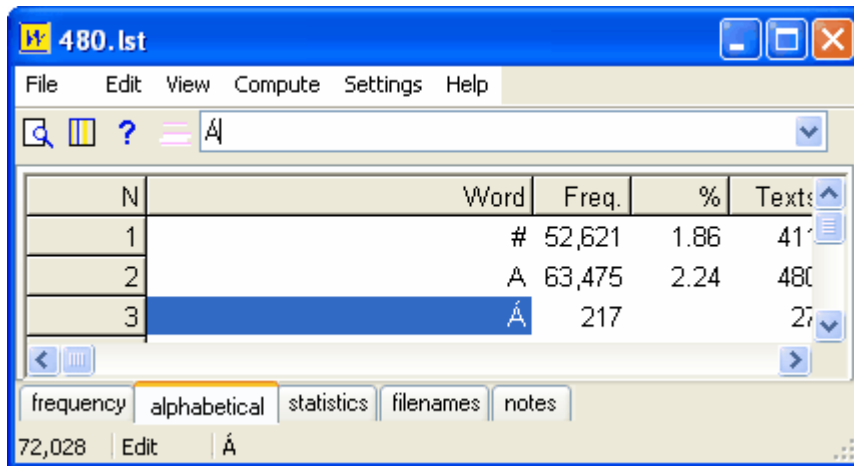
Wählen Sie *Bearbeiten*.

Drücken Sie jetzt in der Spalte, die Sie bearbeiten möchten, auf irgendeinen Buchstaben. Die Symbolleiste wird angezeigt (wenn sie vorher nicht zu sehen war), sodass Sie jetzt die Wortform oder die Frequenz eines Worts ändern können. Wenn Sie das Wort so buchstabieren, dass es mit einem Wort übereinstimmt, das bereits in der Liste vorhanden ist, wird die Liste unter Berücksichtigung Ihrer Veränderungen umgeschrieben.

In diesem Fall soll das Wort **AACUTE** zu **Á** korrigiert werden.



Wenn Sie nun **Á** eingeben, wird das Ergebnis unmittelbar angezeigt:



Wenn Sie auf den Pfeil nach unten klicken, der sich rechts neben dem Kombinationsfeld "Bearbeiten" befindet, wird das ursprüngliche Wort angezeigt für den Fall, dass Sie es doch beibehalten möchten.



Nachdem dem Bearbeiten ist es sinnvoll, [neu zu sortieren](#) (🔍). Wenn Sie ein Wort wie **AAAAAGH** zu dem bereits vorhandenen Wort **AAGH** geändert haben, sollten Sie außerdem die beiden Einträge [zusammenführen](#).

Siehe auch: [Einträge zusammenführen](#), [Quelldateien finden](#).

5.19 relevante Dateien finden

Wozu dient das?

Angenommen Sie stellen fest, dass *Muskel*, *Faser* und *Protein* die Schlüsselwörter eines bestimmten Texts sind. Dann möchten Sie vielleicht herausfinden, ob es noch andere Texte in Ihrem Korpus gibt, in denen diese Wörter vorkommen.

Wie mache ich das?

Diese Funktion kann aus jedem Datenfenster mit der Option **F** aufgerufen werden, z. B. eine [Schlüsselwort](#)-Liste.

Die Funktion ermöglicht die Suche nach allen Textdateien, in denen **jedes** der Wörter, die Sie (mit **MB**) markiert haben, **zumindest einmal** auftritt. Vor dem Anklicken [wählen Sie die Texte](#) aus, die Sie untersuchen wollen.

Ergebnis

Eine Konkordanz basierend auf allen Wörtern, die Sie markiert haben und die anzeigt, in welchen [Textdateien](#) die Wörter gefunden wurden. Es ist jedoch eine "kleinliche" Konkordanz: diejenigen Textdateien, in denen nicht *alle* Wörter enthalten sind, werden ignoriert.

5.20 Unterrichtsanweisungen

When WordSmith is run in a training session, you may want to make certain instructions available to your trainees.

To do this, all you need to do is ensure there is a file called **teacher.rtf** in your main \wsmith4 folder where the WordSmith programs are or in the "instructions folder" explained under [Network Defaults](#). If one is found, it will be shown automatically when WordSmith starts up. To stop it being shown, just rename it! You edit the file using any Rich Text Format word processor, such as MS Word™, saving as an .rtf file.

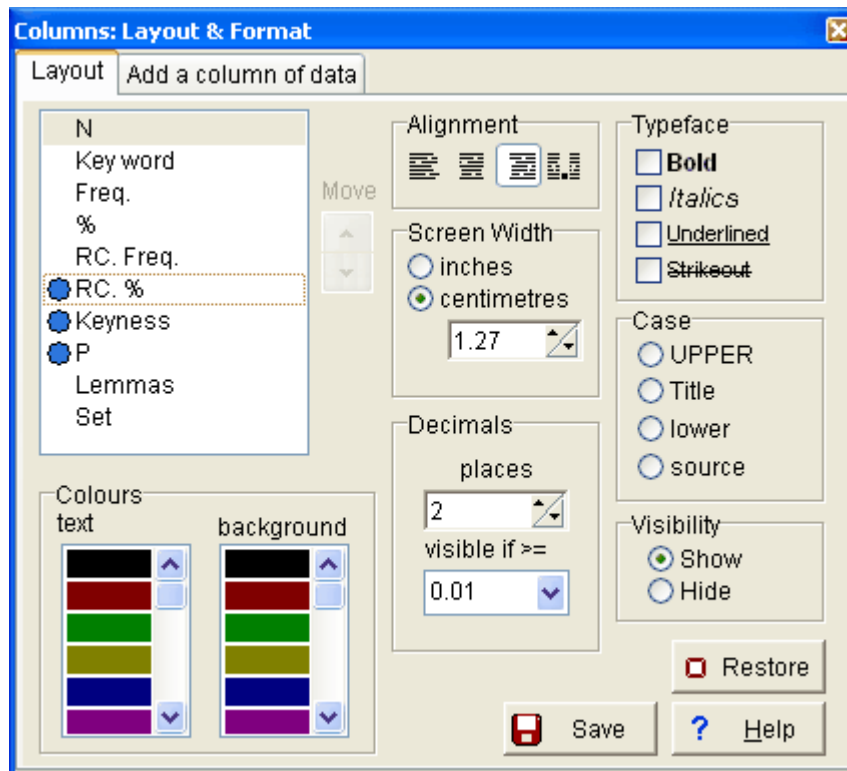
See also: [Network defaults](#)

5.21 Layout & Format

Wenn Sie eine Liste geöffnet haben, klicken Sie mit der rechten Maustaste oder unter dem Menü Ansicht | Layout , um Ihr bevorzugtes Anzeigeformat für jede Datenspalte auszuwählen.

Layout oder Daten hinzufügen?

In der Registerkarte *Layout* können Sie das Layout Ihrer Daten formatieren. Mit der Funktion *Datenspalte hinzufügen* können Sie eine [neue Variable berechnen](#).



Im oberen linken Fenster in Layout sehen Sie die verfügbaren Spaltenüberschriften. Drei der Überschriften wurden durch Klicken aktiviert, sodass die Einstellungen verändert werden können:

Verschieben

Klicken Sie auf die Pfeile, um eine Spalte auf- oder abwärts zu verschieben. So verändert sich die Reihenfolge der Spalten.

Ausrichtung

Hier kann die Ausrichtung des Texts (links, zentriert, rechts, Dezimalausrichtung) für jede Spalte je nach Bedarf festgelegt werden.

Schriftbild

Normal, fett, kursiv und/oder unterstrichen. Wenn Sie keine der Möglichkeiten aktivieren, bleibt das Schriftbild normal.

Fensterbreite

in Ihrer bevorzugten Einheit (Zentimeter oder Zoll).

Groß-/Kleinschreibung

kleinbuchstaben, GROSSBUCHSTABEN, Erster Buchstabe Groß oder Quelle: So, wie das Wort ursprünglich in der Textdatei geschrieben wurde. Standardmäßig sind die meisten Daten auf GROSSBUCHSTABEN eingestellt.

Dezimalstellen

die Anzahl der Dezimalstellen bei numerischen Daten, wo zutreffend.

Anzeige

einblenden oder ausblenden, oder nur einblenden, wenn höher als eine bestimmte Zahl. (Wenn Sie hier *** sehen, dann kann diese Option für die Daten in der aktuell markierten Spalte, nicht ausgeführt werden.)

Farben

Das Fenster links unten zeigt die verfügbaren Farben für Schriftfarbe und Hintergrund. Klicken Sie auf eine Farbe, um die Anzeige für die aktuell ausgewählte Spalte zu verändern.

Wiederherstellen

Die vorherigen Einstellungen werden wiederhergestellt. Außerdem können Sie hier alle gespeicherten benutzerdefinierten Layouts gelöscht werden (siehe unten).

Speichern

Durch Anklicken der Option "Speichern" werden zukünftig alle Listen in diesem Layout erstellt. Sie haben zum Beispiel eine Konkordanz geöffnet. Wenn Sie das Layout nach Bedarf verändern und danach die Konkordanz auf dem üblichen Weg [speichern](#), werden Ihre Layouteinstellungen natürlich für diese Konkordanz gespeichert. Aber wenn Sie das nächste Mal eine Konkordanz erstellen, wird wieder auf das WordSmith Standardlayout zurückgegriffen. Wenn Sie das Layout speichern, wird das Layout der nächsten Konkordanz so aussehen wie das aktuelle Layout. Das benutzerdefinierte Layout wird in dem Verzeichnis `\wsmith4` gespeichert, z. B. **Concordance list customised.dat**.

Sie können aber auch die Datei `wshell.ini` unter Einstellungen so verändern, dass bestimmte Datenspalten immer ausgeblendet oder eingeblendet werden. Wenn Sie z. B. die Spalte Klasse nicht anzeigen möchten, würden Sie in `\wsmith4\wshell.ini` im Abschnitt [Concord] folgendes verändern:

show set column=YES

bis

show set column=NO

Siehe auch: [Standardeinstellungen vornehmen und speichern](#), [Farben](#) im Oxford WordSmith Tools [Controller](#) auswählen.

5.22 Wörter in Liste übereinstimmen

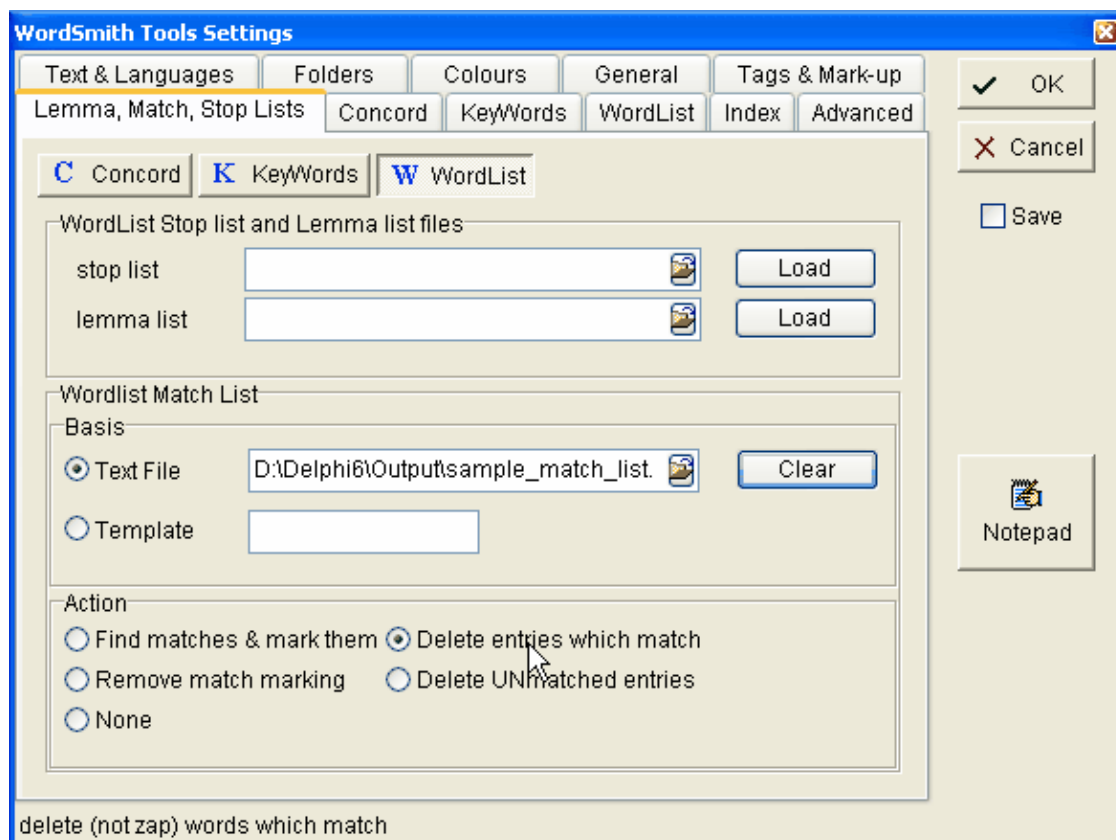
Wozu dient das?

Diese Funktion hilft beim Filtern Ihrer Liste. Sie können die Einträge einer Konkordanz oder Liste (Wortliste, Kollokatorenliste, etc.) mit einer Gruppe spezifischer Wörter, die Sie interessieren,

verknüpfen. Zum Beispiel alle Funktionswörter in Ihrer Liste markieren oder alle, die auf **-ung** enden. Alle Übereinstimmungen werden mit einer Tilde (~) markiert. Sie können nun alle markierten Einträge (oder alle nicht markierten Einträge) löschen oder die Einträge nach der Markierung sortieren.

Wie mache ich das?

Klicken Sie auf die Spalte, deren Daten Sie übereinstimmen möchten. Das ist normalerweise eine Spalte, die Wörter (und nicht Zahlen) enthält. Danach klicken Sie auf die Schaltfläche *Matchliste*. Das Controller-Dialogfeld Einstellungen wird geöffnet.



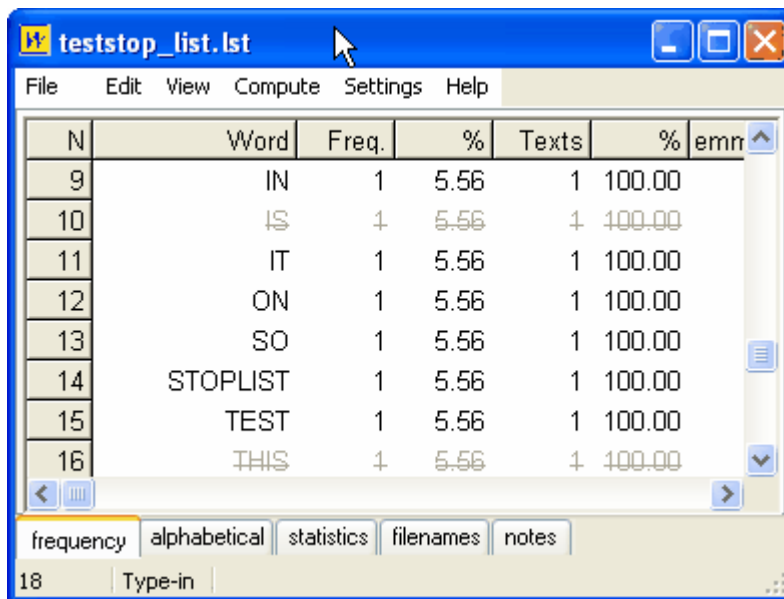
Textdatei oder Vorlage

Wählen Sie nun aus, ob Sie mittels einer Textdatei filtern möchten, die alle notwendigen Wörter enthält (z. B. eine Standardtext-Datei mit Funktionswörtern [nicht in WordSmith Installationsdateien enthalten]), oder anhand eines Vorlagefilters, z. B. ***ung** (jeder Eintrag wird auf die Endung **-ung** überprüft). Wenn Sie eine Datei auswählen, wird diese vom Controller gelesen und Sie werden informiert, wie viele Wörter diese Datei enthält.

Aktion

Vergessen Sie nicht, die entsprechenden Aktionseinstellungen vorzunehmen. Im Screenshot zeigt der Cursor auf die Option "übereinstimmende Einträge löschen".

Das aktuelle Tool überprüft dann jeden Eintrag in der markierten Spalte Ihrer aktuellen Liste darauf, ob er entweder mit der Vorlage oder einem der Wörter in Ihrer Standardtext-Datei übereinstimmt. Die Übereinstimmungen werden entsprechend der Aktionseinstellungen markiert oder gelöscht (im Beispiel unten enthält die Matchliste die Wörter **IS** und **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	

frequency alphabetical statistics filenames notes

18 Type-in

Sie könnten die Statistik der Übereinstimmungen unter der Menüoption [Statistikzusammenfassung](#) aufrufen

Siehe auch: [Worlisten vergleichen](#), [Versionen vergleichen](#), [Stopplisten](#), [Lemmata übereinstimmen](#)

5.23 Drucken und Druckvorschau

Die Druckvorschau wird geöffnet und es wird angezeigt, wie die aktuelle Datenseite aussieht. Aus der Druckvorschau können Sie diese Seite drucken.

Größer und Kleiner

Sie können auf 100% zoomen () , an die Seitenbreite anpassen () oder eine Ansicht aus der Liste auswählen. In der Druckvorschau wird alles genau so angezeigt, wie es auf dem Ausdruck auf Papier später aussieht. Kleine Unterschiede zwischen Druckvorschau und Ausdruck können aufgrund von Unterschieden bei der Schriftart entstehen.

Nächste () und Vorherige () Seite

So können Sie auf die nächste oder vorherige Seite wechseln.

Hochformat () oder Querformat ()?

Hier legen Sie das gewünschte Format für die Seite fest.

Kopfzeile, Fußzeile, Seitenränder

Sie können einen Kopfzeilen- bzw. Fußzeilentext eingeben, der auf jeder Seite erscheint. Wenn Sie <Datum> hinzufügen, wird automatisch das aktuelle Datum eingefügt. Seitenränder werden durch Klicken auf die Pfeile neben den Zahlen verändert -- Sie sehen die Fußzeilen im rechten Dialogfeld.

Drucken ()

Der Windows-Dialog "Drucken" wird geöffnet und standardmäßig ist die Option "Aktuelle Seite" ausgewählt. Sie können in diesem Standarddialogfeld auch andere Seiten auswählen.

Siehe auch: [Druckereinstellungen](#)

5.24 erstmalige Verwendung von WordSmith

Benutzern, die zum ersten Mal mit WordSmith arbeiten, mag dieser Vorgang komplex erscheinen. (Aber schließlich ist auch ein Textverarbeitungsprogramm beim ersten Mal kompliziert.)

Mit der WordSmith-Installation erhalten Sie eine kleine Textdatei, und wenn WordSmith erkennt, dass Sie das Programm zum ersten Mal verwenden, wird dieser Text automatisch für Sie ausgewählt, um Sie mit Concord, WordList, etc. vertraut zu machen. WordSmith erkennt neue Benutzer daran, dass

- 1) bisher keine Konkordanzen oder Wortlisten [gespeichert](#) wurden, und
- 2) keine Gruppe von [bevorzugten Texten](#) für schnellen Zugriff gespeichert wurde.

5.25 vorherige Listen

In diesem Fenster wird die Liste der Ergebnisse angezeigt, die Sie zu einem früheren Zeitpunkt mit WordSmith berechnet haben.

Um diese Ergebnisse anzuzeigen, müssen Sie die Datei auswählen und doppelklicken. Das entsprechende Tool wird geöffnet und die Daten darin angezeigt.

Das Popupmenü für dieses Fenster wird durch das Klicken auf die rechte Maustaste geöffnet.

Um einen Eintrag zu löschen, wählen Sie ihn aus und drücken dann *Entf*.

Um die Einträge alphabetisch neu zu sortieren, wählen Sie *Neu sortieren*.

5.26 WordSmith beenden

Mit der Tastenkombination *Alt+X* können Sie WordSmith beenden.

Wenn Sie den Oxford WordSmith Tools [Controller](#) beenden, werden auch alle anderen Tools geschlossen.

Wenn Sie *Alt+X* drücken oder WordSmith über den Menübefehl beenden, haben Sie vor Beenden des Tools die Möglichkeit, alle Datengruppen zu speichern, die bis zu diesem Zeitpunkt noch nicht gespeichert wurden. Sie müssen das Beenden bestätigen, wenn noch Datenfenster geöffnet sind.

Wenn Sie in Eile sind, schließen Sie mit der Menüoption "Schnell beenden", bei der diese Kontrollen nicht durchgeführt werden.

Standardmäßig wird die letzte Wortliste, Konkordanz oder Schlüsselwortliste, die Sie gespeichert oder abgerufen haben, automatisch wiederhergestellt, sobald Sie Oxford WordSmith Tools wieder starten. Diese Funktion kann in einer Wshell-Menüoption temporär oder in `\smith4\wshell.ini` permanent ausgeschaltet werden.

5.27 Daten auf N Einträge reduzieren

Wenn Sie eine sehr lange Wortliste, Konkordanz, etc. haben, möchten Sie sie vielleicht zufällig reduzieren (z.B. für Stichproben). Mit dieser Menüoption können Sie festlegen, wie viele Einträge Sie in der Liste haben wollen. Wenn Sie die Daten reduzieren, werden die Einträge zufällig [entfernt](#) bis nur noch so viele übrig bleiben, wie Sie wollen. Dieses Verfahren kann nicht rückgängig gemacht werden. D. h. auf der Festplatte wird nichts verändert, aber wenn Sie Ihre Meinung ändern, müssen Sie neu berechnen oder eine zuvor gespeicherte Version öffnen.

Siehe auch: [entfernen](#), [Datenliste bearbeiten](#).

5.28 als Text speichern

Wozu dient das?

"Als Text speichern" bedeutet, dass Sie Ihre Daten als [Standardtext](#)-Datei speichern (im Gegensatz zu dem WordSmith-Format zum späteren Aufruf von Daten). Es ist im Normalfall schneller, den ausgewählten Text in die [Zwischenablage](#) zu kopieren, z. B. wenn Sie einfach die Ergebnisse in ein Textverarbeitungsprogramm einfügen möchten.

Wenn Sie die Daten in Farbe kopieren oder einen Plot exportieren möchten, ist das Kopieren in die [Zwischenablage](#) sicher die einfachste Möglichkeit.

Im Fall einer Konkordanz, wenn Sie nur die in Ihrer Konkordanzzeile sichtbaren Wörter benötigen (nicht die Anzahl der Zeichen, die unten angeführt wird), kopieren Sie die Daten in die Zwischenablage und fügen sie ein als Text oder als Grafik (speziell einfügen).

Wie mache ich das?

Diese Funktion finden Sie unter Speichern als... | Standardtext () , XML-Text () , Excel-Tabelle () oder In Datei drucken (mit F3 oder ) oder Kopieren () in Textdatei.

Es gibt folgende Optionen:

Header	Wörter, die Sie am Anfang der Daten speichern möchten (geben Sie nichts ein, wenn Sie keine Wörter speichern möchten);
nummeriert	ob angezeigte Nummern in der linken Spalte auch gespeichert werden
Spalten trennen	standardmäßig ein Tabstopp, aber Sie können etwas anderes zum Trennen der Spalten festlegen
Zeilen	alle/alle markierten Zeilen/einen bestimmten Bereich, z. B. 1-10, 5-, -3
Spalten	alle/alle markierten Zeilen/einen bestimmten Bereich (die Nummerierung befindet sich in Spalte 1)

Sie können danach Daten in Tabellen, Datenbanken, Textverarbeitung, etc. einfach wieder abrufen. (Wenn Sie die Daten als Tabelle in einem Textverarbeitungsprogramm speichern möchten, speichern Sie die Daten zuerst als Text und wählen Sie dann in Ihrem Textverarbeitungsprogramm, wenn vorhanden, die Option "Von Text in Tabelle umwandeln". Wählen Sie Text bei Tabstopps trennen.)

Anmerkung: In einer Excel-Tabelle () können maximal 65.000 Zeilen gespeichert werden. Das sieht dann etwa so aus:

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

Bei Konkordanzzeilen werden unter "als Text speichern" so viele Zeichen pro Zeile gespeichert, wie Sie festgelegt haben (im Controller unter den [Einstellungen anpassen | Concord](#)). Der Grund dafür ist, dass aufgrund der fixierten Anzahl an Zeichen und der nicht proportionalen Schriftart die Suchwörter schön gleichmäßig untereinander stehen.

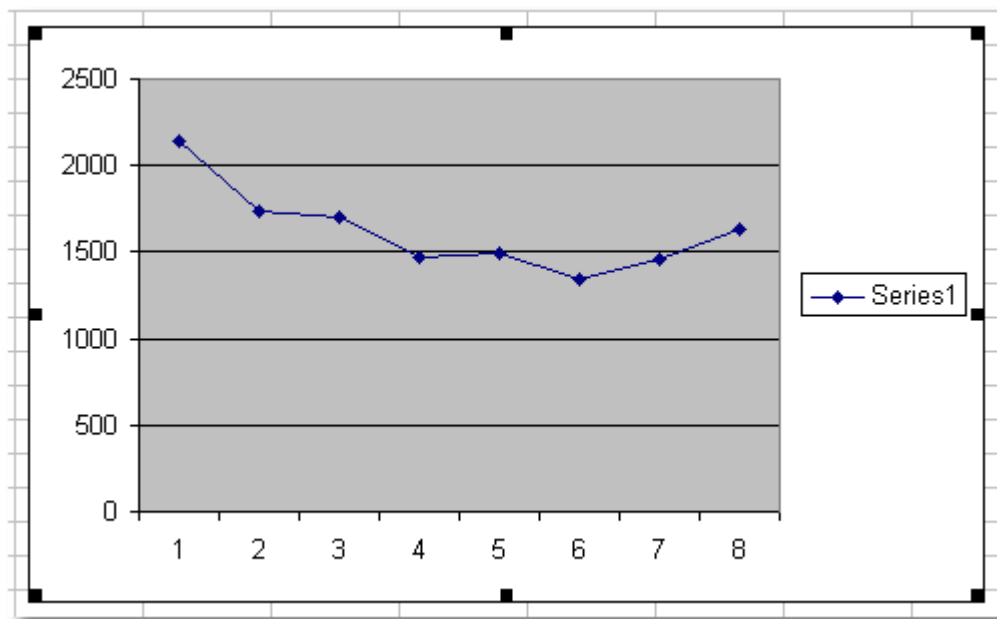
Wenn Ihre Daten einen Plot enthalten, werden Sie noch ein weiteres Arbeitsblatt in Ihrer Exceldatei erhalten, das wie folgt aussieht:

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		0.333333	0.065574	0.213115
hp6.txt	0.333333	0.166667		0.166667				
bn4.txt	0.333333		0.166667	0.166667	0.333333			
asu.txt	0.142857	0.142857	0.571429	0.142857			0.207692	0.215385
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		
aOf.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							

Die Plotdaten sind unterteilt in die Anzahl der Segmente, die für das [Lineal](#) festgelegt wurden (in diesem Fall acht). Die Prozentwerte der einzelnen Segmente stehen in den entsprechenden Spalten. Demnach bedeutet der Wert 23,7% in Zelle B3, dass 23,7% der Daten von cep.txt im ersten Achtel der Textdatei vorkommen. Wenn Sie in Excel die entsprechenden Zellen als Prozent formatieren, sehen Sie in etwa Folgendes:

2140	1735	169
23.77%	15.57%	26.
33.33%	16.67%	
33.33%		16.
14.29%	14.29%	57.
14.67%	22.31%	20.

Oben sehen Sie die Rohdaten, mit denen Sie in Excel eine Grafik erstellen können.



Im Fall eines XML-Texts (🌐) erhalten Sie eine kleine .HTM-Datei und eine große .XML-Datei. Klicken Sie auf die .HTM-Datei, können Sie Ihre Daten Seite für Seite ansehen mit Schaltflächen, um auf die nächste oder vorherige bzw. erste oder letzte Seite zu springen. Dadurch wird auf Ihre XML-Datei zugegriffen, wo die Daten gelesen werden.

5.29 Standardeinstellungen speichern

Die Einstellungen können Sie verändern unter *Einstellungen | Einstellungen anpassen* im Oxford WordSmith Tools [Controller](#). Sie können auch in jedem Tool über das Menüelement Einstellungen die folgenden Einstellungen vornehmen:

Farben, Verzeichnisse, Sprache, Allgemein, Tags, Listen, Concord, KeyWords, WordList

Mit diesen Registerkarten können Sie die Einstellungen auswählen, die ein Tool oder mehrere Tools betreffen.

Farben	Standardfarben anpassen
Verzeichnisse	festlegen, welche Verzeichnisse Sie normalerweise verwenden
e	
Sprache	Zeichensatz , Einstellungen für Bindestriche und Zahlen, Standarddateierweiterungen
Allgemein	letzte Datei wiederherstellen , drucken
Tags	Tags ignorieren, Tagdatei, Tagdatei automatisch laden
Listen	Einstellungen bezüglich Matchlisten , Lemmalisten und Stopplisten für die Tools Concord, KeyWords und WordList
Concord	Anzahl der Einträge, Sortierung, Kollokationsbereich
KeyWords	Verfahren , max. P-Wert , Mindestfrequenz für Datenbank und assoziierte Wörter, Dateiname des Referenzkorpus
WordList	Wortlänge und Frequenzen, Type-Token-Relation , Cluster -Einstellungen

permanente Einstellungen und die Datei wshell.ini

Sie können Ihre veränderten Einstellungen speichern, indem Sie nach dem Anpassen die Schaltfläche Speichern anklicken oder die Datei **wshell.ini** im WordSmith-Verzeichnis bearbeiten. Alle Einstellungen, die Sie oft für alle Tools verwenden, werden so angepasst, z. B. [Verzeichnisse](#) für Ihre Texte und Ergebnisse, [Farben](#), [Schriftarten](#), Standard[spalten](#) in einer Konkordanz, etc.

Sie können **\wsmith4\wshell.ini** anzeigen unter *Einstellungen | Aktuelle anzeigen*.

Hilfdatei anzeigen

In der Registerkarte Allgemein unter "Einstellungen anpassen" sehen Sie das Kontrollkästchen "Hilfdatei anzeigen". Ist dieses Kontrollkästchen aktiviert, wird die Hilfdatei immer angezeigt, wenn WordSmith gestartet wird. Das ist besonders für Benutzer hilfreich, die nur selten mit WordSmith arbeiten, z. B. bei einer Netzwerkinstallation.

Sprichwörter

Alle Sprichwörter, die im [Controller](#) erscheinen, befinden sich in der Datei **wsmith4\sayings.txt**. Mit dem Editor können Sie **wsmith4\sayings.txt** bearbeiten, wenn Sie Sprichwörter entfernen oder hinzufügen möchten.

Netzwerk- und CD-ROM-Standard Einstellungen

Wenn Sie WordSmith direkt von einer CD-ROM starten, können Ihre Standardeinstellungen nicht gespeichert werden, weil es nicht möglich ist, auf einer CD-ROM zu speichern; Windows findet einen passenden Platz für **wshell.ini**, normalerweise im Stammverzeichnis von **c:**. Wenn Sie WordSmith zum ersten Mal verwenden, werden Sie aufgefordert, die Einstellungen anzupassen und [Verzeichnisse](#), [Texteigenschaften](#), [Tag](#)details, etc. auszuwählen. Aktivieren Sie das Kontrollkästchen Speichern, um die Einstellungen für zukünftige Sitzungen zu speichern. Es ist immer möglich, die Einstellungen wieder zu verändern und neu zu speichern.

In einem Netzwerk können Sie im Normalfall die Voreinstellungen nicht permanent verändern, da dies auch andere Benutzer betrifft. Es ist Aufgabe Ihres Netzwerkadministrators das Programm so zu installieren, dass Ihre eigene Kopie von **wshell.ini** bearbeiten können (in einem Verzeichnis, in dem Sie Lese- und Schreibrechte haben). Wenn Oxford WordSmith Tools eine Kopie von **wshell.ini** in diesem Verzeichnis findet, können Sie mit Ihren bevorzugten Einstellungen arbeiten.

5.30 Ergebnisse speichern

Um Ihre korrigierten Ergebnisse zu speichern, wählen Sie im Menü *Speichern* (F2). Alle Ergebnisse werden gespeichert und Sie können die Daten zu einem anderen Zeitpunkt wieder aufrufen. Zuerst können Sie alle [entfernten](#) Elemente löschen.

Gespeicherte Daten haben ein spezielles **Oxford WordSmith Tools**-Format. Das dient nur dazu, die Daten an einem anderen Tag wieder zu verwenden. Sie können die Daten außerhalb der Tools nicht sinnvoll untersuchen. Wenn Sie Daten in eine Tabelle, ein Grafikprogramm, eine Datenbank, ein Textverarbeitungsprogramm exportieren möchten, können Sie die Daten entweder [als Text speichern](#) oder in die [Zwischenablage](#) kopieren.

nur Teile von Daten speichern

Standardmäßig werden mit  und  alle Daten gespeichert, die Sie nicht [entfernt](#) haben. Wenn Sie nur Teile der Daten speichern, aber die restlichen Daten nicht entfernen möchten, wählen Sie [Kopieren](#).

5.31 suchen und ersetzen

Einige Listen, wie zum Beispiel eine Liste von [Dateinamen](#), ermöglichen die Funktion "Suchen und Ersetzen". Wie bei einem Suchvorgang, wird die *aktuelle Spalte* mit Daten durchsucht. Wenn "Groß-/Kleinschreibung beachten" aktiviert ist, muss jeder Buchstabe bezüglich Groß- und Kleinschreibung übereinstimmen. Sie haben die Möglichkeit vom nächsten Eintrag "Abwärts" bis zum Ende oder "Aufwärts" bis zum ersten Eintrag zu suchen. Sie können auch "jede Änderung bestätigen" auswählen.

Ganzes Wort - oder Wildcardsuche

"Ganzes Wort" bedeutet, nach Wörtern mit [Worttrennzeichen](#) am Anfang und am Ende zu suchen. Um einen Suffix oder Präfix zu suchen, verwenden Sie ein Sternchen. Mit ***ung** wird jeder Eintrag gefunden, der auf **-ung** endet; mit **un*** wird jedes Wort gefunden, das mit **un-** beginnt. ***buch*** wird jedes Wort mit **buch** finden (**buchen**, **Textbuch**, **gebucht**, oder **buch** in einem Eintrag mit zwei oder drei Wörtern, wie z. B. **das Hotel buchen**). Wortlisten können nach Suffix sortiert werden: siehe [Sortieren mit WordList](#).

Siehe auch: [Suchen durch Eingabe](#), [Suchen mit F12](#), [Akzentzeichen und Symbole](#).

5.32 suchen durch Eingabe

Wenn eine angezeigte Spalte alphabetisch sortiert ist, können Sie ein Wort schnell durch die Eingabe des Worts finden. Während des Tippens nähert sich **WordSmith** dem Wort. Wenn Sie die ersten fünf Buchstaben eingetippt haben und **WordSmith** eine Übereinstimmung gefunden hat, hören Sie einen Piepton und das Fenster "Bearbeiten" wird geschlossen. Zu diesem Zeitpunkt sollten Sie das Wort bereits sehen.

Siehe auch: [Modus "Bearbeiten" vs. Modus "Eingabe"](#), [Nach einem Wort bzw. Wortteil suchen](#), [Suchen und Ersetzen](#), [WordList-Sortierung](#)

5.33 nach einem Wort bzw. Wortteil suchen

Alle Listen ermöglichen die Suche nach einem Wort, Wortteil oder einer Zahl. Bei dieser Suche wird die *aktuelle Spalte* mit Daten durchsucht. Wenn "Groß-/Kleinschreibung beachten" aktiviert ist, muss jeder Buchstabe bezüglich Groß- und Kleinschreibung übereinstimmen. Sie haben die Möglichkeit vom nächsten Eintrag "Abwärts" bis zum Ende oder "Aufwärts" bis zum ersten Eintrag zu suchen.

Ganzes Wort - oder Wildcardsuche

"Gesamtes Wort" bedeutet, nach Wörtern mit [Worttrennzeichen](#) am Anfang und am Ende zu

suchen. Um einen Suffix oder Präfix zu suchen, verwenden Sie ein Sternchen. Mit ***ung** wird jedes Wort gefunden, das auf **-ung** endet; mit **un*** wird jedes Wort gefunden, das mit **un-** beginnt. ***buch*** wird jedes Wort mit **buch** finden (**buchen**, **Textbuch**, **gebucht**, oder **buch** in einem Eintrag mit zwei oder drei Wörtern, wie z. B. **das Hotel buchen**). Wortlisten können nach Suffix sortiert werden: siehe [Sortieren mit WordList](#).

Siehe auch: [Suchen durch Eingabe](#), [Suchen und Ersetzen](#), [Akzentzeichen und Symbole](#).

5.34 Dateinamen anzeigen

Diese Schaltfläche öffnet ein neues Fenster, in dem der [Dateiname](#) der Datei angezeigt wird, aus der die aktuellen Daten stammen. Falls notwendig, können Sie diese Namen bearbeiten (z. B. wenn die Textdateien verschoben oder umbenannt wurden). Dazu wählen Sie "Ersetzen" ().

Später, wenn Sie [die Ergebnisse speichern](#), wird die Information dauerhaft gespeichert.

Im Falle der Schlüsselwortlisten stammen die Daten aus einer Wortliste. Basiert die Wortliste auf einer Textdatei, ist der Name der Textdatei sichtbar, basiert sie jedoch auf mehreren Textdateien, sieht man den Namen der Wortlistendatei: um mehrere Original-Textdateinamen zu sehen, können Sie die Wortliste öffnen und darin die Schaltfläche "Dateiname" anklicken.

Siehe auch: [Quelldateien finden](#).

5.35 Stopplisten

Stopplisten sind Listen von Wörtern, die Sie nicht in die Analyse einbeziehen möchten. Zum Beispiel möchten Sie möglicherweise eine Wortliste erstellen oder Schlüsselwörter analysieren, und dabei häufige Funktionswörter wie *der*, *von*, *war*, *ist*, *es* ausschließen.

Um Stopplisten zu verwenden, müssen Sie zuerst mit dem Windows **Editor** oder einem beliebigen Standardtext-Editor eine Datei erstellen, in der alle Wörter festgelegt sind, die ignoriert werden sollen. Trennen Sie alle Wörter mit einem Komma oder schreiben Sie jedes Wort in eine neue Zeile. Sie können Großbuchstaben oder Kleinbuchstaben verwenden. Sie können mit einem Semikolon Kommentarzeilen einfügen.

Es gibt eine Datei mit dem Namen **stoplist.stp** (in Ihrem **\wsmith4** Verzeichnis), die Sie als Grundlage verwenden und unter einem neuen Namen speichern können.

Beispiel

```
; Meine Stoppliste für Testzwecke.
DER, DIE, DAS
IT
WIRD
```

Im Menü wählen Sie danach *Stoppliste*, um die Stoppliste(n) festzulegen, die Sie verwenden möchten. Für die Programme **WordList** und **KeyWords** können unterschiedliche Stopplisten verwendet werden. Ist die Stoppliste *aktiviert*, ist sie wirksam: das bedeutet, dass Wörter in der Stoppliste nicht in eine Wortliste aufgenommen werden. Wenn Sie jedes Mal dieselbe Stoppliste (n) verwenden möchten, können Sie diese Liste(n) in **wshell.ini** als [Standardeinstellung](#) festlegen.

Siehe [Matchliste](#) für eine genauere Beschreibung mit Screenshots.

Sie können auch eine Stoppliste erstellen, indem Sie basierend auf einem großen Textkorporus mit **WordList** eine Wortliste erstellen und die Mindestfrequenz sehr hoch einstellen, um nur sehr häufige Wörter zu erhalten. Dann speichern Sie die Wortliste als Textdatei. Danach formatieren Sie die Datei mit **Text Converter** um, und verwenden dazu als [Konvertierungsdatei](#) `stoplist.cod`.

Siehe auch: [Eine Tagdatei erstellen](#), [Matchliste](#), [Lemmatisierung](#).

5.36 Prozess unterbrechen

Während WordSmith die Textdateien verarbeitet oder Daten neu sortiert, sehen Sie im Controller eine Fortschrittsanzeige mit horizontalen Leisten, die den Fortschritt anzeigen. Gegebenenfalls wird auch die Schaltfläche "Abbrechen" angezeigt. Wenn Sie diese Schaltfläche anklicken, haben Sie vier Möglichkeiten:

Fortfahren

Vorgang unverändert fortsetzen

Diese Datei fertig stellen, danach anhalten

eine kontrollierte Unterbrechung. Die "Datei fertig stellen" bedeutet, dass Sie kontrollieren können, welche Dateien bereits verarbeitet wurden, und an welcher Stelle der Vorgang abgebrochen wurde. (Wie? Indem Sie die Dateinamen in der Datei, die Sie gerade erstellen wollten, also eine Wortliste, Konkordanz, etc., prüfen.)

Jetzt stoppen

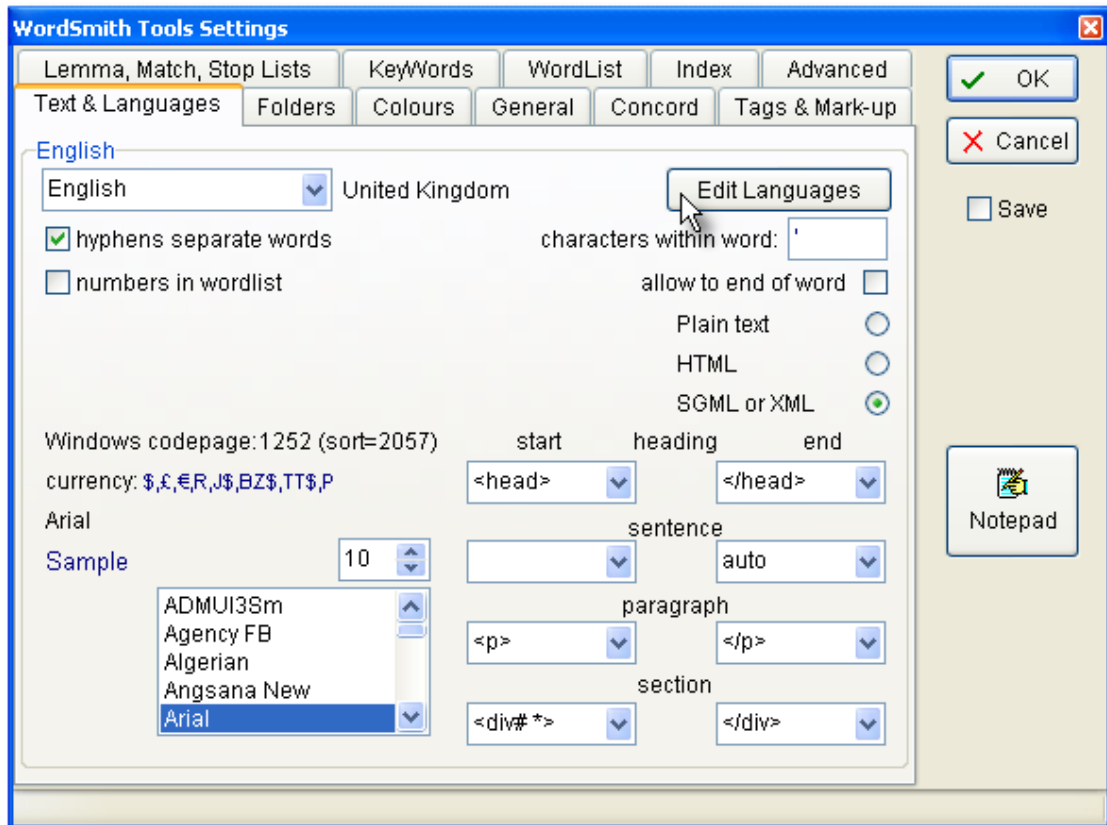
eine weniger kontrollierte Unterbrechung, die besonders bei großen CD-ROM-Dateien nützlich sein kann. WordSmith unterbricht die Verarbeitung der aktuellen Datei sofort, behält aber alle Daten, die bis zu diesem Zeitpunkt gesammelt wurden.

Dieses Tool beenden

ein Notstopp. Das Tool (Concord, WordList, etc.) wird geschlossen und [Speicher](#) für Systemressourcen kann teilweise ungenutzt sein. Der [Controller](#) wird nicht geschlossen.

Klicken Sie erneut auf *Abbrechen*, um Ihre Auswahl zu bestätigen.

5.37 Text und Sprachen



Diese Einstellungen beeinflussen die Verarbeitung Ihrer Texte mit WordSmith. Oben sehen Sie Felder, in denen Sie die Sprachfamilie (z. B. Englisch) und Varietäten der Sprache (UK, Australien, etc.) auswählen können. Diese Auswahl wird von den Einstellungen bestimmt, die Sie zuvor festgelegt haben. Erwartungsgemäß werden Sie nur mit ein paar bevorzugten Sprachen arbeiten, und deshalb können Sie diese Einstellungen einmal festlegen und werden sie später nicht mehr benötigen. Sie nehmen diese Einstellungen vor, indem Sie auf die Schaltfläche [Sprachen bearbeiten](#) klicken.

Diese unteren Einstellungen können für jede Sprache unterschiedlich sein:

Bindestriche und Zahlen

Sie können festlegen, ob Bindestriche als Worttrennzeichen gelten. Wenn das Kontrollkästchen "Bindestriche trennen Wörter" aktiviert ist [X], wird **Midlife-Crisis** als zwei Wörter behandelt.

Sollen Zahlen in Wortlisten wie normale Wörter aufgenommen werden? Wenn Sie das Kontrollkästchen "Zahlen in Wortlisten" nicht aktivieren, werden Wörter wie \$300, 90er oder mp3 in Wortlisten, Schlüsselwörter, Konkordanzen, etc. ignoriert und durch ein # ersetzt. Wenn Sie das Kontrollkästchen aktivieren, werden Wörter, die Zahlen beinhalten, inkludiert.

Zeichen innerhalb eines Worts

WordSmith inkludiert automatisch diejenigen Zeichen als alphabetische Symbole, die vom Betriebssystem als alphabetische Zeichen für die ausgewählte Sprache festgelegt sind. Also für Englisch, A bis Z und übliche Akzentbuchstaben wie é. Für Arabisch oder Japanisch gelten alle Zeichen, die Microsoft als alphabetische Zeichen festgelegt.

Sie möchten möglicherweise bestimmte zusätzliche Zeichen innerhalb eines Worts zulassen. Zum Beispiel wird auf Englisch der Apostroph in **father's** am besten als gültiges Zeichen innerhalb eines Worts festgelegt, da so das gesamte Wort verarbeitet und nicht in der Mitte abgeschnitten wird. (Wenn Sie die Sprache auf Deutsch einstellen, möchten Sie vielleicht nicht, dass Apostrophe als akzeptierte Zeichen in der Mitte eines Worts gezählt werden.)

Beispiele:

' (nur Apostrophe in der Wortmitte zulassen)

'% (sowohl Apostrophe und Prozentzeichen in der Wortmitte zulassen)

'_ (sowohl Apostrophe und Unterstriche in der Wortmitte zulassen)

Sie können bis zu 10 Zeichen festlegen.

Wenn Sie auch **fathers'** zulassen möchten, aktivieren Sie das Kontrollkästchen *am Wortende zulassen*. Wenn dieses Kontrollkästchen aktiviert ist, werden alle diese Symbole an Anfang und Ende eines Worts zugelassen, solange das Zeichen nicht alleine steht (wie in " ' ")

Standardtext/HTML/SGML

Das Format Ihrer Texte ist möglicherweise [Standardtext](#): das Standardformat. Wenn Ihre Texte in [HTML](#), [SGML](#) oder [XML](#) sind und entsprechende [Tags](#) enthalten, sollten Sie einer der Optionen hier auswählen. Dadurch können die Tools Satz-, Absatz- und Überschriftenmarkierungen optimal nutzen.

Windows-Format, etc.

Informationen über Windows [Zeichensätze](#) für die Sprache, mit der Sie arbeiten.

Anfang und Ende von Überschriften

Damit die Tools Überschriften zählen können, müssen Anfang und Ende einer Überschrift markiert werden. Wenn Ihr Text [Tags](#) enthält, z. B. **<h1>** und **</h1>**, geben Sie hier **<h#>** und **</h#>** ein. (# steht für eine beliebige Ziffer, ## für zwei Ziffern, etc.) Bei den von Ihnen eingegebenen Zeichen werden Groß- und Kleinschreibung beachtet: **</H#>** ist nicht dasselbe wie **</h#>**. (Wenn Sie einen [HTML](#)-Text haben, der nicht konsistent ist und manchmal **</h1>** und manchmal **</H1>** verwendet, verwenden Sie [Text Converter](#), um Ihr Text konsistent zu machen).

Anfang und Ende von Abschnitten

Wenn diese Felder z. B. **<div#>** und **</div>** enthalten, werden von den Tools damit Abschnitte identifiziert. Auch hier wird Groß- und Kleinschreibung beachtet.

Anfang und Ende von Sätzen

Wenn dieses Feld das Wort **auto** enthält, werden die Sätze von Tools wie [definiert](#) erkannt (Ende bei Punkt, Fragezeichen oder Ausrufezeichen, jeweils gefolgt von einem Großbuchstaben), aber wenn Ihr Text [Tags](#) enthält, z. B. **<s>** und **</s>**, geben Sie diese hier ein. Auch hier wird Groß- und Kleinschreibung beachtet.

Anfang und Ende von Absätzen

Damit die Tools Absätze erkennen können, müssen Anfang und/oder Ende des Absatzes markiert sein, z. B. zwei Zeilenschaltungen hintereinander (der Autor hat an dieser Stelle zweimal die **Eingabetaste** gedrückt) oder ein Zeilenumbruch gefolgt von einem Tabstopp (der Autor hat zuerst die **Eingabetaste** und danach die **Tabulatortaste** gedrückt). Wenn Ihr Text [Tags](#) enthält, z. B. **<p>** und **</p>**, geben Sie hier **<p>** und **</p>** ein. Auch hier wird Groß-/Kleinschreibung beachtet.

In vielen Fällen ist es ausreichend, das Absatzende zu definieren, da alles vor einem Absatzende Teil des vorangehenden Absatzes ist. Viele HTML-Texte unterscheiden nicht streng zwischen Absatzanfang und Absatzende.

Beachten Sie, dass gesprochene Text im BNC **</u>** anstelle von **</p>** verwenden, aber Sie können **</p>** beibehalten, da in diesem Fall WordSmith automatisch **</u>** verwendet, wenn **</p>** im Text nicht vorkommt.

welche Dateien

. bedeutet alle Textarten anzeigen. Sie können Textdateien auch z. B. auf *.txt beschränken.

Siehe auch: [Texte mit Tags](#), [Stopplisten](#), [Eine neue Sprache auswählen](#), [Verarbeitung von Texten auf Chinesisch, etc.](#)

5.38 Window Management

Wenn Oxford WordSmith Tools ausgeführt wird, befindet sich der [Controller](#) auf Ihrem Bildschirm links oben, und nimmt etwa ein Viertel des Bildschirms ein. Mit Ausnahme von Viewer und Concord ist das Hauptfenster jedes Tools gleich groß und wird rechts vom Controller geöffnet. Jedes Tool-Hauptfenster wird knapp unter den vorhergehenden Fenstern geöffnet. Einzelne Ergebnisfenster in jedem Tool werden innerhalb des Tool-Hauptfensters angezeigt, und können nebeneinander oder überlappend angeordnet werden. Verwenden Sie die Taskleiste (oder Alt+Tabulator, womit Sie einfach zwischen den Fenstern wechseln können).

"Ein neues Concord-Fenster öffnen"?

Diese Meldung erscheint, wenn Sie bereits ein Datenfenster geöffnet haben und *Neu* anklicken, um eine neue Konkordanz zu erstellen. Sie können eine beliebige Anzahl an Fenstern für jedes Tool geöffnet haben, jedes davon mit unterschiedlichen Daten.

Fenster minimieren, verschieben, Größe anpassen

Alle Fenster können vergrößert oder verkleinert werden, indem Sie den Cursor den Rand des Fensters positionieren und ziehen. Sie können leicht verschoben werden, indem Sie die Titelleiste (wo sich die Beschriftung befindet) mit der Maus an eine andere Stelle ziehen. Sie können ein Fenster minimieren: das Fenster wird dann zu einem Symbol; Sie können das Fenster wiederherstellen, indem Sie auf das Symbol klicken. Wenn Sie das Fenster maximieren, füllt es den gesamten Bildschirm des entsprechenden Tools aus. Dies sind Standard-Windowsfunktionen. Wenn Sie die einzelnen Tools verwenden, ist es unter Umständen günstig, den [Controller](#) zu minimieren.

Nebeneinander oder überlappend anordnen

Die aktiven Fenster werden in allen Haupttools unter dem Menüpunkt *Fenster* im Hauptmenü aufgelistet. Das aktive Fenster ist mit einem Häkchen versehen. Um ein anderes Fenster nach oben zu verschieben, klicken Sie einfach auf den Namen in der Liste.

Um eine Vielzahl verschiedener Fenster neu anzuordnen, können Sie die Fenster (bei fünf oder weniger Fenstern) *nebeneinander anordnen* (gleich große Fenster). Bei mehr als fünf Fenstern können Sie die Fenster *überlappend anordnen* (vertikal untereinander). Sie können im **Oxford WordSmith Tools Controller** auch *die Tools nebeneinander* oder *überlappend* anordnen.

Unübersichtlicher Bildschirm

Es passiert schnell, dass der Bildschirm unübersichtlich wird aufgrund zahlreicher Konkordanzen, jede davon mit geöffnetem Plot-, [Cluster](#)-, Kollokatoren- und Musterfenster. Beachten Sie, dass diese Fenster alle vom übergeordneten Fenster, der Konkordanz selbst, abhängen. Ein Schlüsselwort-Plot ist auch der Schlüsselwortliste untergeordnet. Sie können diese Fenster jederzeit schließen und erneut öffnen, solange das übergeordnete Fenster geöffnet ist.

Wenn Sie eine Konkordanz mit Kollokatoren und Mustern geöffnet haben, empfiehlt es sich, das Konkordanzfenster zu minimieren, und dann die Fenster nebeneinander anzuordnen. So werden Kollokatoren und Muster angezeigt, während die Konkordanz keinen Platz benötigt.

letzte Datei wiederherstellen

Eine Funktion für Bequemlichkeit: die letzte Datei, die Sie gespeichert oder geöffnet haben, wird standardmäßig wiederhergestellt, wenn Sie Oxford WordSmith Tools das nächste Mal

ausführen. Um einen unübersichtlichen Bildschirm zu vermeiden, wird nur eine Datei wiederhergestellt. Diese Funktion kann über eine Option in den Einstellungen temporär deaktiviert werden oder permanent in **wshe11.ini** (in Ihrem **\wsmith4** Verzeichnis).

5.39 nicht gewollte Zeilen entfernen

Um die richtige Reihenfolge Ihrer Daten nach mehrmaligem Bearbeiten wiederherzustellen, klicken Sie auf die Schaltfläche "Entfernen" ( oder drücken Strg+Z). So werden alle Zeilen, die Sie (durch Drücken der Taste Entf) entfernt haben, permanent entfernt, es sei denn, Sie haben die Daten (durch Drücken der Taste Einfg) wiederhergestellt.

Bei Wortlisten wird auch die gesamte Datei neu nach Frequenz geordnet. Alle entfernten Einträge sind zu diesem Zeitpunkt bereits unwiderruflich gelöscht. Diejenigen, die als Lemmata eines Hauptwortes festgelegt wurden, können vor oder nach dem Speichern noch angezeigt werden. Lemmata können jedoch nach dem Entfernen nicht mehr wiederhergestellt werden.

Siehe auch: [Daten auf N Einträge reduzieren](#).

Concord

Section



VI

6 Concord

6.1 Zweck



Concord is a program which makes a [concordance](#) using [DOS](#), [Text Only](#), [ASCII](#) or [ANSI](#) text files.

To use it you will specify a [search word](#), which Concord will seek in all the text files you have chosen. It will then present a concordance display, and give you access to information about collocates of the search word, dispersion plots showing where the search word came in each file, cluster analyses showing repeated clusters of words (phrases) etc.

The point of it...

The point of a concordance is to be able to see lots of examples of a word or phrase, in their contexts. You get a much better idea of the use of a word by seeing lots of examples of it, and it's by seeing or hearing new words in context lots of times that you come to grasp the meaning of most of the words in your native language. It's by seeing the contexts that you get a better idea about how to use the new word yourself. A dictionary can tell you the meanings but it's not much good at showing you how to use the word.

Language students can use a concordancer to find out how to use a word or phrase, or to find out which other words belong with a word they want to use. For example, it's through using a concordancer that you could find out that in academic writing, a *paper* can *describe*, *claim*, or *show*, though it doesn't *believe* or *want* (**this paper wants to prove that ...*).

Language teachers can use the concordancer to find similar patterns so as to help their students. They can also use Concord to help produce vocabulary exercises, by choosing two or three search-words, [blanking](#) them out, then [printing](#).

Researchers can use a concordancer, for example when searching through a database of hospital accident records, to see whether *fracture* is associated with *fall*, *grease*, *ladder*. Or to examine historical documents to find all the references to land ownership.

6.2 was ist eine Konkordanz

a set of examples of a given word or phrase, showing the context. A concordance of *give* might look like this:

```
... could not give me the time ...
... Rosemary, give me another ...
... would not give much for that ...
```

A concordancer searches through a text or a group of texts and then shows the concordance as output. This can be saved, printed, etc.

6.3 ausblenden

In a concordance, to blank out the search-words with asterisks, just press the spacebar (or choose *View | Blanked out*). Press it again to restore them.

The point of it...

A blanked-out concordance is useful when you want to create an exercise. This one has *give* and *put* mingled:

```
... 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 ...

```

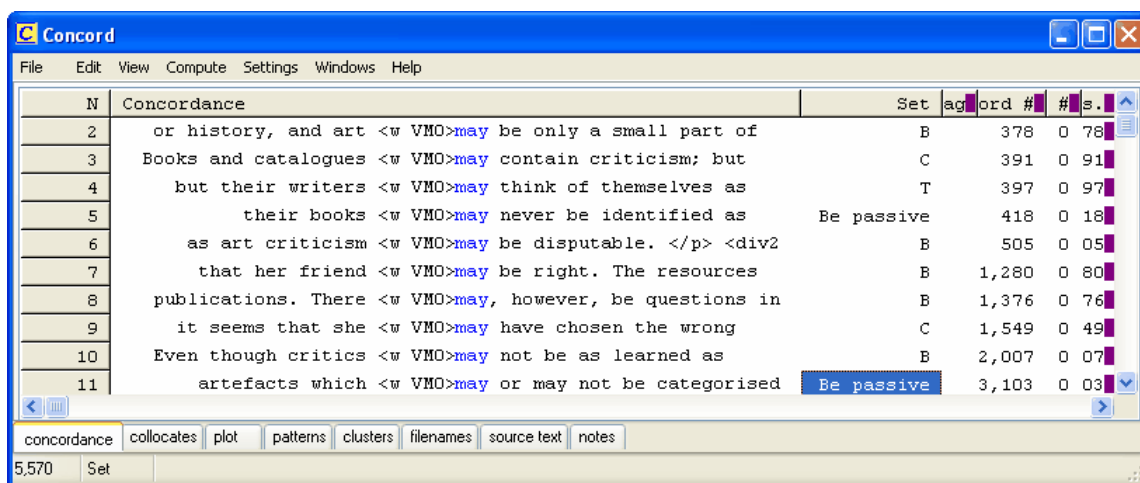
Concord will give equal space to the blanks so that the size of the blank doesn't give the game away.

See also : [Hide tags and other main Controller settings for Concord](#)

6.4 Kategorien

Wozu dient das?

Sie möchten möglicherweise Einträge mit einem eigenen System klassifizieren, z. B. indem Sie adjektivisch verwendete Wörter von nominal verwendeten Wörtern trennen, oder die Wörter entsprechend unterschiedlicher Bedeutungen sortieren.



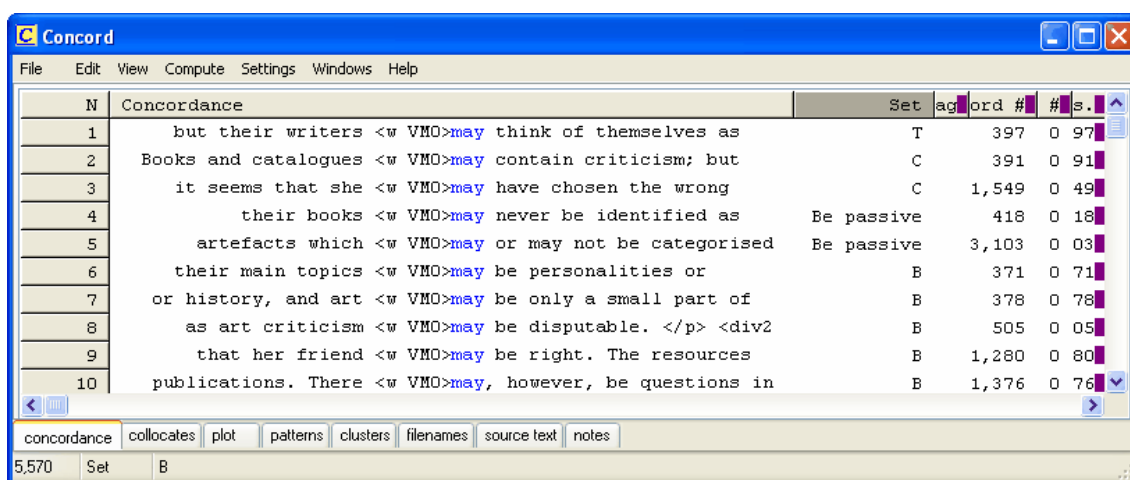
In diesem Beispiel hat der Benutzer den Buchstaben **B** für Einträge verwendet, bei denen auf das Wort **may** das Verb **BE** folgt. Es wird zwischen **BE** als Hauptverb bzw. **BE** als Hilfsverb in Passivkonstruktionen unterschieden; andere mit **BE** verwendete Verben werden nach ihrem Anfangsbuchstaben klassifiziert.

Wenn Sie sich im Modus Klasse (von [Eingabe vs. Klasse vs. Bearbeiten](#)) befinden (wie im Screenshot oben zu sehen ist) und Sie einen Buchstaben oder eine Ziffer eingeben, wird die Konkordanzzeile, in der Sie sich gerade befinden, mit diesem Zeichen markiert.

Wenn Sie eine längere Zeichenfolge eingeben möchten, klicken doppelt Sie auf die Spalte "Klasse" und Sie können mehrere Zeichen eingeben.

Um einen Fehler zu korrigieren, drücken Sie die Leertaste.

Sie können die Konkordanzzeilen später anhand dieser Kategorien [sortieren](#), indem Sie auf die Spaltenüberschrift *Klasse* klicken (siehe Screenshot unten).



The screenshot shows the Concord software window with a menu bar (File, Edit, View, Compute, Settings, Windows, Help) and a toolbar. The main window displays a concordance table for the search term 'may'. The table has columns for line number (N), concordance text, Set, tag, word, frequency (#), and score (s.). Below the table is a toolbar with buttons for concordance, collocates, plot, patterns, clusters, filenames, source text, and notes. The status bar at the bottom shows '5,570' and 'Set B'.

N	Concordance	Set	tag	word	#	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

Siehe auch: [Modus "Bearbeiten" vs. Modus "Eingabe"](#).

6.5 Kollokationsbereich

The collocate horizons represent the number of collocates Concord will find to the left and right of your search word, and the distance used by **KeyWords** in searching out [plot-links](#). The [default](#) is 5,5 (5 to left and 5 to right) but you can go up to 25 on either side.

To set collocation horizons and other **Concord** settings, in the main **WordSmith Controller** menu at the top, choose *Adjust Settings*, then *Concord*.

See also: [Collocate Settings](#)

6.6 Kollokationseinstellungen

To set collocation horizons and other **Concord** settings, in the main **WordSmith Controller** menu at the top, choose *Adjust Settings*, then *Concord*.

Minimum Specifications

The minimum length is 1, and minimum frequency is 1 (default is 10). You can specify here how frequently it must have appeared in the neighbourhood of the Search Word. Words which only come once or twice are less likely to be informative. So specifying 5 will only show a collocate which comes 5 or more times in the neighbouring context.

Similarly, you can specify how long a collocate must be for it to be stored in memory, e.g. 3 letters or more would be 3.

Horizons

Here you specify how many words to left and right of the Search Word are to be included in the collocation search: the size of the "neighbourhood" referred to above. The maximum is 25 left and 25 right. Results will later show you these in separate columns so you can examine exactly how many times a given collocate cropped up say 3 words to the left of your Search Word.

The most frequent will be signalled in **red**.

Breaks

These are

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

which you will see in the bottom right corner of the screen visible in the [Controller Concord Settings](#).

When the collocates are computed, if the setting is to stop at sentence breaks, collocates will be counted within the above horizons but taking sentence breaks into account.

For example, if a concordance line contains

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

and the search-word is **however**,

only

when we compared these two

will be used for collocates because there is a sentence break to the left of the search word. If the setting is "stop at punctuation", then nothing will come into the collocate list for that line (because there is a more major break than punctuation to the left of it, and no word to the right of the search-word before a punctuation symbol).

6.7 Kollokatoren-Anzeige

The point of it...

The point of all this is to work out characteristic lexical patterns. It can be hard to see overall trends in your concordance lines, especially if there are lots of them. By examining collocations in this way you can see common lexical and grammatical patterns of co-occurrence. Collocational linkages which involve grammatical items are often referred to as *colligation*.

Display

The collocation display initially shows the collocates in frequency order.

Beside each word and the search-word which the concordance was based on, you'll see the [strength of relationship](#) between the two (or 0.000 if it hasn't yet been computed).

Then, the total number of times it co-occurred with the search word in your concordance, and a total for Left and Right of the search-word. Then a detailed break-down, showing how many times it cropped up 5 words to the left, 4 words to the left, and so on up to 5 words to the right. The centre position (where the search word came) is shown with an asterisk.

The number of words to left and right depends on the [collocation horizons](#).

The numbers are:

- the total number of times the word was found in the neighbourhood of the search word
- the total number of times it came to the left of the search-word
- the total number of times it came to the right of the search-word
- a set of individual frequencies to the left of the search word (5L, i.e. 5 words to the left, 4L .. 1L)
- a Centre column, representing the search-word
- a set of individual frequencies to the right of the search word (1R, 2R, etc.)

The number of columns will depend on the collocation word horizons. With 5,5 you'll get five columns to the left and 5 to the right of the search word. So you can see exactly how many times

each word was found in the general neighbourhood of the search word and how many times it was found exactly 1 word to the left or 4 words to the right, for example. The most frequent will be signalled in **red**. In the screenshot below, **differences** comes 44 times in total but **39** of these are in 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

The screenshot above shows collocation results for a concordance of **BETWEEN/AMONG** sorted by the *Relation* column, where items like **differentiate**, **difference** etc. are found to be most strongly related to **between**. Further down the listing, some links concerning **among** (**growing**, **refugees**) are to be seen.

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

The frequency display can be [re-sorted](#) (✖) and you can recalculate the collocates (📊) if you [zap](#) entries from the concordance or change the [horizons](#).

See also: [Collocation](#), [Collocation Relationship](#), [Mutual Information](#)

6.8 Verbindung zwischen Kollokatoren

The point of it...

The idea is to find out how strongly each collocate relates to the search-word near which it was found.

Choose a wordlist

The collocation display initially shows the collocates without any relation information (eg. 0.00); this is because to compute the relationship we need to know

- a) how often each collocate appears in the corpus we're using,
- b) how often the [search-word](#) appears in the corpus, and
- c) how often they come together within the [horizons](#) selected.

The problem is that although b) and c) are known at the time the concordance is computed, a) is not known without doing a concordance or wordlist for each collocate....

So you are asked to choose an appropriate wordlist (created by WordSmith of course!), which will know the frequencies for each word. It's up to you to choose a wordlist which actually relates to the concordance you've done!

Type of Relation

Choose in the main [Controller Concord settings](#) which type of relation you wish to compute.

See also: [Collocation](#), [Collocate display](#), [Mutual Information](#)

6.9 Kollokation

What's a "collocate"?

Collocates are the words which occur in the neighbourhood of your search word. Collocates of *letter* might include *post*, *stamp*, *envelope*, etc. However, very common words like *the* will also collocate with *letter*.

The point of it...

By examining the collocates you can find out more about "the company the word keeps", which helps to show its meaning and its usage.

Options

You may compute a concordance with or without collocates: without is slightly quicker and will take up less room on your hard disk. The default is to compute with collocates.

The number of collocates stored will depend on the [collocation horizons](#).

You can re-compute collocates after editing your concordance.

If you want to filter your collocate list, use a [match list](#) or [stop list](#).

[Re-sort](#) a collocate list in a variety of ways.

You can see the [strength of relationship](#) between the word and the search-word which the concordance was based on.

Collocates can be [viewed](#) after the concordance has been computed.

Technical Note

The [literature](#) on collocation has never distinguished very satisfactorily between collocates which we think of as "associated" with a word (letter - stamp) on the one hand, and on the other, the words which do actually co-occur with the word (letter - my, this, a, etc.).

We could call the first type "coherence collocates" and the second "neighbourhood collocates" or "horizon collocates". It has been suggested that to detect coherence collocates is very tricky, as once we start looking beyond a horizon of about 4 or 5 words on either side, we get so many words that there is more noise than signal in the system.

KeyWords allows you to study [Associates](#), which are a pointer to "coherence collocates".

Concord will supply "neighbourhood collocates". **WordList** allows you also to study [Mutual](#)

[Information](#).

See also: [collocation display](#), [collocation settings](#), [collocation relationship](#), [mutual information display](#).

6.10 Concord: Cluster

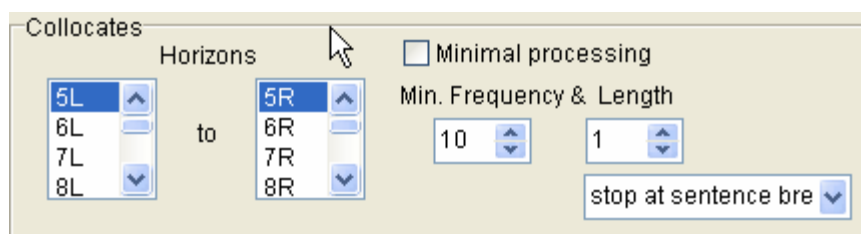
The point of it...

These word clusters help you to see patterns of repeated phraseology in your concordance, especially if you have a concordance with several thousand lines. Naturally, they will usually contain the search-word itself, since they are based on concordance lines.

Another feature in **Concord** which helps you see patterns is [Patterns](#).

How it does it...

Clusters are computed automatically if this is not disabled in the main [Controller](#) settings for Concord (*Adjust Settings | Concord*) where you will see something like this:



where your usual default settings are controlled. "Minimal processing", if checked, means do not compute collocates, clusters, patterns etc. when computing a concordance. (They can always be computed later if the source text files are still present.)

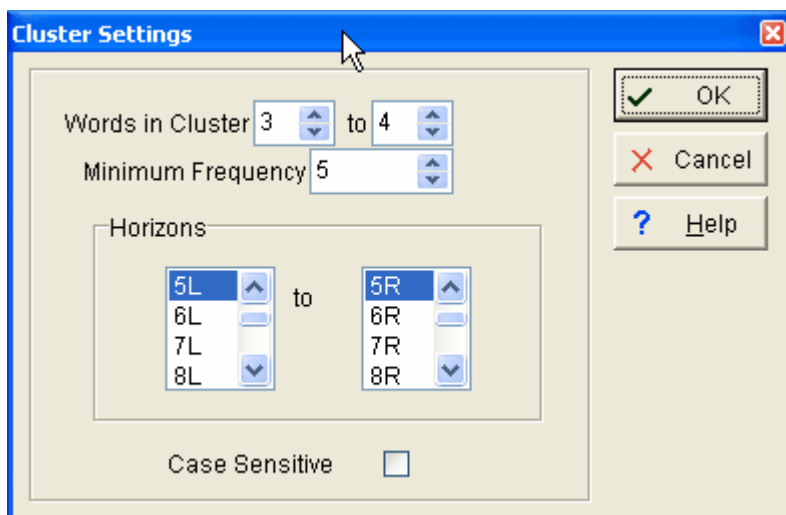
Clusters are sought within these limits: default: 5 words left and right of the search word, but up to 25 left and 25 right allowed. The default is for clusters to be three words in length and you can choose how many of each must be found for the results to be worth displaying (say 3 as a minimum frequency).

Clusters are calculated using the existing concordance lines. That is, any line which has not been deleted or zapped is used for computing clusters.

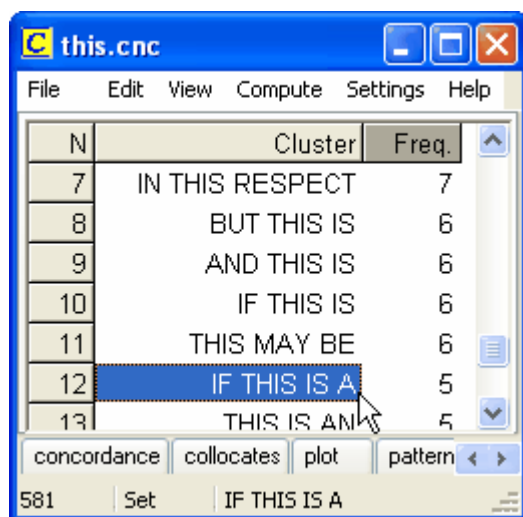
As with [WordList index clusters](#), the idea of "stop at sentence breaks" (there are other alternatives) is that a cluster which spans across two sentences is not likely to make sense.

Re-computing clusters

The default clusters computed may not suit, (and you may want to recompute after deleting some lines), so you can also choose *Compute | Clusters*  in the Concord menu, so as to choose how many words a cluster should have (cluster size 2 to 4 words recommended), and alter the other settings.



When you press OK, clusters will be computed. In this case we have asked for 3- or 4-word clusters and get results like this:



Most of the clusters are 3-word clusters but the highlighted one has 4 words here. Line 10 shows there are 6 cases of **if this is**, and line 12 shows that most of them are **if this is a**.

It's a dependent window

Each Cluster window is dependent on the Concordance from which it was derived. If you close the original concordance down, they will disappear.

See also: [general information on clusters](#), [WordList Clusters](#).

6.11 Concord: Verteilung

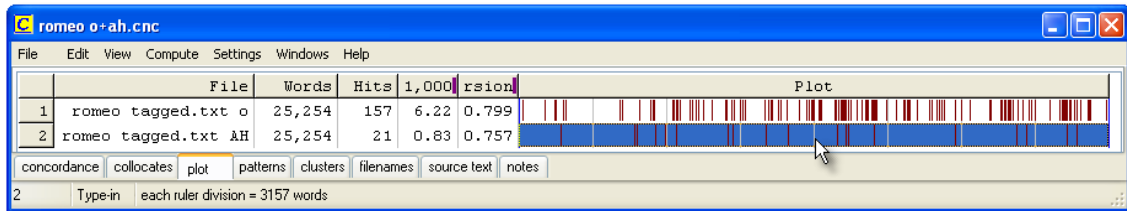
The point of it...

This shows where the search word occurs in the file which the current entry belongs to. That way you can see where mention is made most of your search word in each file.

What you see

The plot shows:

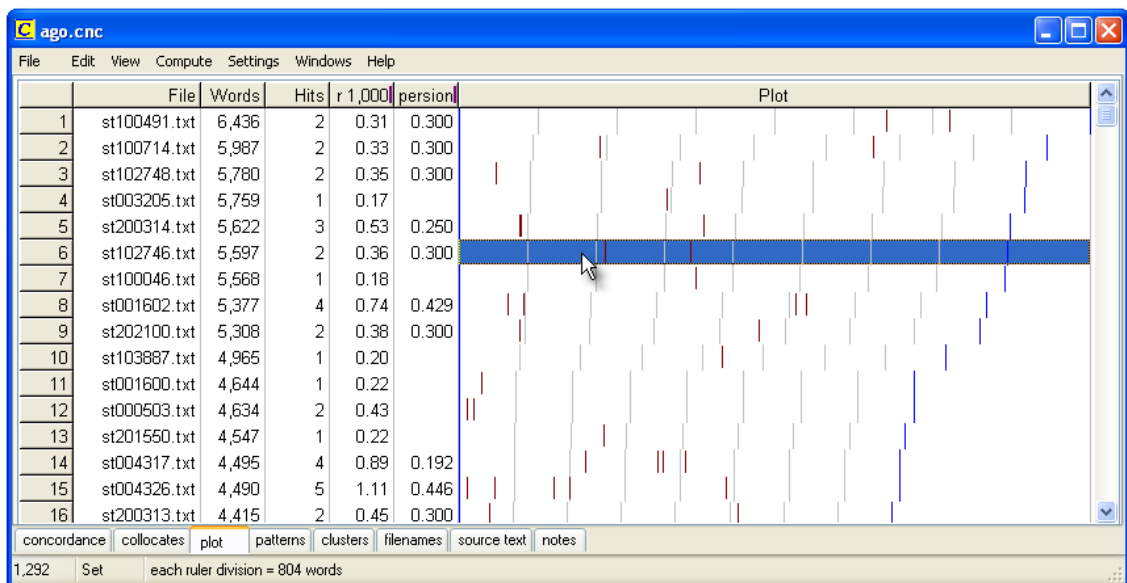
File source text file-name
Words number of words in the source text
Hits number of occurrences of the search-word
per 1,000 how many occurrences per 1,000 words
Dispersion the plot [dispersion value](#)
Plot a plot showing where they cropped up, where the left edge of the plot represents "Once upon a time" and the right edge is "happily ever after".



Here we see a plot of "O" and another of "AH" from the play Romeo and Juliet. They are on 2 separate lines because there were 2 search-words. There are more "O" exclamations than "AH"s. There is a "ruler" splitting the display into 8 segments, and the status bar tells us each segment represents about 3150 running words of the play.

The plot is initially [sorted](#) by no. of words per 1,000.

There are two ways of viewing the plot, the default, where all plotting rectangles are the same length, or *Uniform Plot* (where the plot rectangles reflect the original file size -- the biggest file is longest). Change this in the *View* menu at the top.



The screenshot shows "uniform plot" -- as the statusbar says, each ruler segment represents 800 words in these dispersion plots of "ago". If you look at the Words column, you will see that the number of words in each file varies, which is why the blue right plot edge and the ruler marks vary in position.

If you don't see as many marks as the number of hits, that'll be because the hits came too close together for the amount of screen space in proportion to your screen resolution. You can stretch the plot by dragging the top right edge of it. You can export the plot using [Save As](#) and can get your spreadsheet to make graphs etc, as [explained here](#).

Each plot window is dependent on the concordance from which it was derived. If you close the original concordance down, it will disappear. You can *Print* the plot. There's no *Save* option because the data come from a concordance which you should [Save](#), or *Print to File*. You can *Copy* to the [clipboard](#) (Ctrl-Ins) and then put it into a word processor as a graphic, using Paste Special.

See also: [plot and ruler colours](#), [plot dispersion value](#).

6.12 Concord: speichern und drucken

Sie können die Konkordanz (und deren Kollokatoren sowie andere dazugehörige Ergebnisse, die beim Erstellen der Konkordanz gespeichert wurden) sowohl als Textdatei (z. B. zum Importieren in ein Textverarbeitungsprogramm) als auch als Ergebnisdatei speichern, die Sie zu einem späteren Zeitpunkt im Hauptmenü *öffnen* und anzeigen können. Wenn Sie **Concord** verlassen, werden Sie aufgefordert, die Konkordanz zu speichern, falls Sie dies nicht bereits getan haben.

Wenn Sie eine Konkordanz speichern, können Sie diese später wieder aufrufen und Kollokatoren, Verteilungsplots und Cluster untersuchen.

Sie können eine Konkordanz mit dem Windows-Drucker [drucken](#), der an Ihr System angeschlossen ist. Sie haben die Möglichkeit, die Anzahl der Seiten festzulegen, die Sie drucken möchten. Die Schriftart ist ähnlich wie die am Bildschirm dargestellte Schriftart. Wenn Ihr Drucker in Farbe oder Graustufen drucken kann, wird Ihr Ausdruck an die Bildschirmfarben angepasst. Wenn Sie einen Schwarzweißdrucker haben, werden farbig markierte Elemente *kursiv* gesetzt, vorausgesetzt Ihr Drucker bietet diese Möglichkeit.

Concord kann so viele Konkordanzen inklusive dazugehöriger Informationen drucken, wie es Ihre Druckpapiereinstellungen erlauben. Die entsprechenden Seitenränder werden in der [Druckvorschau](#) angezeigt.

6.13 Concord: Anzeigeeoptionen

These menu options toggle on and off. When on, they're checked. They include:

Sentence Only

This will show only the sentence in which the search-word appears.

Tags and Spaces Cut

If you have specified any [tags to retain](#), these will normally be visible in your concordance. If you wish to hide them, toggle this menu option. The same option will also cut out any redundant spaces in your concordance line; these might be caused by the presence of [tags](#) which have been ignored.

See also: [showing nearest tags](#), [Blanking out](#) the search-word.

6.14 Concord: Audio und Video wiedergeben

The point of it

Suppose you do a concordance of "elephant" and want to hear how the word is actually spoken in context. Is the last vowel a schwa? Does the second vowel sound like "i" or "e" or "u" or a schwa?

How to do it...

If you have defined tags which refer to multimedia files, and if there are any such tags in the "tag-context" of a given concordance line, you can hear or see the source multimedia. To play the media file, press Control/M or choose *File | Play media file*. The tag will be [parsed](#) to identify the file needed, if necessary downloading it from a web address, and then played.

See also: [Handling Tags](#), [Making a Tag File](#), [Showing Nearest Tags in Concord](#), [Tag Concordancing](#), [Types of Tag](#), [Viewing the Tags](#), [Using Tags as Text Selectors](#), [Tags in WordList](#)

6.15 Concord: Anzeige und Funktionen

You have a listing showing all the concordance lines in a window. You can scroll up and down and left or right with the mouse or with the cursor keys. Click for information on the [menu](#).

The Columns

These show the details for each entry: the entry number, the concordance line, set, tag, word-position (e.g. 1st word in the text is 1), source text [filename](#), and how far into the file it comes (as a percentage). See below for an explanation of the **purple blobs**.

Set

This is where you can classify the entries yourself, using any letter, into [user-defined categories](#). Supposing you want to sort out verb uses from noun uses, you can press V or N. To type more (eg. "Noun"), double-click the entry in the set column and type what you want. If you have more than one [search-word](#), you will find the Set column filled with the search-word for each entry. To clear the current entry, you can type the number 0. To clear the whole Set column, choose *Edit | Clear Set column*.

Tag

This column shows the [tag context](#).

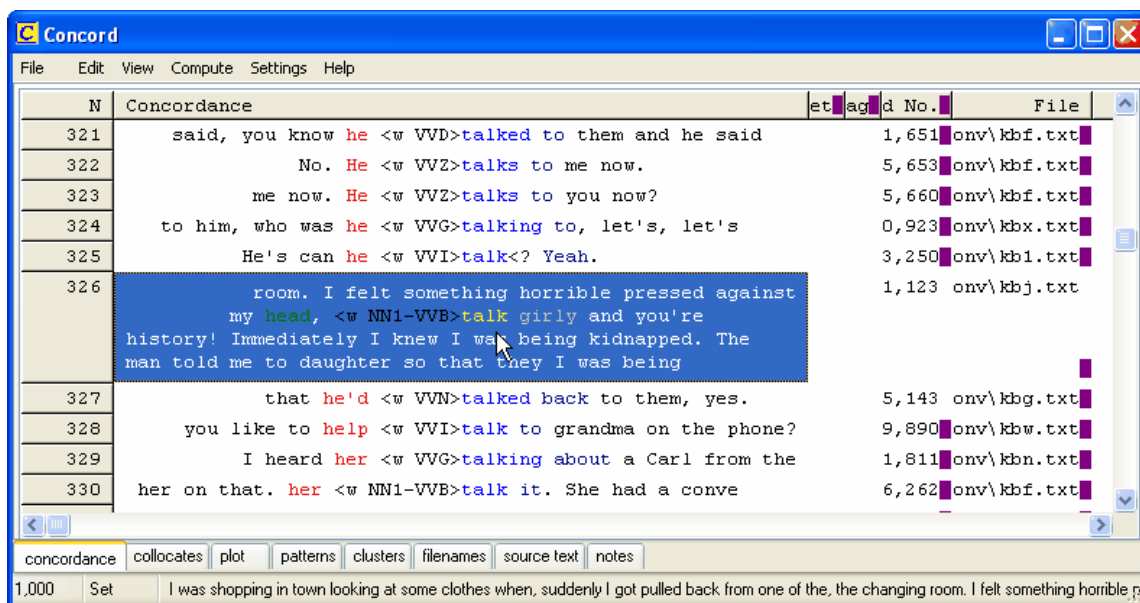
Stretching the display to see more

You can pull the concordance display to widen its column. Just place the mouse cursor on the bar between one column and another; when the cursor changes shape you can pull the whole column.

Stretch one line to see more context

The same applies to each individual row: place the mouse cursor between one row and another in the grey numbered area, and drag.

Or press  (F8) to "grow" all the rows, or  (Ctrl/F8) to shrink them. Or press the numeric key-pad 8 to grow the current line as shown below. (Use numeric key-pad 2 to shrink it.)



Purple marks

In the screenshot you will see purple marks where any column is not wide enough to show all the data. The reason is that numbers are often not fully visible and you might otherwise get the wrong impression. For example in the concordance below, the *Word #* column shows **4,569** but the true number might be **14,569**. Pull the column wider and the purple lines disappear.

	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

Viewing the original file (📄)

Available if it is still on the disk where it was when the concordance was originally created.

Status bar

The [status bar](#) panels show

- the number of entries (1,000 in the screenshot)
- whether we're in "Set" or "Edit" mode;
- the current concordance line from its start.

See also:

[Re-sorting](#) your concordance lines

[User-defined categories](#)

[Altering the View](#)

[Blanking out](#) the search-word

[Collocation](#) (words in the neighbourhood of the search-word)

[Plot](#) (plots where the search-word came in the texts)

[Clusters](#) (groups of words in your concordance)

[Text segments in Concord](#)

[Editing the concordance](#)

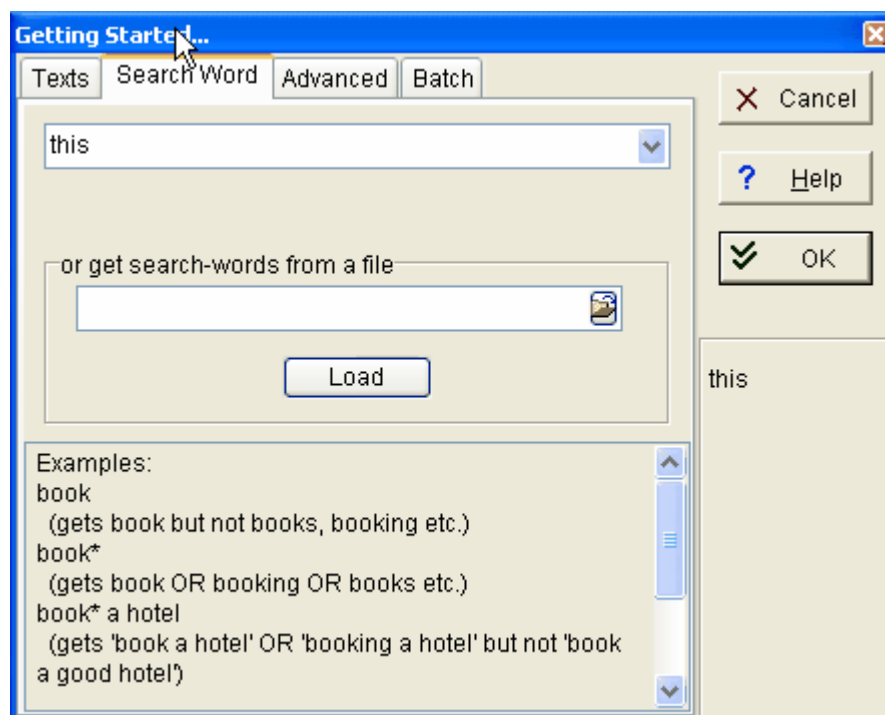
[Zapping entries](#)

[Saving and printing](#)

6.16 Konkordanzeinstellungen

Search Word or Phrase and/or Tag

Type the [word or phrase](#) Concord will search for when making the concordance, or the name of a [file of search words](#). You may also choose from a [history list](#) of your previous search words. For details of syntax, see [Search Word Syntax](#) or the set of examples shown in this screenshot:

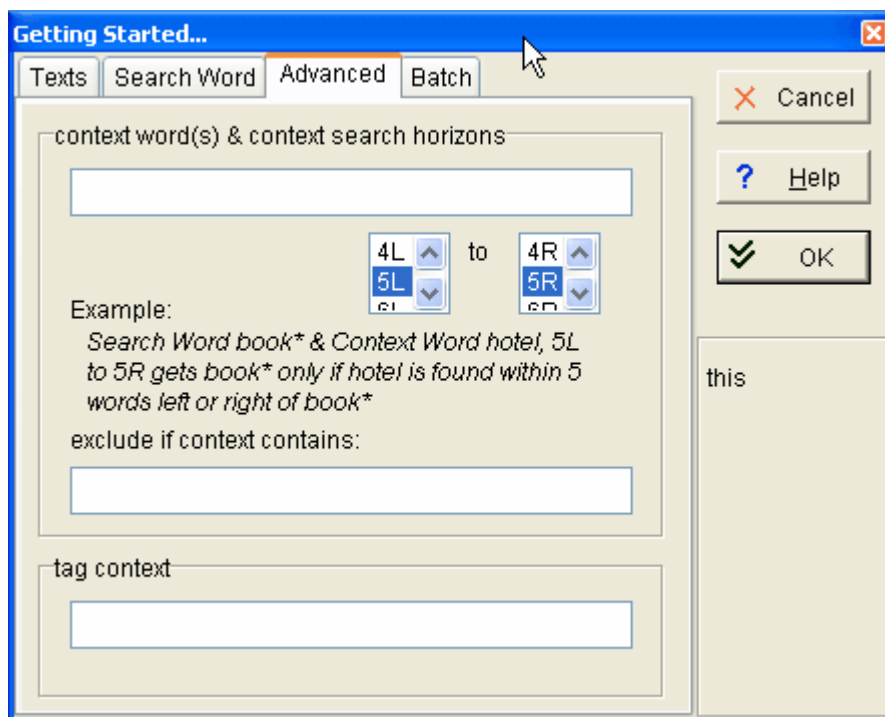


If you want to do many concordances in a [file-based search](#), first prepare a small text file containing the search words, e.g. containing

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

Press the file button to locate your text file, then press the *Load* button. This will then change its name to something like *Clear 4*, where 4 means as in the example above that there are 4 different search-words to be concordanced. See "Batch" below for details on saving each one under a separate filename, otherwise all the searches will be combined into the same concordance.

Advanced



Context Word(s) and Context Search Horizons

You may wish to find a word or phrase depending on the context. In that case you can specify context word(s) which you want, or which you do not want (and if found will mean that entry is not used).

For example, if the search word is **book*** and the *context* word is **hotel**, you'll get **book**, **books**, **booked**, **booking**, **bookable**, but only if **hotel** is found within your [Context Search Horizons](#). Or if the search word is **book*** and the *exclusion* word is **hotel**, you'll get **book**, **books**, **booked**, **booking**, **bookable**, as long as **hotel** is *not* found within your context search horizons. Or if the search word is **book*** and the exclusion word is **booked**, you'll get **book**, **books**, **booking**, **bookable**, but not **booked**.

Tag Context

The tag context is the context of tags as defined in your [tag-file](#), to the left of the search-word. In this example, the tag context is **<u speaker=Simon>**:

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

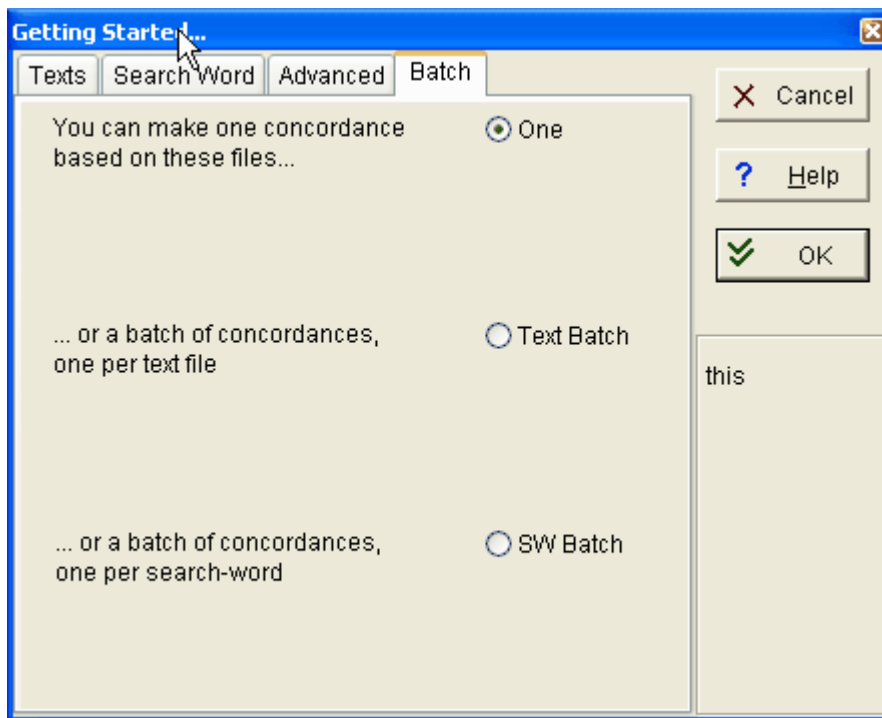
The tag context actually can include a bit more than that -- if this chunk is in **<section 1>**, and there is a tag showing the end of section 1 (probably **</section>**) which occurs anywhere to the right of the search-word, then the full tag context might be **<section 1><u speaker=Simon>** or even more. So a tag context is actually the whole set of tags which are still in operation at the point where your search-word was found.

You may wish to find a word or phrase depending on the tag context. In that case you specify tag attributes which you require.

For example, if the search word is **book*** and the tag context is **Simon**, you'll get **book**, **books**, **booking**, **bookable**, **booked** only if **Simon** is found in the tag context.

Batch

Suppose you're concordancing **book*** in 20 text files: you might want *One* concordance based on all 20 files (the default), or instead 20 separate concordances [in a zipped batch](#) which can be viewed separately (*Text Batch*). If you have multiple search-words in a [file-based search](#) as explained above, you may want each result saved separately (*SW Batch*).



Other settings affecting a concordance are available too: see [WordSmith Controller Concordance Settings](#); [Accented characters](#); [Choosing Language](#); [Context Word\(s\) & Context Search Horizons](#)

6.17 Konkordanz mit Tags

Wozu dient das?

Angenommen Sie möchten bestimmte Strukturen identifizieren (im Gegensatz zu einem vorgegebenen Wort oder einer Phrase), zum Beispiel drei Substantive in Folge (**Noun+Noun+Noun**). Sie können die Tags eingeben, für die Sie die Konkordanz erstellen möchten (mit oder ohne Wörtern).

Wie mache ich das?

Im Suchwortfeld in Concord geben Sie die Tags ein, die Sie suchen, oder Sie definieren die Tags in einer [Tagdatei](#).

Beispiele

<w NN1>table findet *table* als Substantiv im Singular (nicht als Verb).

<w NN1>* <w NN1>* findet alle Stellen im BNC Sampler, an denen zwei Gattungsnamen direkt hintereinander stehen.

Beachten Sie, dass **<w NN1>table** nur dann *table* findet, wenn Ihr Text mit Tags in spitzen Klammern (< und >) markiert ist. Wenn Sie [und] als Tagsymbole festgelegt haben, wird nach **[w NN1]table** gesucht.

Unter [Suchwort-Syntax](#) finden Sie weitere Beispiele dazu.

Es spielt hierbei keine Rolle, ob Sie eine [Tag-Datei](#) verwenden, da WordSmith Ihre Tags automatisch erkennt. (Das funktioniert natürlich nur dann, wenn Sie den Text, den Sie verwenden, zuvor mit Tags markiert haben.)

Im zweiten Beispiel wird ein Sternchen verwendet, da im BNC die Tags direkt vor dem Wort stehen, auf das sie sich beziehen: wenn Sie das Sternchen vergessen, geht das Programm davon aus, dass Sie ein Tag mit einem [Trennzeichen](#) vorne und hinten suchen.

Erstellen Sie eine Konkordanz für Tags?

Wenn diese Meldung angezeigt wird und Sie eine Konkordanz mit Tags erstellen möchten, klicken Sie auf "Ja". Wenn nicht, werden in Ihrem Suchwort die Symbole < oder > mit Anführungszeichen ("<") markiert, wie auch unter [Suchwort-Syntax](#) erklärt wird.

Groß-/Kleinschreibung beachten

Bei Tags wird nur Groß- und Kleinschreibung beachtet, wenn dies auch bei Suchwort oder Suchphrase der Fall ist. Bei Suchwörtern wird standardmäßig keine Groß- und Kleinschreibung beachtet. Also im ersten Beispiel erhalten Sie die Ergebnisse *table*, *Table* und *TABLE*, wenn diese als Substantive verwendet werden (aber Sie erhalten keine Ergebnisse, wenn Ihr Text nicht mit Tags markiert ist).

Siehe auch: [Tag-Übersicht](#), [Tag-Handling](#), [Nächstes Tags in Concord anzeigen](#), [Suchwörter oder Suchphrasen](#), [Tagtypen](#), [Tags anzeigen](#), [Tags als Selektoren verwenden](#)

6.18 Kontextwort

You may restrict a concordance search by specifying a context word which either must or may not be present within a certain number of words of your search word.

For example, you might have **book** as your search word and **hotel*** as the context word. This will only find **book** if **hotel** or **hotels** is nearby.

Or you might have **book** as your search word and **paper*** as an exclusion criterion. This will only find **book** if **paper** or **papers** is *not* within your Context Search Horizons.

Context Search Horizons

The context horizons determine how far Concord must look to left and right of the search word when checking whether the search criteria have been met. The [default](#) is 5,5 (5 to left and 5 to right of the search word) but this can be set to up to 25 on either side. 0,2 would look only to the right within two words of the search word.

If you have specified a context word, you can re-sort on it. Also, the context words will be in their own special [colour](#).

Syntax is like that of the [search word or phrase](#),

* means disregard the end of the word and can be placed at either end of your context word.

== means case sensitive

/ separates alternatives. You can specify up to 15 alternatives within an 80-character overall limit.

If you want to use *, ?, ==, ~, : \ or / as a character in your search word, put it in double quotes, e.g. "*"

6.19 Konkordanzen bearbeiten

The point of it...

You may well find you have got some entries which weren't what you expected. Suppose you have done a search for **SHRIMP*/PRAWN*** -- you may find a mention of *Shrimpton* in the listing. It's easy to clean up the listing by simply pressing **Del** on each unwanted line. (Do a sort on the search word first so as to get all the *Shrimptons* next to each other.) The line will turn a light grey colour.

Pressing **Ins** will restore it, if you make a mistake. To delete or restore ALL the lines from the current line to the bottom, press the grey - key or the grey + key by the numeric keypad. When you have finished marking unwanted lines, you can choose (**Alt-Z** or ) to [zap](#) the deleted lines. If you're a teacher you may want to [blank](#) out the search words: to do so, press the spacebar. Pressing the spacebar again will restore it, so don't worry!

See also: [Window Management](#)

6.20 dateibasierte Suchwörter

Wozu dient das?

Um bei der Eingabe mehrerer Suchwörter Zeit zu sparen.

Möglicherweise möchten Sie öfters eine Standardsuche in verschiedenen untergeordneten Korpora durchführen.

Da Concord eine unbegrenzte Anzahl von Einträgen ermöglicht, möchten Sie vielleicht eine Konkordanz durchführen, die viele [Suchwörter oder -phrasen](#) einbezieht.

Die Länge für mehrere Suchwörter ist auf 80 Zeichen (inklusive /, etc.) begrenzt. Wenn Ihre bevorzugten Suchwörter diese Länge übersteigen oder Sie eine standardisierte Suche durchführen wollen, können Sie eine Datei erstellen, die alle Suchwörter enthält.

Wie mache ich das?

Eine Beispieldatei (`\wsmith4\concordance_search_words.txt`) finden Sie bei den Installationsdateien.

Verwenden Sie ein Windows Textverarbeitungsprogramm (z. B. Editor), um Ihre eigene Datei zu erstellen. Jedes Wort muss in einer eigenen Zeile Ihres Dokuments stehen. Kommentarzeilen können nicht eingefügt werden, das Einfügen leerer Zeilen erhöht jedoch die Lesbarkeit.

Wenn Sie das Wort in einem bestimmten Kontext suchen, geben Sie **context:=** ein, zum Beispiel:

buchen context:=Hotel

Bei dieser Suche nach dem Wort **buchen** werden nur die Ergebnisse angezeigt, in deren Kontextbereich das Wort **Hotel** vorkommt.

Anstatt jedes Wort oder jede Phrase nun in dem Dialogfeld "Suchwort" einzugeben, laden Sie einfach die Datei.

Lemmata und dateibasierte Konkordanz

Wenn Sie Concord aus WordList [aufrufen](#) und das hervorgehobene Wort in der Wortliste ist der Haupteintrag der [Lemmata](#), dann wird eine temporäre Datei erstellt, in der alle Lemmata aufgelistet werden, und auf Grundlage dieser dateibasierten Suchwortliste erstellt Concord eine Konkordanz. Die temporäre Datei wird in Ihrem `\wsmith4`-Verzeichnis gespeichert, es sei denn Sie arbeiten auf dem Netzwerk. In diesem Fall wird die Datei in Ihrem temporären Windows-Verzeichnis gespeichert, z. B. `\windows\temp`. Sollten Sie die temporäre Datei nicht benötigen, können Sie diese löschen.

Automatische dateibasierte Konkordanz

Unter *SW Batch* [Konkordanzeinstellungen](#) kann Concord viele verschiedene Suchwörter gleichzeitig verarbeiten, und jedes Ergebnis wird während des Vorgangs automatisch gespeichert.

6.21 nächstes Tag

Mit Concord können Sie die am nächsten gelegenen Tags anzeigen, wenn Sie eine [Tagdatei](#) angegeben haben, in der Ihre bevorzugten Tags aufgeführt sind. Danach wird in einer geöffneten Konkordanz das Tag in einer der Spalten des Konkordanzfensters angezeigt.

Wozu dient das?

Der Vorteil ist, dass Sie in Ihrer Konkordanz sehen können, welche Beziehung zwischen dem Suchwort und dem Text in Tags besteht. Zum Beispiel, wenn Sie Gesprächsbeiträge von Robert mit **[Rob]** bzw. Beiträge von Barbara mit **[Barbara]** markiert haben, können Sie in einer Konkordanz dieses Gesprächs zwischen Barbara und Robert sehr schnell herausfinden, wer von den beiden welche Wörter verwendet hat.

Sie können auch ihren Text mit **<Einleitung>**, **<Hauptteil>** und **<Schlussteil>** markieren. "Nächstes" Tag wird jede Zeile folgendermaßen anzeigen:

```
1 ... Die Damen geben sich und ihren ... <Einleitung>
2 ... Wein auch geben. Ein tiefer ... <Hauptteil>
3 ... froh, der geben mag. Ein Tag ... <Hauptteil>
4 ... mir alles geben; Sobald du dir ... <Schlussteil>
```

Um den Text so zu markieren, erstellen Sie eine [Tagdatei](#) mit den Markierungen (Tags) für die Abschnitte und kennzeichnen sie als Abschnitte, wie in diesen Beispielen:

```
<ABSTRACT> /description "Abschnitt"
</VORWORT>
<INTRODUCTION> /description "Abschnitt"
</EINLEITUNG>
<SECTION 1> /description "Abschnitt"
</SECTION 1>
```

Sie können die Liste nach den nächsten Tags [sortieren](#). In dem Screenshot wurde eine Konkordanz des Worts **such** anhand des BNC World text berechnet. In manchen Fällen wurde **such** mit dem Tag <w PRP> markiert (**such as**), in anderen mit dem Tag <w DTO>. Die Tagsspalte zeigt das nächste Tag an, und die gesamte Liste wurde nach dieser Spalte sortiert.

Concord									
File Edit View Compute Settings Help									
N	Concordance					et	Tag	ord #	# s.
18	ATO>a <w NN1>scheme <w PRP>such as <w ATO>a <w NN1>mural						<w PRP>	1,993	01 20
19	<w NNO>works <w PRP>such as <w NN2>altarpieces <w						<w PRP>	2,310	12 5
20	not <w VVI>date, <w PRP>such as <w NN1>art <w						<w PRP>	4,403	00 16
21	<w NN1>programme, <w PRP>such as <w ATO>the <w						<w PRP>	1,417	07 5
22	<w NN1>title, <w PRP>such as <w NN1>futurism. <s						<w PRP>	1,958	30 55
23	<w NN2>techniques <w PRP>such as <w AJO>infra-red <w						<w PRP>	6,063	04 33
24	TOO>to <w VVI>make <w DTO>such <w NN2>identifications,						<w DT>	2,792	08 13
25	<w CJC>or <w DTO>such <w NN2>questions <w						<w DT>	3,133	20 16
26	<w CJC>and <w DTO>such <w NN2>treatises <w						<w DT>	3,170	21 35
27	VMO>may <w VVI>take <w DTO>such <w ATO>a <w						<w DT>	3,570	36 6
28	<w NN1>creation of <w DTO>such <w NN2>theories <w						<w DT>	3,722	41 25
29	<w PRP>by <w DTO>such <w ATO>an <w						<w DT>	4,857	95 17
30	<w PRP>in <w DTO>such <w NN2>papers <w CJS>as						<w DT>	4,951	99 5
31	<w NN1>criticism of <w DTO>such <w NN1>art <w PRP>within						<w DT>	7,534	01 5
32	DTO>This <w VVD>took <w DTO>such <w NN2>forms <w CJS>as						<w DT>	7,724	08 3

Wenn in Ihrer Konkordanz keine Tags angezeigt werden, haben die [ignorierten Tags](#) dasselbe Format. Zum Beispiel, wenn unter Tags ignorieren <*> festgelegt ist, werden alle Tags wie <Titel>, <Zitat>, etc. bei der Konkordanzsuche ignoriert, es sei denn, Sie legen diese Tags in einer [Tagdatei](#) fest. In diesem Fall müssen Sie eine Tagdatei angeben und die Konkordanz neu berechnen.

Sie können die Tags auch farbig markieren oder ausblenden -- das Wort in Tags bleibt dabei farbig markiert. Hier sehen Sie eine Konkordanz von **this** basierend auf BNC World Edition text mit farbig markierten Tags:

N	Concordance									
3	<wNN1>availability. <wDTO>This <wVBZ>is <wAVD>so <wPNP>we									
4	<wNN1>spread <wPRF>of <wDTO>this <wAJO>terrible <wNN1>disease									
5	<wNN1>need <wPRP>for <wDTO>this <wNN1>service. <wNN1>Home									
6	<wNN1>opening <wPRF>of <wDTO>this <wNN1>office <wPRP>in									
7	<wCJS>even though <wDTO>this <wVBZ>is <wAVD>increasingly									
8	<wPRP>on 21 <wNPO>June <wDTO>this <wNN1>year. <wAJO>Official									
9	<wTOO>To <wVDI>do <wDTO>this, <wPNP>we <wVMO>would									
10	<wPRP>in <wDTO>this <wNN1>way, <wPRP>with <wATO>									
11	<wVVB-NN1>need <wDTO>this <wVVI>support <wTOO>to									
12	<wNPO>November <wPRF>of <wDTO>this <wNN1>year. <wATO>The									
13	<wVVI>play <wPRP-AVP>in <wDTO>this <wNN1>process. &equo; <wPNP>He									
14	3&percent; <wPRP>by 1989. <wDTO>This <wNN1>change <wVHZ>has									
15	<wATO>the <wNN1>neck. <wDTO>This <wVMO>can <wVVI>make									
16	<wNN1>creation <wPRF>of <wDTO>this <wAJO>now <wNN1>next <wATO>									

In dieser Ansicht werden dieselben Daten angezeigt, mit der Option *Ansicht | Tags ausblenden* aktiviert:



Die Tags selbst sind nicht mehr sichtbar, und nur sechs verschiedene Tags wurden farbig markiert.

Siehe auch: [Tag-Übersicht](#), [Tag-Handling](#), [Eine Tagdatei erstellen](#), [Texte mit Tags](#), [Tagtypen](#), [Tags anzeigen](#), [Tags als Selektoren verwenden](#)

6.22 Muster

Wenn Sie ein Kollokationsfenster geöffnet haben, sehen Sie die Registerkarte "Muster". Dort sehen Sie die Kollokatoren (Wörter neben dem Suchwort), die innerhalb jeder Spalte nach Frequenz geordnet sind. So steht am Anfang jeder Spalte das Wort, das am häufigsten in dieser Position gefunden wurde. Das zweite Wort ist gleichzeitig auch das Zweithäufigste.

Die Mindestfrequenz, ab der ein Wort angezeigt wird, ist die [Mindestfrequenz für Kollokatoren](#).

Wozu dient das?

Dadurch stehen die Wörter, die sich am häufigsten in der Nähe des Suchworts befinden, am Anfang der Liste. Wie bei Kollokationen können Sie dadurch die lexikalischen Muster in der Konkordanz besser erkennen.

6.23 neu sortieren

Wie mache ich das?

Neu sortieren können Sie entweder durch Anklicken der obersten Reihe einer Liste, durch Drücken von F6 bzw. Strg+F6 oder mit der Menüoption "Neu sortieren".

Beim Erstellen einer Konkordanz erscheinen die Einträge sortiert nach den Textdateien, aus denen sie stammen: Dateisortierung. Sobald die Konkordanz fertig gestellt ist, wird sie entsprechend den aktuellen [Standardeinstellungen](#) neu sortiert. Wenn Sie nicht neu sortieren möchten, legen Sie die Standardeinstellungen auf file,file fest.

Wozu dient das?

Neu sortieren hilft dabei, charakteristische lexikalische Muster zu erkennen. Es ist manchmal schwierig, allgemeine Tendenzen in Konkordanzzeilen zu erkennen, speziell wenn die Konkordanz sehr lang ist. Indem Sie die Zeilen sortieren, können Sie multiple Suchwörter trennen und den direkten Kontext links und rechts untersuchen. Wenn bei vielen Einträgen direkt vor dem Suchwort "in dem" oder "in einem" oder "in meinem" steht, kann man Klarheit schaffen, indem man die Konkordanz nach dem zweiten Wort links des Suchworts sortiert. Sie können sortieren nach den Wörtern links oder rechts des Suchworts (L1 [=1 Wort links des Suchworts], L2, L3, L4, L5, R1 [=1 Wort rechts des Suchworts], R2, R3, R4, R5), dem Suchwort selbst, dem Kontextwort (sofern Sie eines angegeben haben), dem [nächsten Tag](#), dem Abstand zum nächsten Tag, der [Kategorie "Klasse"](#), die Sie selbst festlegen können, oder der ursprünglichen Dateisortierung (Datei).

Primärsortierung

Die Liste kann nach drei Kriterien gleichzeitig sortiert werden. Eine Primärsortierung nach L1 (1. Wort links des Suchworts) sortiert die Einträge alphabetisch nach dem Wort direkt links des Suchworts. Mit einer Sekundärsortierung (... dann nach ...) nach R2 passiert folgendes: nur die Einträge, bei denen die Wörter auf L1 (direkt links des Suchworts) exakt übereinstimmen, werden zusätzlich alphabetisch nach R2 (2. Wort rechts des Suchworts) geordnet. Für sehr lange Konkordanzen kann eine dritte Sortierung (...zuletzt nach) hilfreich sein: falls auch nach der Sekundärsortierung Einträge auf der Position R2 exakt übereinstimmen, möchten Sie diese vielleicht zusätzlich sortieren.

In den meisten Fällen sind Sekundärsortierungen nicht notwendig, und werden ignoriert, wenn die Primär- und Sekundärsortierung gleich sind (z. B. L1 und L1) oder wenn das Kontrollkästchen "aktiviert" NICHT aktiviert ist.

Sortieren nach Klasse ([benutzerdefinierte Kategorien](#))

Sie können auch nach "Klasse" sortieren, wenn Sie beschlossen haben, die Konkordanzzeilen nach Ihrem eigenen System zu klassifizieren, und zwar mit Buchstaben von **A** bis **Z** oder **a** bis **z**. Durch die Sortierung werden die klassifizierten Zeilen an den Anfang gerückt, geordnet nach Kategorien. Nicht klassifizierte Zeilen werden an das Ende der Konkordanz verschoben und hellgrau markiert. Siehe [Nächstes Tag](#) für nähere Informationen bezüglich der Sortierung nach Tags.

Die Farbe des Suchworts verändert sich entsprechend dem verwendeten Sortiersystem.

Alle

Standardmäßig werden alle Zeilen sortiert; Sie können aber auch z. B. 5-49 eingeben, um die Sortierung auf die Zeilen 5 bis 49 zu beschränken.

Aufsteigend

Wenn dieses Kontrollkästchen aktiviert ist, wird von **A** bis **Z** sortiert, ansonsten von **Z** bis **A**.

Siehe auch: [WordList-Sortierung](#), [KeyWords-Sortierung](#), [Sprachen auswählen](#)

6.24 neu sortieren: Kollokatoren

The frequency-ordered collocation display can be re-sorted to reveal the frequencies sorted by their *total* frequencies overall (the default), by the left or right frequency total, or by any individual frequency position, from 25 words to the left of the search word to 25 words to the right. Just press the header of a column to sort it. Press again to toggle the sort between ascending to descending.

The point of it...

is to find patterns of collocation, so as to more fully understand the company your search-word keeps.

The choices depend on the [collocation horizons](#).

See also: [Collocation](#), [Collocation Display](#)

6.25 neu sortieren: Verteilungsplot

This automatically re-sorts the dispersion plot, rotating through these options:

alphabetically (by file-name)

in *frequency* order (in terms of hits per 1,000 words of running text)

by first occurrence in the source text(s): *text order*

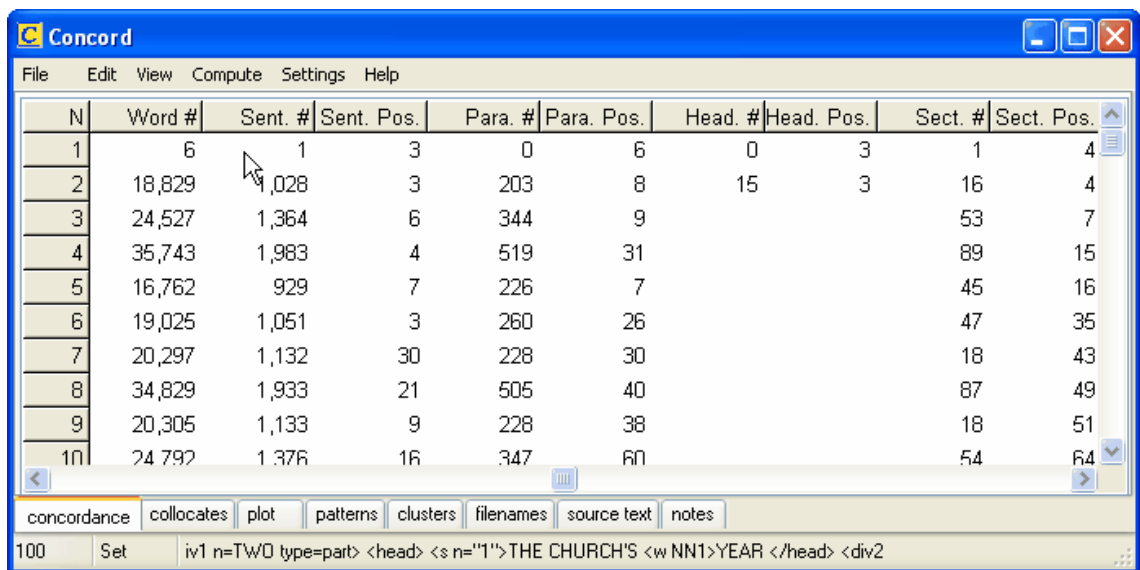
by *range*: the gap between first and last occurrence in the source text.

see also: [Dispersion Plot](#)

6.26 Textsegmente in Concord

Eine Konkordanzzeile enthält Informationen über das Textsegment, in dem sie gefunden wurde.

In dem Screenshot unten wurde eine Konkordanz für das Wort **year** erstellt; die Liste wurde nach Heading Position (Position in Überschrift) sortiert -- in den ersten zwei Zeilen wurde **year** an dritter Stelle in einer Überschrift gefunden. Der Vorteil dieser Funktion ist die Möglichkeit, Suchwörter zu identifizieren, die am Anfang eines Satzes, eines Abschnitts, einer Überschrift oder eines Absatzes vorkommen.



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

concordance collocates plot patterns clusters filenames source text notes

100 Set iv1 n=TW0 type=part <head> <s n='1'>THE CHURCH'S <w NN1>YEAR </head> <div2

Siehe auch: [Anfang und Ende von Textsegmenten](#).

6.27 Suchwort-Syntax

Standardmäßig werden von Concord ganze Wörter gesucht und Groß- und Kleinschreibung nicht beachtet.

Beispiele

Suchwort	findet
book	<i>Book oder book oder BoOk</i>
book	<i>book, books, booking, booked</i>
*book	<i>textbook (aber nicht textbooks)</i>
bo* in	<i>book in, books in, booking in (aber nicht 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 (aber nicht BOOK oder Book)</i>
book/paperback	<i>book oder paperback</i>

Symbol	bedeutet	Beispiele
*	Wortende ignorieren, ein ganzes Wort ignorieren	tele* *ness *happi*
?	ein beliebiges Zeichen (auch Interpunktion) kann an dieser Stelle stehen	book * hotel Engl??? ?50.00

#	eine beliebige Ziffer, 0 bis 9	\$### £##.00 Fr^nc^
^	ein beliebiger Buchstabe des Alphabets kann an dieser Stelle stehen	
==	Groß-/Kleinschreibung wird beachtet	==French== ==Fr*==
: \	Datei für viele Suchwörter verwenden (siehe dateibasierte Suchwörter)	c:\text\frd.txt
/	trennt alternative Suchwörter. Sie können Alternativen festlegen innerhalb eines 80-Zeichen-Gesamtlimits	may/can/will
<>	Anfang und Ende von Tags	<w NN1>

Wenn Sie *, ? , == , #, ^ , : \ , > , < oder / als Zeichen in Ihrem Suchwort verwenden möchten, setzen Sie es in Anführungszeichen. Beispiele:

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

Vergessen Sie nicht, dass Fragezeichen in Interrogativsätzen am Ende des letzten Worts stehen, Sie benötigen also *"?"

Tags

Wenn Ihr Text Tags enthält, können Sie Tags in Ihrem Suchwort festlegen. Beispiele:

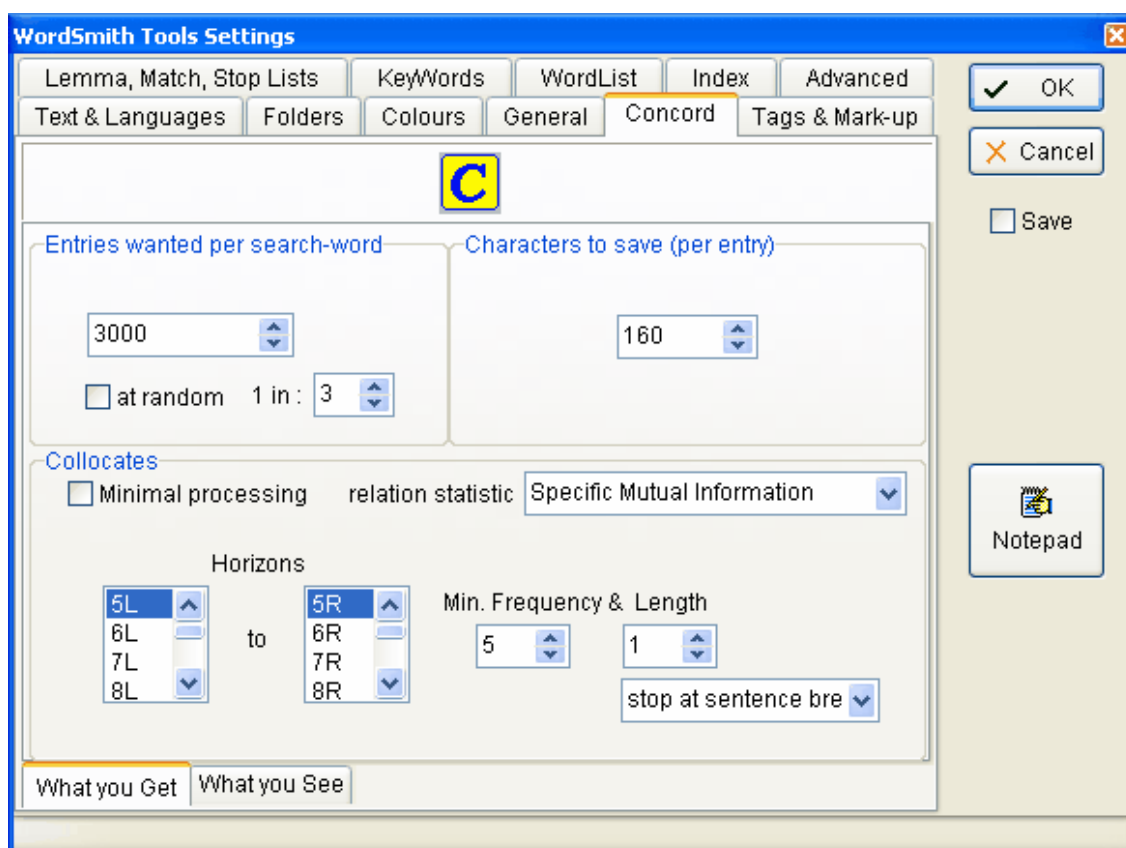
Symbol	bedeutet	Beispiele
<w NN1>*	Gattungsname im Singular (BNC)	<i>book, chair, elephant</i>
<w NN?>*	Gattungsname im Singular oder Plural	<i>book, chairs</i>
<w NN1>t*	jedes Substantiv im Singular, das mit T oder t beginnt	<i>table, teacher</i>
<w NN1>* <w NN1>*	zwei Gattungsnamen im Singular hintereinander	<i>campaign manager</i>

Siehe auch: [Tag-Konkordanz erstellen](#), [Kontextwort](#)

6.28 WordSmith Controller: Concord: Einstellungen

These are found in the main [Controller](#) under *Adjust Settings | Concord*.

This is because some of the choices -- e.g. [collocation horizons](#) -- may affect other Tools.



WHAT YOU GET and WHAT YOU SEE

There are 2 tabs for settings affecting *What you get* in the concordance and *What you see* in the display. There is a screenshot below showing the options under *What you see*.

WHAT YOU GET

Entries Wanted

The maximum is more than 2 billion lines. This feature is useful if you're doing a number of searches and want, say, 100 examples of each. The 100 entries will be the first 100 found in the texts you have selected. If you search for more than 1 search-word (eg. **book/paperback**), you will get 100 of **book** and 100 of **paperback**.

"*at random*" is a feature which allows you to randomise the search. Here **Concord** goes through the text files and gets the 100 entries by giving each hit a random one-in-three chance of being selected. To get 100 entries **Concord** will have found around 250-350 hits. You can set the randomiser anywhere from 1 in 2 to 1 in 1,000.

Characters to save

Here is where you set how many characters in a concordance line will be saved as text (as opposed to sending them to the [clipboard](#)). The default is 80 (minimum 20 and maximum 8,000). The reason for this is that you will probably want a fixed number of characters so that when using a non proportional font, such as Courier or Lucinda Console, the search-words line up nicely.

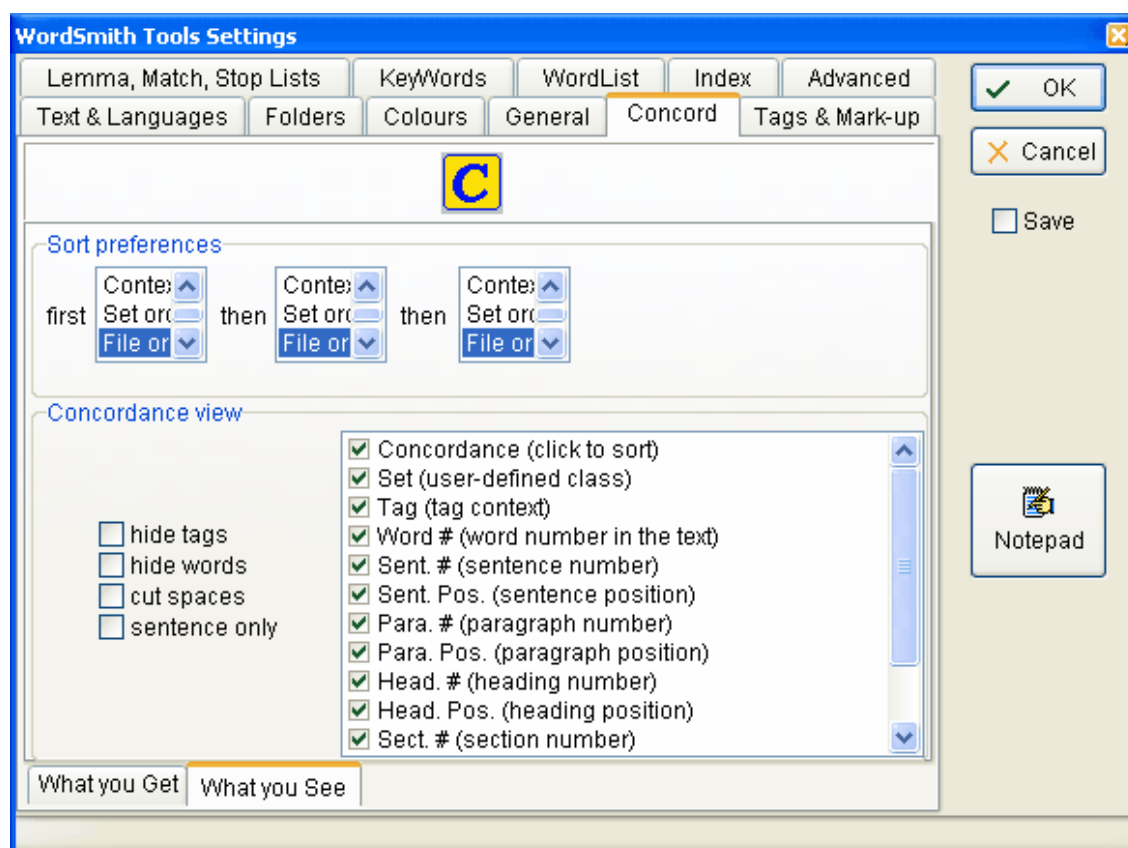
Collocates

By default, **Concord** will compute collocates as well as the concordance, but you can set it not to if you like (*Minimal processing*). For further details, see [Collocate Horizons](#) or [Collocation](#)

Collocates relation statistic

Choose between Specific Mutual Information, MI3, Z Score, Log Likelihood. See [Mutual Information Display](#) for examples of how these can differ.

WHAT YOU SEE



Sort preferences

By default, **Concord** will sort a new concordance in original file order, but you can set this to different values if you like. For further details, see [Sorting a Concordance](#).

Concordance view

You can choose different ways of seeing the data, and a whole set of choices as to what columns you want to display for each new concordance. You can re-instate any later if you wish by changing the [Layout](#).

hide search-word = blank it out eg. to make a [guess-the-word exercise](#)

hide undefined tags = hide those not defined in your [tag file](#)

hide tag file tags = hide all tags including undefined ones

hide words = show only the tags

cut spaces = remove any double spaces

sentence only = show the context only up to its left and right [sentence boundaries](#)

raw numbers = show the raw data instead of percentages e.g. for sentence position

See also: Concord Help Contents, [Collocation Settings](#).

WordSmith Tools

Key Words

Section

VIII

7 KeyWords

7.1 Zweck



Dieses Programm dient dazu, die Schlüsselwörter eines Texts oder mehrerer Texte zu identifizieren. Schlüsselwörter sind solche Wörter, die im Vergleich mit einem Normwert ungewöhnlich häufig auftreten. Klicken Sie hier, um ein [Beispiel](#) zu sehen.

Wozu dient das?

Schlüsselwörter geben aufschlussreiche Informationen über einen Text oder eine Textart. Mögliche Anwendungsbereiche sind: Sprachunterricht, forensische Linguistik, Stilistik, Inhaltsanalysen, Abruf von Texten.

Dieses Programm vergleicht zwei bereits vorhandene Wortlisten, die mit dem Programm WordList erstellt wurden. Eine dieser Listen ist größer als die andere und dient somit als Referenzdatei. Die andere Wortliste ist kürzer und basiert auf dem Text, den Sie untersuchen wollen.

Ziel ist es herauszufinden, welche Wörter typisch für den Text sind, für den Sie sich interessieren (der kürzere der beiden ausgewählten Texte). Der längere Text dient als Referenzkorpus und liefert die Hintergrundinformationen für den Vergleich.

Schlüsselwörter und deren [Links](#) können in einem [Plot](#) grafisch dargestellt sowie entsprechend ihrer [assoziierten Wörter](#) gruppiert werden. Außerdem können Sie aus Schlüsselwörtern eine [Datenbank](#) erstellen.

7.2 Analyse von zwei Wortlisten

The usual kind of **KeyWords** analysis. It compares the one text file you're chiefly interested in, with a reference corpus based on a lot of text.

Choose Word Lists

In the dialogue box you will choose 2 files. The text file in the box above and the reference corpus file in the box below.

See also [How Key Words are Calculated](#), [KeyWords Settings](#)

7.3 Definition assoziierte Wörter

An "associate" of key-word X is another key-word (Y) which co-occurs with X in a number of texts. It may or may not co-occur in proximity to key-word X. (A *collocate* would have to occur within a given distance of it, whereas an associate is "associated" by being key in the same text.) For example, in a key-word database of *Guardian* newspaper text, *wine* was found to be a key word in 25 out of 299 stories from the Saturday "tabloid" page, thus a [key key word](#) in this section. The top associates of *wine* were: *wines, Tim, Atkin, dry, le, bottle, de, fruit, region, chardonnay, red, producers, beaujolais*.

It is strikingly close to the early notion of "collocate".

Association operates in various ways. It can be strong or weak, and it can be one-way or two-way. For example, the association between *to* and *fro* is one-way (*to* is nearly always found near *fro* but it is rare to find *fro* near *to*).

See also: [Definition of Key Word](#), [Associates](#), [Definitions](#), [Mutual Information](#)

7.4 assoziierte Wörter

"Associates" is the name given to key-words associated with a [key key-word](#).

The point of it...

The idea is to identify words which are commonly associated with a key key-word, because they are key words in the same texts as the key key-word is. An example will help.

Suppose the word *wine* is a key key-word in a set of texts, such as the weekend sections of newspaper articles. Some of these articles discuss different wines and their flavours, others concern cooking and refer to using wine in stews or sauces, others discuss the prices of wine in a context of agriculture and diseases affecting vineyards. In this case, the associates of *wine* would be items like *Chardonnay*, *Chile*, *sauce*, *fruit*, *infected*, *soil*, etc.

The listing shows associates in order of frequency. A menu option allows you to re-sort them.

You can set a minimum number of files for the association procedure, using *Settings | Database*.

See also: [definition of associate](#).

7.5 Dateien auswählen

Wortliste des aktuellen Texts

Im oberen Feld wählen Sie eine Wortlistendatei aus.

Um mehr als eine Wortliste auswählen zu können, halten Sie während des Anklickens die Taste Strg (für nicht nebeneinander liegende Listen) bzw. die Umschalttaste (für einen ganzen Bereich) gedrückt.

In diesem Feld wird bestimmt, in welchen Wortlisten nach Schlüsselwörtern gesucht wird.

Wortliste des Referenzkorpus

In dem unteren Feld wählen Sie die [Referenzkorpus](#)liste aus. (In den Einstellungen im Controller kann dieser Referenzkorpus permanent festgelegt werden.)

Keine Wortlisten werden angezeigt

Wenn Sie keine Wortlisten in der Anzeige sehen, wechseln Sie entweder das Verzeichnis bis Wortlisten angezeigt werden, oder Sie gehen zurück zu WordList und erstellen mindestens zwei Wortlisten: Für diesen Vorgang sind mindestens zwei Wortlisten erforderlich, um diese miteinander vergleichen zu können.

7.6 assoziierte Gruppen

"Clumps" is the name given to groups of key-words [associated](#) with a [key key-word](#).

The point of it (1)...

The idea here is to refine associates by grouping together words which are found as key in the same sub-sets of text files. The example used to explain associates will help.

Suppose the word *wine* is a key key-word in a set of texts, such as the weekend sections of newspaper articles. Some of these articles discuss different wines and their flavours, others concern cooking and refer to using wine in stews or sauces, others discuss the prices of wine in a context of agriculture and diseases affecting vineyards. In this case, the associates of *wine* would be items like *Chardonnay*, *Chile*, *sauce*, *fruit*, *infected*, *soil*, etc. The associates procedure shows all such items unsorted.

The clumping procedure, on the other hand, attempts to sort them out according to these different uses. The reasoning is that the key words of each text file give a condensed picture of its "aboutness", and that "aboutnesses" of different texts can be grouped by matching the key word lists. Thus sets of key words can be clumped together according to the degree of overlap in the key word lexis of each text file.

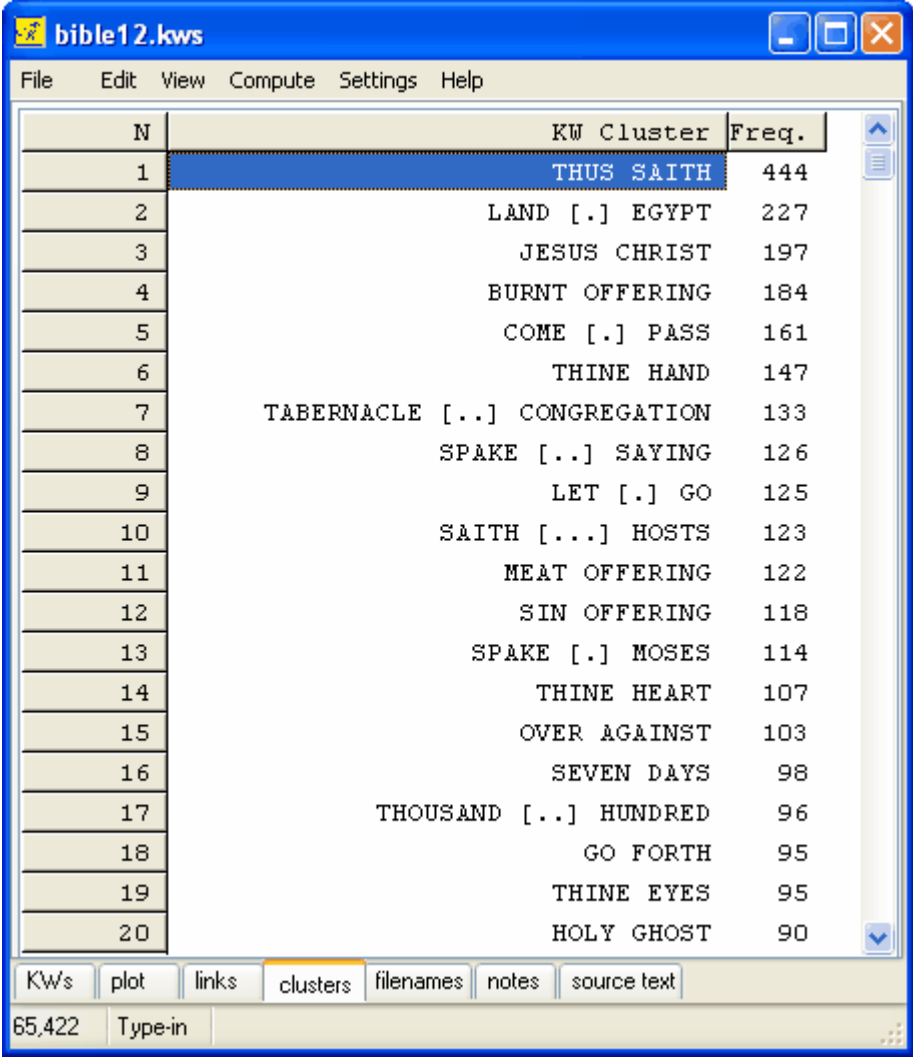
Two stages

The **initial clumping process does no grouping**: you will simply see each set of key-words for each text file separately. To [group clumps](#), you may simply join those you think belong together (by dragging), or regroup with help by pressing . The listing shows clumps sorted in alphabetical order. You can re-sort by frequency (the number of times each key word in the clump appeared in all the files which comprise the clump). See also: [definition of associate](#), [regrouping clumps](#)

7.7 KeyWords: Cluster

Ein KeyWords-Cluster, wie auch ein WordList-Cluster, repräsentiert zwei oder mehr Wörter, die in den Texten wiederholt nahe beieinander stehen. Ein KeyWords-Cluster besteht nur aus Schlüsselwörtern.

Ein Screenshot soll dies verdeutlichen:



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

Diese Cluster wurden anhand der Bibel als Quelltext berechnet. Basierend auf dem Vergleich mit einem Referenzkorpus ist jedes dieser Wörter ein Schlüsselwort. Die Cluster sind Fälle, in denen diese Schlüsselwörter zusammen innerhalb des aktuellen Kollokationsbereichs

auftauchen. Die [...] Klammern repräsentieren Fälle, in denen die Schlüsselwörter nicht gemeinsam auftreten, z. B. bei **come** [...] **pass** steht der Punkt für das Wort **to**, und die häufige Phrase ist **come to pass**.

Siehe auch: [Plot-Berechnung](#).

7.8 Konkordanz

Wenn Sie bereits eine Schlüsselwortliste geöffnet haben, können Sie  anklicken, um eine Konkordanz der aktuell ausgewählten Wörter zu erstellen. Die Konkordanz wird basierend auf der ursprünglichen Textdatei erstellt, aus der Sie die Schlüsselwortliste generiert haben.

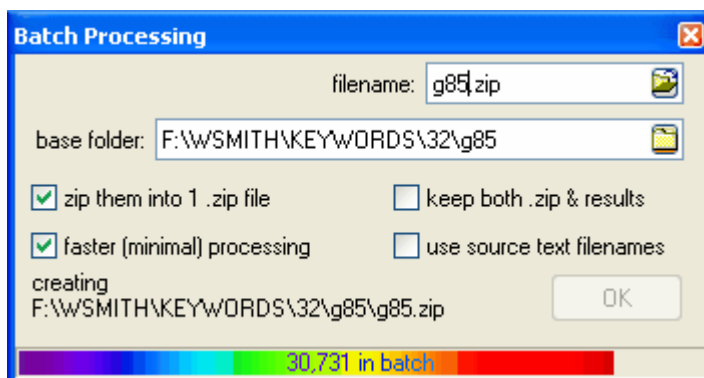
Wozu dient das?

Damit werden dieselben Schlüsselwörter in ihren ursprünglichen Kontexten angezeigt.

7.9 Erstellung einer Datenbank

Um eine Schlüsselwortdatenbank zu erstellen, benötigen Sie eine Gruppe von Schlüsselwortlisten. Für eine Datenbank mit einer großen Datenmenge ist es empfehlenswert, folgendermaßen vorzugehen:

1. Erstellen Sie ein [Batch](#) aus Wortlisten.
2. Verwenden Sie diese Wortlistengruppe, um daraus ein [Batch](#) von Schlüsselwortlisten zu erstellen. Aktivieren Sie "Schnelle Minimalverarbeitung" wie in diesem Screenshot, um auf die Berechnung von Plots, etc. zu verzichten und so den Vorgang zu beschleunigen.



3. Wählen Sie nun in **KeyWords Neu | Datenbank**.

So können Sie alle Schlüsselwortlisten gleichzeitig auswählen.

Beachten Sie bitte, dass in eine Datenbank nur [positive](#) Schlüsselwörter aufgenommen werden. In den [Controller-KeyWords-Einstellungen](#) können Sie bestimmen, wie viele Schlüsselwörter pro Text in die Datenbank aufgenommen werden.

Geben Sie dann an, in welchem Verzeichnis Sie die erstellte Datenbank speichern möchten.

Nachdem Sie die Datenbank erstellt haben, können Sie die Datenbank *öffnen* und sich [gewichtete Schlüsselwörter](#) und [assoziierte Wörter](#) ansehen.

7.10 Beispiel Schlüsselwörter

Sie besitzen eine Sammlung von verschiedenen Zeitungsartikeln. Basierend auf diesen Artikeln erstellen Sie eine Wortliste und stellen fest, das häufigste deutsche Wort ist *der*. Unter den eher seltenen Wörtern in dieser Liste finden sich Beispiele wie *hüpfen*, *Modem*, *ehrgeizig*, etc.

Dann untersuchen Sie einen Artikel mit 1.000 Wörtern aus der Sammlung und erstellen eine Wortliste. Das häufigste Wort ist wiederum *der*. Es sind keine großen Unterschiede festzustellen.

Danach analysieren Sie mit **KeyWords** diese beiden Wortlisten. **KeyWords** analysiert die Texte und die häufigsten "Schlüssel"wörter sind: *Polizei*, *Häftling*, *Zusammenbruch*, *Rat*, *angeklagt*, *Müller*, *Gemeinschaft*.

Diese "Schlüssel"wörter sind nicht die häufigsten Wörter (wie etwa *der*), sondern die außergewöhnlich häufigen Wörter in einem Artikel mit 1.000 Wörtern. Schlüsselwörter vermitteln im Normalfall einen ziemlich guten Eindruck über das Thema des Texts.

7.11 Definition eines gewichteten Schlüsselworts

Ein "gewichtetes Schlüsselwort" ist ein Wort, das in mehreren zusammenhängenden Texten als "Schlüsselwort" auftritt. Die Wichtigkeit eines Schlüsselworts steigt mit der Anzahl der Texte, in dem es als Schlüsselwort erscheint. Dies wird entscheidend durch die Homogenität der Themen innerhalb des zu untersuchenden Korpus beeinflusst. In einem Korpus, der aus Stadtnachrichten besteht, werden gewichtete Schlüsselwörter wie *Bank*, *Gewinn*, *Unternehmen* vorkommen, nicht jedoch *Computer*, das möglicherweise nur das Schlüsselwort in einigen wenigen Artikeln über IBM oder Microsoft Aktiengeschäfte auftritt.

Siehe auch: [Berechnung von Schlüsselwörtern](#), [Definition eines Schlüsselworts](#), [Erstellung einer Datenbank](#), [Definitionen](#)

7.12 Definition der Gewichtung

Der Begriff "Schlüsselwort" ist in der Linguistik nicht definiert, obwohl er sehr häufig verwendet wird. Dieses Programm identifiziert maschinell die Schlüsselwörter, indem es die Frequenzmuster vergleicht. (Ein Mensch hingegen wählt möglicherweise eine Phrase oder einen Oberbegriff als Schlüsselwort.)

Ein Wort ist dann ein "Schlüsselwort", wenn:

- a) es in einem Text mindestens so häufig auftritt, wie vom Benutzer in der Mindestfrequenz festgelegt wurde.
- b) seine Frequenz in einem Text verglichen mit der Frequenz in einem Referenzkorpus entsprechend der statistischen Wahrscheinlichkeit, die mittels [entsprechender Verfahren](#) errechnet wird, gleich hoch oder niedriger ist als der vom Benutzer festgelegte [P-Wert](#).

Positive und negative Gewichtung

Ein Wort mit einer *positiven* Gewichtung tritt *häufiger* auf, als dieses Wort im Vergleich mit dem Referenzkorpus zufällig auftreten würde.

Ein Wort mit *negativen* Gewichtung tritt *weniger häufig* auf, als dieses Wort im Vergleich mit dem Referenzkorpus zufällig auftreten würde.

Typische Schlüsselwörter

KeyWords berechnet normalerweise drei Arten von "Schlüsselwörtern".

1. Eigennamen. Eigennamen sind in Texten oft Schlüsselwörter, obwohl etwa in einem Text über Pferderennen die Namen der Pferde zwar gewichtet, jedoch eher nebensächlich sind. Das

kann durch eine höhere Mindestfrequenz vermieden werden.

2. Wörter, die von Menschen als Schlüsselwörter identifiziert werden. Dieses Programm eignet sich sehr gut für diese Schlüsselwörter, die einen guten Einblick in das Thema eines Texts vermitteln. (Das Programm kann jedoch Synonyme nicht gruppieren, und Wörter, die in einem Text nur einmal vorkommen, sind für das menschliche Verständnis manchmal Schlüsselwörter. Außerdem wird **KeyWords** Schlüsselphrasen nicht erkennen, es sei denn Sie vergleichen Wortlisten, die auf [Wortgruppen](#) basieren.)

3. Häufig auftretende Wörter wie "weil", "wird" oder "schon". Solche Wörter werden vom Leser in der Regel nicht als Schlüsselwörter identifiziert. Sie geben eher Aufschluss über den Stil als über das Thema eines Texts. Wenn KeyWords solche Wörter identifiziert, sollten Sie den Text eventuell mit Concord (auf  klicken) auf die ungewöhnliche Frequenz dieser Wörter untersuchen.

Siehe auch: [Berechnung von Schlüsselwörtern](#), [Definition eines gewichteten Schlüsselworts](#), [Definitionen](#), [KeyWords-Einstellungen](#)

7.13 KeyWords-Datenbank

(default file extension .KDB)

The point of it...

The point of this database is that it will allow you to see the "key-key-words" in your set of files. That is, the key-words which are most frequent over a number of files.

For example, if you have 500 business reports, each one will have its own key words. These will probably be of two main kinds. There will be key-words which are key in one text but are not generally key (names of the firms and words relating to what they individually produce); and other, more general words (like consultant, profit, employee) which are typical of business documentation generally.

By making up a database, you can sort these out. The ones at the top of the list, when you view them, will be those which are most typical of the genre. The list is ordered in terms of "key key-ness" but can be toggled into alphabetical order and back again.

You can set a minimum number of files that each word must have been found to be key in, using *Settings | KeyWords | Database*.

When viewing a database you will be able to investigate the [associates](#) of the key key-words. Under Statistics, you will also be able to see details of the key words files which comprise the database (file name and number of key words per file), together with overall statistics on the number of different types and the tokens (the total of all the key-words in the whole database including repeats).

See also : [Creating a database](#), [Definition of key key-word](#)

7.14 KeyWords: Tipps

1. Es ist nicht sinnvoll, einen Schlüsselwort-Plot basierend auf mehreren Textdateien zu erstellen. In dem Plot werden immer nur die Wörter aus der ersten Textdatei dargestellt. Wenn Sie den Plot eines bestimmten Worts oder einer bestimmten Phrase in verschiedenen Dateien erstellen, verwenden Sie dazu die Funktion [Concord-Verteilung](#).
2. Es gibt keine Garantie, dass die Schlüsselwörter Ihren Vorstellungen von Schlüsselwörtern entsprechen. Ein "wichtiges" Wort kommt möglicherweise nur einmal im

gesamten Text vor. Schlüsselwörter in WordSmith sind nur diejenigen Wörter, die verglichen mit dem Referenzkorpus besonders häufig oder selten vorkommen.

3. Versuchen Sie nicht, Äpfel mit Birnen zu vergleichen. Folgen Sie bei der Auswahl Ihres Referenzkorpus bestimmten Prinzipien. Diese Aufgabe können Ihnen weder der Computer noch WordSmith abnehmen.

7.15 KeyWords: Berechnung

Die "Schlüsselwörter" werden berechnet, indem die Frequenz jedes Worts in der Wortliste des betreffenden Texts mit der Frequenz desselben Worts in der Referenzwortliste verglichen wird. Es werden alle Wörter einbezogen, die in der kleineren Liste auftreten, es sei denn, sie befinden sich in einer [Stoppliste](#).

Wenn z. B. das Wort **der** zu 5% in kleinen Wortlisten und zu 6% in dem Referenzkorpus auftritt, wird es nicht zu den Schlüsselwörtern gezählt, obwohl es sehr wahrscheinlich das häufigste Wort ist. Wenn der Text von der Anatomie der Spinne handelt, so ist es sehr wohl möglich, dass die Namen der Forscher und Wörter wie **Spinne**, **Bein**, **acht**, usw. häufiger auftreten als im Referenzkorpus. (Es sei denn, im Referenzkorpus werden nur Spinnen behandelt!)

Um die "Gewichtung" eines Worts zu berechnen, muss das Programm folgende Parameter bestimmen:

- seine Frequenz in der kleinen Wortliste
- die gesamte Wortanzahl in der kleinen Wortliste
- seine Frequenz im Referenzkorpus
- die gesamte Wortanzahl im Referenzkorpus

und diese miteinander verknüpfen.

Statistische Verfahren:

- der klassische Chi-Quadrat-Test mit Yates-Korrektur für eine Vierfeldertafel
- [Ted Dunnings](#) Log-Likelihood-Test, der bessere Schätzwerte für die Gewichtung liefert, speziell wenn man lange Texte oder eine Sammlung von Texten der gleichen Textsorte mit dem Referenzkorpus vergleicht.

In diese Liste gelangen Wörter, wenn sie im Vergleich mit der großen Wortliste ungewöhnlich häufig (oder ungewöhnlich selten) auftreten.

Ungewöhnlich *seltene* Schlüsselwörter heißen "negative Schlüsselwörter", und erscheinen in einer anderen Farbe am Ende der Liste. Beachten Sie, dass negative Schlüsselwörter automatisch aus einer KeyWords-[Datenbank](#) und einem Plot entfernt werden.

7.16 KeyWords: Links

Wozu dient das?

Mit dieser Funktion finden Sie heraus, welche Schlüsselwörter am engsten mit einem vorgegebenen Schlüsselwort verknüpft sind.

Ein Plot zeigt an, an welcher Stelle in der Originaldatei das Schlüsselwort vorkommt. Er zeigt außerdem an, wie viele Links es zwischen den Schlüsselwörtern gibt.

Was sind Links?

Links sind "innerhalb des Kollokationsbereichs gleichzeitig auftretende Schlüsselwörter". Anhand ein Beispiels soll das verdeutlicht werden:

Angenommen die beiden Wörter *Elefant* und *Wasser* sind Schlüsselwörter eines Texts über Afrika. Wenn *Elefant* und *Wasser* nicht mehr als fünf Wörter voneinander entfernt vorkommen,

dann werden diese beiden Wörter als "Links" bezeichnet. Wie häufig diese beiden Wörter im Text verlinkt sind, wird im Links-Fenster angezeigt.

Anzeige

In diesem Links-Fenster wird die Anzahl der Links angezeigt, gefolgt von einer Spalte mit dem Titel "in" und einem Prozentsatz. Dieser Prozentsatz ergibt sich aus der Gesamtanzahl der Links dividiert durch die Gesamtanzahl der Frequenz des betreffenden Worts (die Zahl in der Spalte "in"). Wenn Sie also die Links von *Elefant* sehen möchten, und das Wort *Elefant* erscheint insgesamt zehn Mal in Ihrem Originaltext, und jedes Mal erscheint es in der Nähe des Worts *Wasser* (obwohl *Wasser* insgesamt 40 Mal erscheint), werden 100% angezeigt. Wenn Sie die Links von *Wasser* sehen möchten, werden neben *Elefant* 25% angezeigt.

Der Kollokations"[bereich](#)" kann in **Concord** festgelegt werden, und bis zu 25 Wörter links bzw. rechts umfassen. Die [Standardeinstellung](#) ist 5/5.

Doppelklicken Sie auf irgendein Wort in der [Plotliste](#) um ein Fenster aufzurufen (bis zu maximal 20 Fenster), in dem die verknüpften Schlüsselwörter angezeigt werden.

Siehe auch: [Plot-Berechnung](#), [KeyWords: Cluster](#)

7.17 Wortliste aus KeyWords-Daten erstellen

Wenn Sie bereits eine Schlüsselwortliste geöffnet haben, können Sie auf  klicken, um Ihre Daten als Wortliste zu speichern (zum späteren Vergleich, etc. mit Hilfe der **WordList**-Funktionen).

7.18 P-Wert

Der **P**-Wert wird bei Standard-Chi-Quadrat-Tests und anderen statistischen Tests verwendet. Dieser Wert reicht von 0 bis 1. Ein Wert von 0,01 bedeutet, dass mit einer Wahrscheinlichkeit von 1% die berechnete Beziehung nicht vorliegt, bei einem Wert von 0,05 liegt die Gefahr eines Irrtums bei 5%. In den Sozialwissenschaften ist ein 5%-Risiko normalerweise akzeptabel. Im Fall einer Schlüsselwortanalyse, wo die Möglichkeit eines Irrtums weniger wichtig ist als die Selektivität, ist es vielleicht oft sinnvoll, einen vergleichsweise niedrigen **P**-Grenzwert festzulegen, wie etwa 0,000001 (eins in einer Million), so weniger Schlüsselwörter zu erhalten. Oder Sie legen im [Controller](#) eine niedrige Anzahl für "max. Ergebnisse" fest, unter *Einstellungen anpassen* | *KeyWords*.

Wird das [Chi-Quadrat-Verfahren](#) angewandt, dann wird der berechnete P-Wert nur angezeigt, wenn alle entsprechenden statistischen Anforderungen zutreffen (alle erwarteten Werte ≥ 5).

Siehe auch: [Definitionen](#)

7.19 Plot-Berechnung

Wozu dient das?

Das Plot dient zur Darstellung der Verteilung von Schlüsselwörtern in einem Text. Befinden sich die Schlüsselwörter größtenteils in der Mitte oder am Anfang des Texts?

Wie mache ich das?

Mit dieser Funktion werden alle Beziehungen zwischen den Schlüsselwörtern, die bis zu diesem Zeitpunkt erkannt wurden, berechnet. Nicht einbezogen werden Schlüsselwörter, die gelöscht oder [entfernt](#) wurden.

1. Basierend auf dem Text wird eine Konkordanz erstellt, bei der nach jedem Schlüsselwort gesucht wird.

2. Danach wird berechnet, welches der anderen Schlüsselwörter sich innerhalb des Kollokationsbereichs, der unter *Einstellungen* festgelegt wurden, befindet. Berücksichtigt wird dabei die breitere der beiden Grenzen.
3. Dann werden alle Wörter und ihre Positionen in der Originaldatei grafisch dargestellt. (Anhand eines "Lineals" wird dargestellt, wie viele Wörter sich in jedem Teil der Datei befinden.)
4. Es wird berechnet, wie viele andere Schlüsselwörter innerhalb des aktuellen Kollokationsbereichs auftreten.
5. es wird ein [Plot-Verteilungswert](#) berechnet.

Anmerkung: Für diesen Vorgang muss KeyWords auf die [Quelltexte](#), auf denen Ihre originale Wortliste basiert, zugreifen.

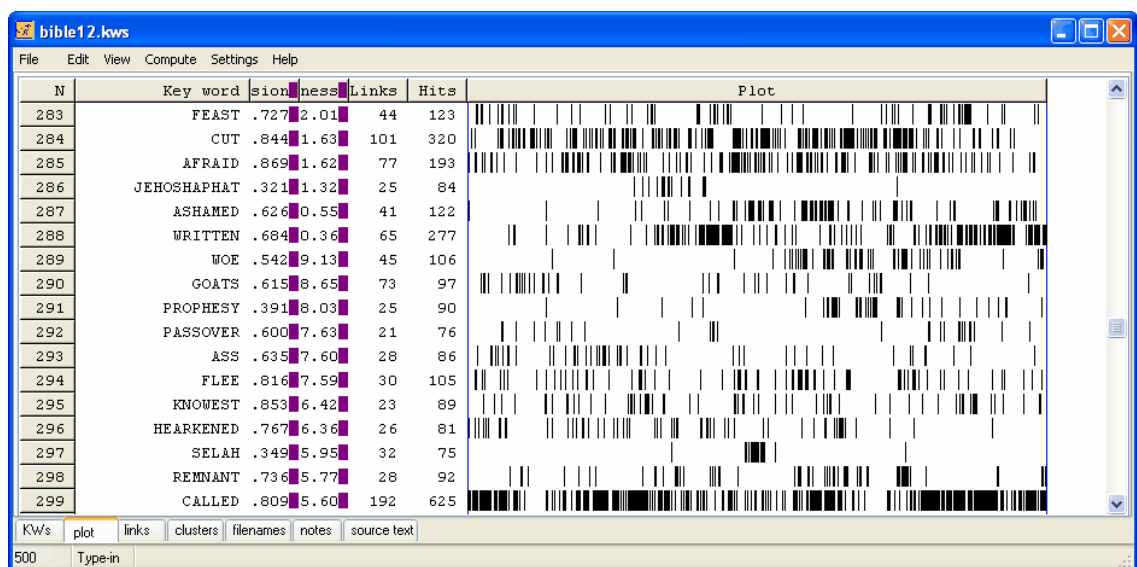
Gegebenenfalls können Sie den [Plot](#) exportieren und andere grafische Darstellungen erstellen, wie unter [Speichern als](#) erklärt wird.

Siehe auch: [Links grafisch darstellen](#), [Plotanzeige der Schlüsselwörter](#)

7.20 Plotanzeige

Der Plot bietet Ihnen wichtige Informationen darüber, wie häufig und an welchen Stellen die verschiedenen Schlüsselwörter im Text vorkommen. Der Plot wird danach [sortiert](#), wo im Text die Schlüsselwörter auftreten. Zuerst werden die Wörter aufgelistet, die am Anfang des Texts vorkommen (z. B. in der Einleitung) und danach diejenigen, die erst später vorkommen.

Der folgende Screenshot zeigt Schlüsselwörter aus der Bibel und wo diese Wörter vorkommen. Z.B. der Name **Jehoshaphat** erscheint erst im zweiten Drittel des Texts.



neu sortieren

Sie können die Liste mit  [neu sortieren](#). Es gibt folgende Möglichkeiten neu zu sortieren:

- erstes Auftreten des Schlüsselworts im Text
- Verteilung im Text
- originale Plotordnung (basierend auf der Gewichtung des Worts)
- alphabetische Sortierung

Gesamtzahl der verknüpften Schlüsselwörter

Links

Hier die die Gesamtzahl der [Links](#) mit anderen Schlüsselwörtern innerhalb desselben Texts im Rahmen des aktuellen Kollokationsbereichs ([Standardeinstellung](#) = 5/5). Diese Zahl gibt an, wie oft das Schlüsselwort maximal fünf Wörter entfernt (nach links oder rechts) von einem anderen Schlüsselwort in Ihrem Plot stand.

Treffer

Diese Spalte zeigt an, wie häufig das Schlüsselwort im Text vorkommt.

Haben Sie einen Plot erstellt, können Sie erkennen, wie bestimmte Wörter mit anderen verknüpft sind. Um das zu tun, klicken Sie auf die Registrierkarte "Links" am unteren Rand. Dort werden alle Schlüsselwörter angezeigt, die mit dem ausgewählten Schlüsselwort am stärksten [verknüpft](#) sind. Das sind die Wörter, die am häufigsten innerhalb des von Ihnen festgelegten Kollokationsbereichs auftreten. Anhand des Fensters "Links" bekommen Sie Einblicke in lexikalische Beziehungen.

Jedes Plotfenster hängt von der Schlüsselwortliste ab, aus der es erstellt wurde. Wenn Sie diese schließen, wird auch das Plotfenster verschwinden. Sie können es *ausdrucken*. Es gibt keine *Speicheroption*, da der Plot aus einer Schlüsselwortliste stammt, die Sie *speichern* sollten ("Speichern" oder "Speichern als"). Es gibt keine Option [als Text](#) speichern, da im Plot Grafiken enthalten sind, die mit Textsymbolen nicht entsprechend dargestellt werden können. Sie können die Grafik jedoch in die [Zwischenablage](#) (Strg+Einfg) *kopieren* und dann als Grafik in ein Textverarbeitungsprogramm einfügen. Sie können auch mit der Option *Output | Daten als Textdatei* Ihre Plotdaten speichern. Nach jedem Wort wird die Gesamtanzahl der Wörter in der Datei angezeigt sowie an wievielter Stelle das Wort jeweils auftritt.

Mithilfe des [Lineals](#) im Menü  können Sie den Plot eines Texts in acht gleiche Segmente bzw. den Plot mehrerer Texte in Textdateiabchnitte unterteilen.

Siehe auch: [Schlüsselwort-Plot](#), [Plot-Verteilungswert](#)

7.21 assoziierte Schlüsselwortgruppen neu gruppieren

Wie mache ich das?

Sie können durch Ziehen Gruppen assoziierter Wörter zusammenführen, z. B. wenn sich deren Schlüsselwörter semantisch ähnlich sind.

Wenn Sie  anklicken, wird Ihnen **KeyWords** mitteilen, welche zwei Gruppen am besten zusammenpassen. Sie sehen dann zwei Listen mit den Wörtern jeweils einer der zwei Gruppen und in der Mitte eine Liste übereinstimmender Wörter. Sie entscheiden aufgrund der Übereinstimmungen, ob die Gruppen zusammengeführt werden oder nicht.

Wenn Sie nicht sicher sind, wählen Sie **Abbrechen**.

Wenn Sie die Gruppen zusammenführen möchten, wählen Sie **Zusammenführen**.

Wenn Sie sich sicher sind, dass Sie die Gruppen **nicht** zusammenführen möchten, und dieses Paar von **KeyWords** nicht mehr vorgeschlagen werden soll, wählen Sie **Überspringen**. Mit **KeyWords** können Sie maximal 50 Paare speichern, die übersprungen werden. Um die Elemente, die übersprungen werden, aus dem Speicher zu entfernen, wählen Sie **Übersprungene Elemente löschen**.

Wozu dient das (2)?

[Scott](#) (1997) zeigt anhand einer Reihe von Sonderbeiträgen aus dem *Guardian*, wie assoziierte Schlüsselwortgruppen die Wahrnehmung von verschiedenen Frauen beeinflussen.

Siehe auch: [assozierte Gruppen](#)

7.22 neu sortieren: KeyWords

Wie mache ich das?

Neu sortieren können Sie entweder durch Anklicken der obersten Reihe einer Liste, durch Drücken von F6 bzw. Strg+F6 oder mit der Menüoption "Neu sortieren". Sortieren Sie erneut, um von aufsteigender Sortierung zu absteigender Sortierung zu wechseln.

Eine **Schlüsselwortliste** kann sortiert werden nach:

Gewichtung (das *wichtigste* Schlüsselwort erscheint am Anfang der Liste)
 alphabetische Sortierung (von A bis Z)
 Frequenz in der kürzeren Liste (das häufigste Wort steht an erste Stelle)
 Frequenz in der Referenzliste (das häufigste Wort steht an erste Stelle)

Ein **Schlüsselwort-Plot** kann sortiert werden nach:

Gewichtung (das *wichtigste* Schlüsselwort erscheint am Anfang der Liste)
 alphabetische Sortierung (von A bis Z)
 Frequenz (das häufigste Wort erscheint an erster Stelle)
 Anzahl der Links (das Wort mit den meisten Links steht an erste Stelle)
 erstes Auftreten des Schlüsselworts im Text
 Bereich (Wörter, die nur in kleinen Abschnitten des Texts auftreten, kommen am Anfang)

Eine **Datenbank gewichteter Schlüsselwörter** kann sortiert werden nach:

Frequenz (das *häufigste* Schlüsselwort erscheint am Anfang der Liste)
 alphabetische Sortierung (von A bis Z)

Eine Liste [assoziierter Wörter](#) kann sortiert werden nach:

Frequenz (Assoziation zwischen Titelwort und Wort)
 alphabetische Sortierung (von A bis Z)
 Frequenz (Assoziation zwischen Wort und Titelwort)

7.23 Anzeige der Schlüsselwörter

In der Anzeige sind zu sehen:

1. jedes Schlüsselwort
2. seine Frequenz in den Quelltexten, in denen es auftritt, in *Kursivschrift*.
3. der Name der Quelltextdatei (oder der Wortlistendatei, wenn es mehrere gibt) und %, auch in *Kursivschrift*.
4. seine Frequenz im Referenzkorpus
5. der Name des Referenzkorpus (oder der Korpuswortlistendatei, wenn diese auf mehreren Texten basiert) und %.
6. Gewichtung (Chi-Quadrat- oder Log-Likelihood-[Statistik](#))
7. [P-Wert](#).

Wie ungewöhnlich die Frequenz ist, wird anhand der angewandten [statistischen Verfahren](#) berechnet. Die Statistik wird rechts angezeigt. Wenn der Chi-Quadrat-Test angewandt wird, und es herrschen die üblichen Bedingungen (erwarteter Wert ≥ 5 in allen vier Zellen), bzw. wenn das Log-Likelihood-Verfahren angewandt wird, so wird die Wahrscheinlichkeit (p für engl. 'probability') rechts vom Chi-Quadrat-Wert angezeigt.

Das Kriterium dafür, wie "ungewöhnlich" die Frequenz ist, basiert auf dem minimalen Wahrscheinlichkeitswert, der vor der Berechnung der Schlüsselwörter angegeben wurde. Je kleiner der Wert ist, desto weniger Schlüsselwörter werden angezeigt. Im Normalfall sind 40 Schlüsselwörter vollkommen ausreichend.

Die Wörter erscheinen [geordnet](#) danach, wie ungewöhnlich die Frequenz ihres Auftretens ist. Die Frequenz der Wörter am Anfang der Liste ist am ungewöhnlichsten. Am Ende der Liste finden Sie die Wörter, die in ihrer Frequenz ungewöhnlich [selten](#) sind (negative Schlüsselwörter), in einer anderen Farbe.

Schaltfläche "Ansicht"

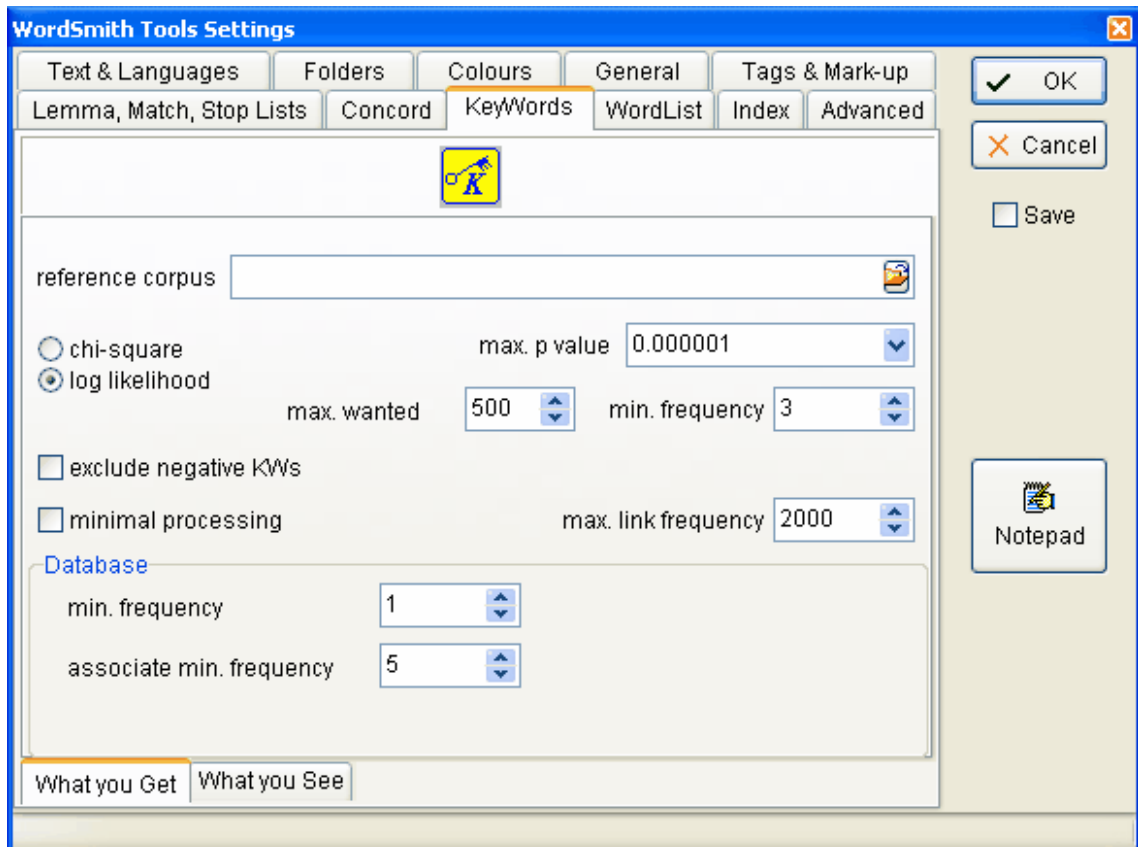
Klicken Sie diese Schaltfläche an, wird mit dem Utility [Viewer](#) der Originaltext angezeigt und die

Schlüsselwörter werden optisch hervorgehoben.

Gehen Sie zu  [Layout](#), um die verschiedenen Farben oder Schriftarten für die einzelnen Datenspalten zu ändern, z. B. wenn Ihnen die Kursivschrift nicht gefällt.

7.24 WordSmith Controller: KeyWords-Einstellungen

These are found in the main [Controller](#) under *Adjust Settings | KeyWords*.



This is because some of the choices may affect other Tools. KeyWords and WordList both use similar routines: KeyWords to calculate the key words of a text file, and WordList when comparing [comparing word-lists](#).

Procedure

Chi-square or Log Likelihood. The default is Log Likelihood. See [procedure](#) for further details.

Max. p value

The default level of significance. See [p value](#) for more details.

Max. wanted (500) and Min. frequency (3)

You may want to restrict the number of key words (KW's) identified so as to find for example the ten most "key" for each text. The program will identify all the key words, sort them by key-ness, and then throw away any excess. It will thus favour [positive key words](#) over negative ones. The minimum frequency is a setting which will help to eliminate any words or clusters which are unusual but infrequent. For example, a proper noun such as the name of a village will usually be extremely infrequent in your reference corpus, and if mentioned only once in the text you're

analysing, it is likely not to be "key". The default setting of 3 mentions as a minimum helps reduce spurious hits here. In the case of short texts, less than 600 words long, a minimum of 2 will automatically be used.

Exclude negative KWs

If this is checked, KeyWords will not compute negative key words (ones which occur significantly *infrequently*).

Minimal processing

If this is checked, KeyWords will not compute [plots](#), [links](#) or [KW clusters](#) as it computes the key words (they can always be computed later assuming you do not move or delete the original text files). This is useful if computing a lot of KW files in a batch, eg. to make a database.

Max. link frequency

To compute a plot is hard work as all the KWs have to be concordanced so as to work out where they crop up. To compute links between each KW is much harder work again and can take time especially if your KWs include some which occur thousands or hundreds of times in the text. To keep this process more manageable, you can set a default. Here 2000 means that any KW which occurs more than 2000 times in the text will not be used for computing [links](#). (It will still appear in the plots and list of KWs, of course.)

Database: minimum frequency

The default is 1. See [database](#).

Database: associate minimum frequency

The default is 5. See [associates](#).

See also: KeyWords Help Contents, [KeyWords calculation](#).

WordSmith Tools

WordList

Section

VIII

8 WordList

8.1 Zweck



Dieses Programm generiert Wortlisten basierend auf einer oder mehreren [ANSI](#)- oder [ASCII](#)-Textdateien. Die Wortlisten werden automatisch generiert, sowohl alphabetisch und als auch nach Frequenz sortiert. Sie haben auch die Option, eine [Wortindex](#)liste zu erstellen.

Wozu dient das?

Diese Indexliste kann man dazu verwenden

- 1 das in den Texten verwendete Vokabular zu untersuchen;
- 2 häufige Wort-[Cluster](#) zu identifizieren;
- 3 die Frequenz eines Worts in verschiedenen Textdateien bzw. Textsorten zu vergleichen;
- 4 die Frequenzen von verwandten Wörtern oder Übersetzungsäquivalenten in [verschiedenen Sprachen](#) zu vergleichen.

Innerhalb von **WordList** können Sie zwei [Listen](#) vergleichen oder für einen stilistischen Vergleich eine Konsistenzanalyse durchführen ([einfach](#) oder [detailliert](#)).

Diese Wortlisten können auch als Input für das Programm [KeyWords](#) verwendet werden, in dem die Wörter in einem bestimmten Text analysiert und deren Frequenzen mit einem Referenzkorpus verglichen werden, um daraus Listen mit "Schlüsselwörtern" und "gewichteten Schlüsselwörtern" zu erstellen.

Siehe auch: [WordList-Anzeige](#)

8.2 Lemmata automatisch zusammenführen

Mit der Menüoption *Automatisch verbinden* können Sie Strings wie **S** oder **S;ED;ING** festlegen, und dann werden alle Einträge aus der gesamten Wortliste lemmatisiert, in denen sich ein Wort vom nächsten Wort nur durch ein **S** oder **ED** oder **ING** am Ende unterscheidet. (Verwenden Sie **;** für mehrere separate Suffixe.)

Präfix / Suffix / Infix

Standardmäßig werden alle eingegebenen Strings als Suffixe gewertet; um Präfixe zu verbinden, fügen Sie ein Sternchen (*) am Ende des Präfixes hinzu. Wenn Sie nach Infixen suchen möchten (z. B. **bloody** in **absobloodylutely** [Sprachen wie Suaheli verwenden häufig Infixe]), fügen Sie jeweils ein Sternchen vorne und hinten hinzu.

Beispiele

S;ED;ING wird das Wort **book** mit den Wörtern **books**, **booked** und **booking** verbinden.

***S; *ED; *ING** wird das Wort **book** mit den Wörtern **books**, **booked** und **booking** verbinden.

UN*;ED;ING wird das Wort **do** mit **undo** verbinden, und **booked** und **booking** mit **book** verbinden.

BLOODY wird das Wort **absobloodylutely** mit dem Wort **absolutely** verbinden.

Dieser Vorgang kann entweder automatisch im Hintergrund ablaufen, oder Sie müssen das Verbinden der einzelnen Wörter jedes Mal bestätigen. Automatische Lemmatisierung, wie auch die Rechtschreibprüfung mit automatischem Suchen und Ersetzen, kann zu kuriosen Ergebnissen führen.

Um während dem automatischen Zusammenführen anzuhalten, drücken Sie Escape.

Tipp

Mit einer zuvor gespeicherten Liste können Sie ausprobieren, die Einträge automatisch *ohne* Bestätigung der Änderungen zusammenzuführen (oder Sie wählen während des Vorgangs *Ja, alle*). Danach wählen Sie die alphabetische Version der Liste (anstelle der nach Frequenz geordneten Liste) und sortieren die Liste nach Lemmata (indem Sie auf die Überschrift der *Lemmaspalte* klicken). Sie werden dann alle verbundenen Einträge am Anfang der Liste sehen. Es ist unter Umständen einfacher, mögliche Fehler durch [Trennen](#) (Strg+F4) rückgängig zu machen, als jedes Verbinden einzeln zu bestätigen. Abschließend nach *Wort* sortieren und speichern.

Siehe auch: [Lemmatisierung](#)

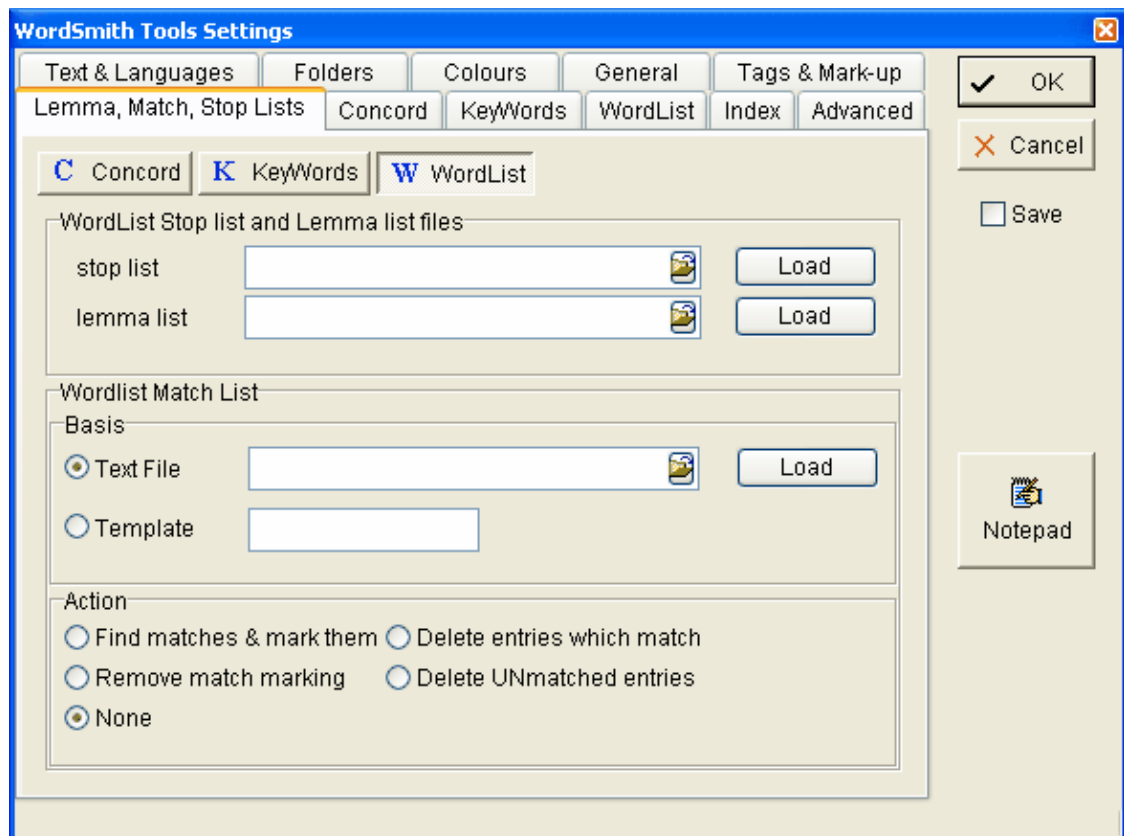
8.3 Lemmadatei auswählen

Wozu dient das?

Sie können alle Elemente in Ihrer aktuellen Wortliste lemmatisieren, indem Sie eine Standardtext-Datei erstellen, in der Wörter, die zusammen gehören, gruppiert werden (sein -> bin, bist, ist, war, waren, etc.). Obwohl es ziemlich lange dauert, diese Textdatei zu erstellen, wird sie sehr nützlich sein, wenn Sie viele Wortlisten lemmatisieren müssen, und diese Methode ist dazu sehr viel zuverlässiger als [automatisch zusammenführen](#).

Wie mache ich das?

Im Controller unter *Einstellungen* | *Einstellungen anpassen* | *Listen*, sehen Sie folgendes Fenster:



Klicken Sie auf die entsprechenden Schaltfläche (für Concord, KeyWords oder WordList) und geben Sie den Dateinamen ein oder suchen Sie nach der Datei.

Die Datei sollte eine einfache Textliste beinhalten, bestehend aus Lemmata mit Elementen wie den folgenden:

SEIN -> BIN, BIST, WAR, WAREN, IST
GEHEN -> GEHT, GEHT, GING, GINGEN

Im **Controller** wird diese Datei dann gelesen und die Elemente werden angezeigt (oder nur eine Stichprobe daraus, wenn es eine lange Liste ist).

Der eigentliche Vorgang geschieht nur dann, wenn Sie in WordList, Concord oder KeyWords die Menüoption *Lemmata übereinstimmen* () auswählen. Siehe [Matchliste](#) für eine genauere Beschreibung mit Screenshots.

Was, wenn meine Textdateien SEIN nicht beinhalten?

Angenommen Sie möchten **BIN, BIST**, etc. mit **SEIN** wie in der Liste oben übereinstimmen, aber Ihre Texte enthalten das Wort **SEIN** gar nicht. WordList kann dieses Wort also nicht finden, und deshalb auch nicht mit den anderen Wörtern zusammenführen. Die beste Lösung dafür ist die Erstellung einer neuen [Wortliste basierend auf einer Standardtext-Datei](#) (in der Sie **SEIN** und andere Basisformen angeben). Speichern Sie diese Wortliste und führen Sie die Liste mit Ihrer bereits bestehenden Wortliste [zusammen](#). Nun sollte WordList auch die Form **SEIN** finden und **BIN, BIST, WAR**, etc. dazu hinzufügen.

Siehe auch: [Lemmatisierung](#), [Matchliste](#), [Stopplisten](#)

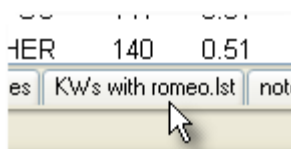
8.4 Wortlisten vergleichen

Damit sollen stilistische Vergleiche erleichtert werden. Angenommen Sie untersuchen die verschiedenen Versionen einer Geschichte oder verschiedene Übersetzungen dieser Geschichte. Wenn in einer Version das Wort *töten* und in einer anderen Version das Wort *ermorden* verwendet wird, können Sie das mit dieser Funktion herausfinden.

In diesem Verfahren werden alle Wörter in *beiden* Listen verglichen, und diejenigen Wörter angezeigt, die in einer Version deutlich häufiger vorkommen als in der anderen Version, inklusive der Wörter, die in einer Version über einer Mindestfrequenz liegen, in der anderen Version jedoch gar nicht vorkommen.

Wie mache ich das?

1. Öffnen Sie eine Wortliste.
2. Im Menü wählen Sie *Datei | 2 Wortlisten vergleichen*.
3. Wählen Sie die Wortliste, mit der Sie die bereits geöffnete Wortliste vergleichen möchten. Die Ergebnisse werden in einer der Registerkarten am unteren Rand des Bildschirms angezeigt.



Die Mindestfrequenz (die Sie im [Controller](#) unter *Einstellungen anpassen* in der Registerkarte *KeyWords* verändern können), kann auf 1 festgelegt werden. Wird dieser Wert auf 3 erhöht, werden bei dem Vergleich alle Wörter ignoriert, die weniger als dreimal in mindestens einer der Listen vorkommen.

Legen Sie den Signifikanzwert fest (*alle*, oder ein [P-Wert](#) zwischen 0,1 und 0,000001 oder beliebiger anderer Wert). Je niedriger der [P-Wert](#) ist, desto selektiver ist der Vergleich. Mit anderen Worten, wird der P-Wert auf 0,1 festgelegt, dann werden mehr Wörter angezeigt als bei 0,0001.

Das [Anzeige](#)format ist ähnlich wie das Format, das in [KeyWords](#) verwendet wird.

Siehe auch: [Konsistenzanalyse](#), [Matchliste](#)

8.5 Wortlisten zusammenführen

Wozu dient das?

Sie möchten zwei Wortlisten (oder Konkordanzen, Transinformationslisten, etc.) zusammenführen, weil die Erstellung jeder Liste zu lange dauert oder weil Sie schrittweise eine Master-Wortliste oder -Konkordanz basierend auf einer Vielzahl von Texten aus verschiedenen Bereichen oder Textsorten erstellen möchten.

Wie mache ich das?

Wenn Sie eine Wortliste (oder Konkordanz) geöffnet haben, wählen Sie *Datei* | *Zusammenführen mit* an und wählen die zweite Liste aus.

Was ist dabei zu beachten?

Wenn Sie zwei Wortlisten zusammenführen möchten, ist davon auszugehen, dass die Wortlisten auf unterschiedlichen Quelltexten basieren. Beinhalten diese zwei Wortlisten nun Informationen über dieselbe Textdatei, wird WordSmith zwar den Vorgang durchführen, aber die Informationen über Frequenz bzw. in welchen Texten die Wörter vorkommen, werden wahrscheinlich falsch sein.

Zusammenführung von Listen mit Wörtern in verschiedenen Sprachen: Wenn Sie mit einer spanischen Liste beginnen, und diese mit einer englischen Liste zusammenführen möchten, wird die englische Liste z. B. bezüglich Sortierreihenfolge wie eine spanische Liste behandelt. Wenn Sie eine deutsche Liste mit einer arabischen Liste zusammenführen, werden wahrscheinlich alle Formen angezeigt, aber das Ergebnis wird ein anderes sein, als wenn Sie die arabische Liste zu der englischen Liste hinzufügen (alle arabischen Wörter werden behandelt, als wären es englische Wörter).

8.6 Vergleichsanzeige

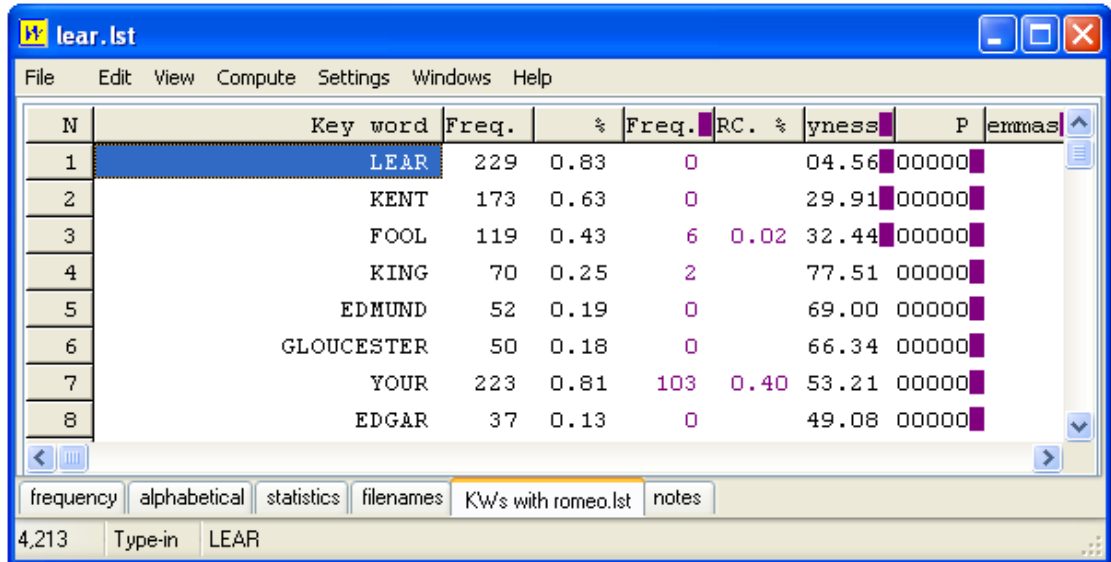
Here is a comparison window, where we have compared Shakespeare's *King Lear* with *Romeo and Juliet*.

The display shows frequency in the text you started with, here *King Lear*, (with % if > 0.01%) -- then, to the right frequency in the other text, here *Romeo & Juliet*, (with % if > 0.01%) -- then, to the right [chi-square or log likelihood](#), and p [value](#).

The criterion for what counts as "outstanding" is based on the minimum probability value entered before the lists were compared. The smaller this probability value the fewer words in the display. The words appear sorted according to how outstanding their frequencies of occurrence are. Those near the top are outstandingly frequent in your main wordlist. At the end of the listing you'll find those which are outstandingly infrequent in the first text chosen: in other words, key in the second text.

This comparison is similar to the analysis of "key words" in the [KeyWords](#) program. The KeyWords analysis is slightly quicker and allows for batch processing.

The word **Lear** is the most key of all, it scores 304 on the keyness column. (It looks like 04.56 because the column hasn't been pulled any wider.)

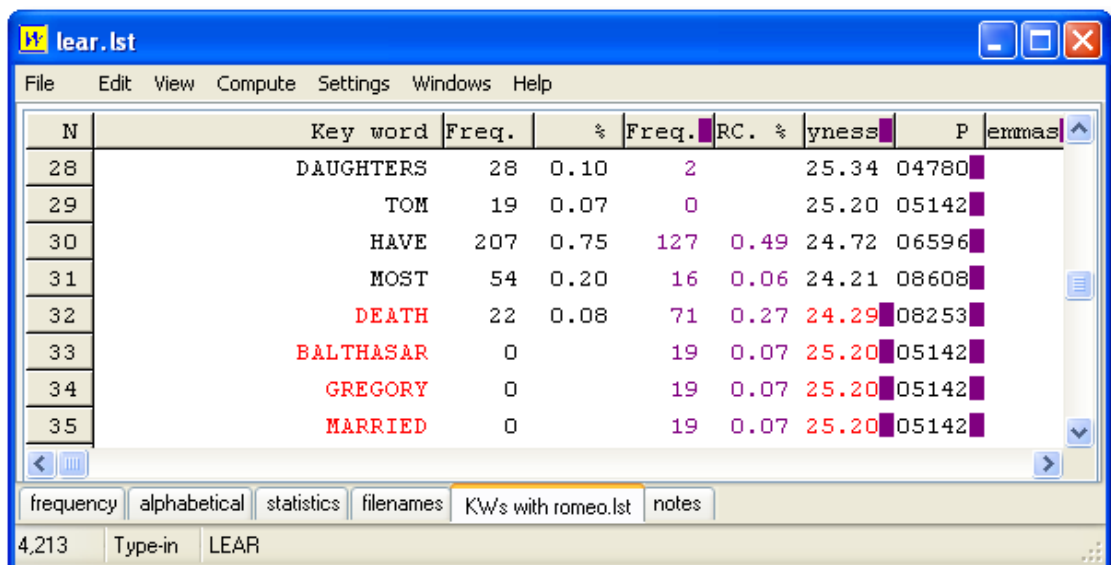


The screenshot shows the WordSmith Tools window titled 'lear.lst'. The menu bar includes File, Edit, View, Compute, Settings, Windows, and Help. The main table displays key words for the file 'lear.lst'. The columns are N, Key word, Freq., %, Freq., RC. %, yness, P, and emmas. The data is as follows:

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	

At the bottom, the status bar shows '4,213 Type-in LEAR'.

The words above, in black, are key to Lear. Below, we see the middle of the listing --- the words in red are those which are key to Romeo. The word **most** is the last key word of Lear, and **death** the least key in Romeo; both have a keyness value of around 25 (positive or negative).

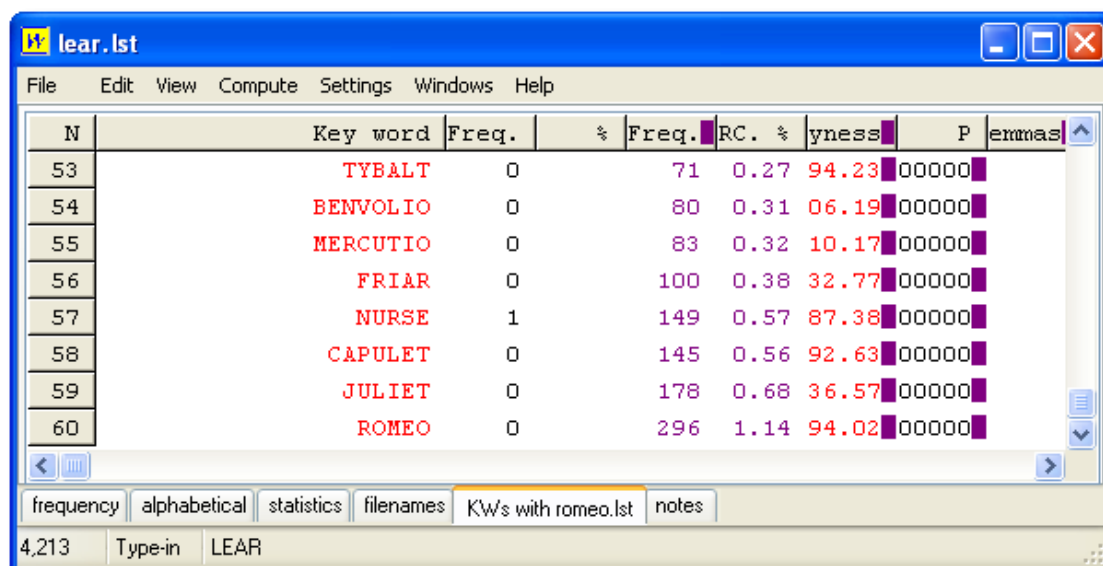


The screenshot shows the WordSmith Tools window titled 'lear.lst'. The menu bar includes File, Edit, View, Compute, Settings, Windows, and Help. The main table displays key words for the file 'lear.lst'. The columns are N, Key word, Freq., %, Freq., RC. %, yness, P, and emmas. The data is as follows:

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	

At the bottom, the status bar shows '4,213 Type-in LEAR'.

Here at the bottom we see the words which are most key to the play *Romeo and Juliet*.



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	

frequency alphabetical statistics filenames Kw's with romeo.lst notes

4,213 Type-in LEAR

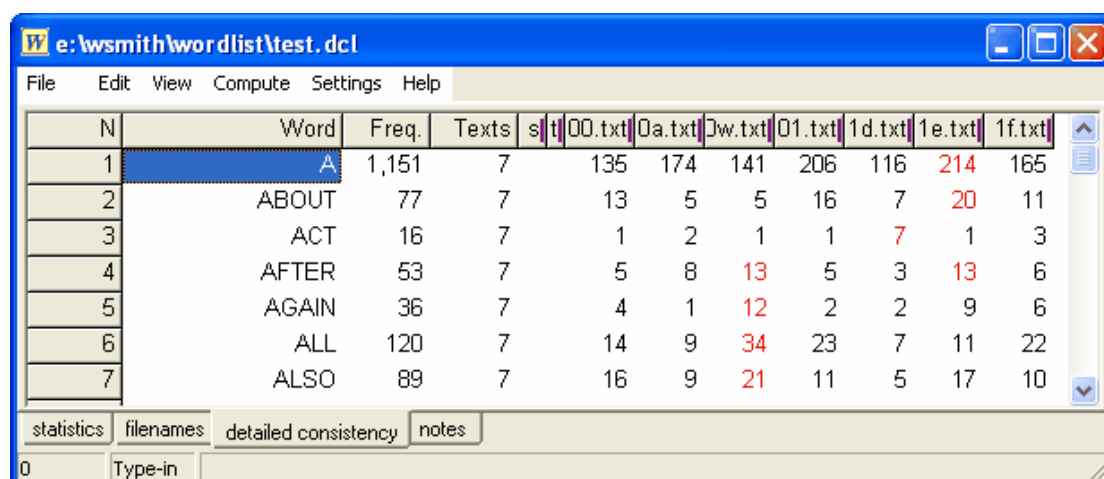
The word which is most outstanding (key) here is Romeo, with a keyness score of 394 (the column needs to be pulled wider).

8.7 Konsistenzanalyse (detailliert)

Diese Funktion unterscheidet sich von der [einfachen Konsistenzanalyse](#) nur dadurch, dass sie detailliertere Informationen bietet.

Wozu dient das?

Damit sollen stilistische Vergleiche erleichtert werden. Angenommen Sie untersuchen die verschiedenen Versionen einer Geschichte oder verschiedene Übersetzungen dieser Geschichte. Diese Funktion ermöglicht es, alle Wörter anzuzeigen, die in den geöffneten Wortlisten enthalten sind. Die Wörter werden in der Anzeige so geordnet, dass die erste Gruppe alle Wörter enthält, die in allen Versionen vorkommen, danach kommen alle Wörter, die nur in einer Version nicht vorkommen, und so weiter bis zu den Wörtern, die nur in einer Version vorkommen.



N	Word	Freq.	Texts	s	t	00.txt	0a.txt	0w.txt	0l.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		

statistics filenames detailed consistency notes

0 Type-in

Innerhalb jeder Gruppe werden die Wörter alphabetisch geordnet. Die Spalte Freq. zeigt an, wie oft jedes Wort insgesamt vorkommt, Texte zeigt an, in wie vielen Textdateien das Wort vorkommt. Danach gibt es zwei Spalten (Nr. der Lemmata und Klasse, die sich wie in einer Wortliste verhalten) und danach eine Spalte für jeden Text. In diesem Fall kam das Wort **about** in allen sieben Texten insgesamt 77 Mal vor, davon am häufigsten, nämlich **20** Mal, in **1e.txt**. Statistik und Dateinamen für die sieben verwendeten Texte können durch Klicken auf die Registerkarten unten angezeigt werden. Notizen können bearbeitet und gemeinsam mit der detaillierten Konsistenzliste gespeichert werden.

Beachten Sie bitte, dass der Dateiname **text.dcl** (**detailed consistency list**) lautet. In Bezug auf die Anzahl der Text, die bei diesem Vorgang bearbeitet werden können, gibt es keine Begrenzung, außer dem begrenzten Arbeitsspeicher.

Wie mache ich das?

Wenn Sie in dem angezeigten Fenster auf **Neu...** () klicken, wird eine Registerkarte für die detaillierte Konsistenzanalyse angezeigt. Wählen Sie Ihre Wortlisten aus und klicken Sie auf **Detaillierte Konsistenz jetzt berechnen**.

Jede Spalte kann durch Anklicken der Spaltenüberschrift (**Wort**, **Freq.**, etc.) sortiert werden. Um die Wörter, die in allen sieben Texten vorkamen, an den Anfang der Liste zu verschieben, klicken Sie auf **Texte**.

Siehe auch: [Konsistenzanalyse \(einfach\)](#), [Vergleichsanzeige](#), [Wortlisten vergleichen](#), [Matchliste](#), [Spaltensummen](#)

8.8 Konsistenzanalyse (einfach)

Diese Funktion (von [Paul Nation](#) "range" benannt) wird mit jeder Wortliste aufgerufen.

In jeder Wortliste gibt es eine Spalte mit der Überschrift "Texte". Hier wird die Anzahl der Texte angezeigt, in denen das Wort vorkommt (maximal die Gesamtanzahl der Texte, aus denen Sie die Wortliste erstellt haben).

Wozu dient das?

Mit dieser Funktion kann man herausfinden, welche Wörter wiederholt in vielen Texten derselben Textsorte vorkommen. Zum Beispiel in einer Sammlung englischer Geschäftsberichten erscheint das Wort **consolidate** in sehr vielen Texten. Es kommt in den einzelnen Texten nicht sehr häufig vor, aber in Geschäftsberichten erscheint dieses Wort konsistenter als in einer gemischten Textsammlung.

Logischerweise sind Wörter wie **the** fast in allen englischen Texten konsistent.

(Interessanterweise ich bei meiner Arbeit mit Geschäftsberichten einen Text gefunden, in dem das Wort **the** nicht vorkam. Ich habe auch herausgefunden, dass einer meiner Texte Italienisch war: aber das war nicht derselbe Text, in dem **the** nicht vorkam! Der Text ohne **the** war eine Wahlergebnisliste, in denen **Cons.**, **Lab.** und Ortsnamen häufig vorkamen, aber **the** kein einziges Mal.)

Bei der Analyse häufiger grammatischer Wörter wie **und**, kann eine Konsistenzliste sehr hilfreich sein. Dennoch werden einige gebräuchliche lexikalische Elemente erstaunlich konsistent auftreten.

Um gebräuchliche Wörter auszuschließen und nur diejenigen zu finden, die charakteristisch für Ihre Textsorte oder Textsammlung sind, müssen Sie herausfinden, welche der Wörter signifikant konsistent sind. Speichern Sie Ihre Wortliste und [vergleichen](#) Sie die Liste mit anderen Listen in WordList oder in KeyWords. Auf diese Weise können Sie bestimmen, welche Wörter in Ihrer Textsorte oder untergeordneten Textsorte signifikant konsistent sind.

Siehe auch: [Konsistenzanalyse \(detailliert\)](#), [Worlisten vergleichen](#), [Matchliste](#)

8.9 Lemmata

Möglicherweise möchten Sie einige Einträge gemeinsam speichern: z. B. **arbeiten**, **arbeite**, **arbeitest**, **arbeitete** als verschiedene Formen desselben Lemmas.

Manuelles Zusammenführen

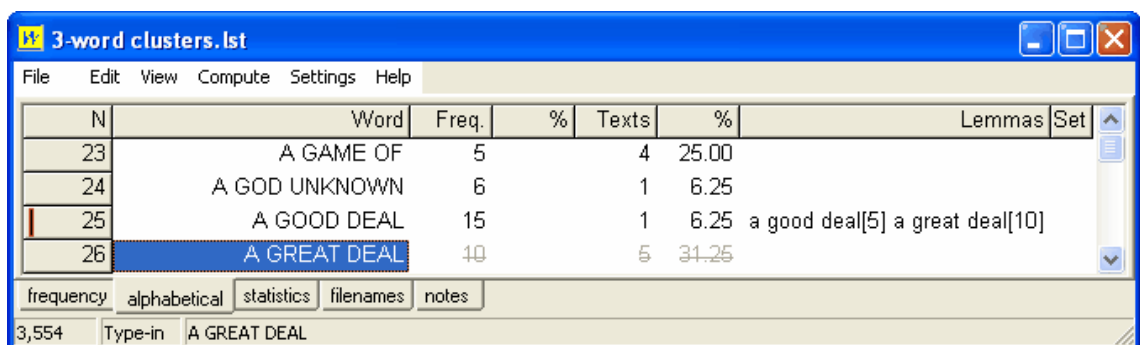
Durch Ziehen eines Eintrags zu einem anderen Eintrag werden die Einträge zusammengeführt. Angenommen Ihre Wortliste enthält

ARBEITEN
ARBEITE
ARBEITEST

Dann können Sie mit der Maus den Eintrag **arbeite** oder **arbeitest** einfach auf **arbeiten** ziehen.

(Siehe [Lemmadatei auswählen](#), wenn Sie diese Einträge mit einem Eintrag zusammenführen möchten, der nicht in der Liste ist.)

Sowohl die alphabetische Liste als auch die Frequenzliste werden ordnungsgemäß aktualisiert, obwohl die Frequenzliste möglicherweise nicht richtig sortiert ist, solange die Datei nicht durch das [Entfernen von Einträgen](#) neu sortiert wurde. Ein lemmatisierter Haupteintrag ist mit einer roten Markierung am linken Rand gekennzeichnet. Die anderen markierten Einträge werden farblich so markiert, als wären sie gelöscht worden. Die verknüpften Einträge, die zum Haupteintrag hinzugefügt wurden, werden rechts angezeigt.



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		

frequency alphabetical statistics filenames notes

3,554 Type-in A GREAT DEAL

Hier sieht man eine Wortliste basierend auf [3-Wort-Clustern](#) (bestehend aus drei Wörtern), von denen **a good deal** ursprünglich eine Frequenz von 5 aufwies. Nach der Zusammenführung zu dem Eintrag **a great deal** (Frequenz = 10) beträgt deren Frequenz nun 15.

Wenn nicht alle Elemente, die Sie zusammenführen möchten, in einem Bildschirm angezeigt werden, können Sie die Einträge mittels der Funktionstasten zusammenführen.

1. Mit F5 markieren Sie die Einträge, die Sie zusammenführen möchten. Der zuerst markierte Eintrag ist der "Haupteintrag". Während dieses Vorgangs wird die aktuelle Zeile am Rand grün markiert. Bei allen weiteren Einträgen, die durch erneutes Drücken von F5 mit dem Haupteintrag verknüpft werden, ist die Markierung weiß. (Durch erneutes Drücken von F5 kann die Markierung wieder aufgehoben werden.)

2. Zum Zusammenführen aller markierten Einträge drücken Sie F4. Das Programm zählt die Frequenzen aller Einträge mit der Frequenz des zuerst markierten Eintrags (dem Haupteintrag) zusammen.

Trennen

Wenn Sie ein Element auswählen, das rechts sichtbare Lemmata aufweist, und Strg+F4 drücken, werden die Einträge wieder getrennt.

Dateibasiertes Zusammenführen

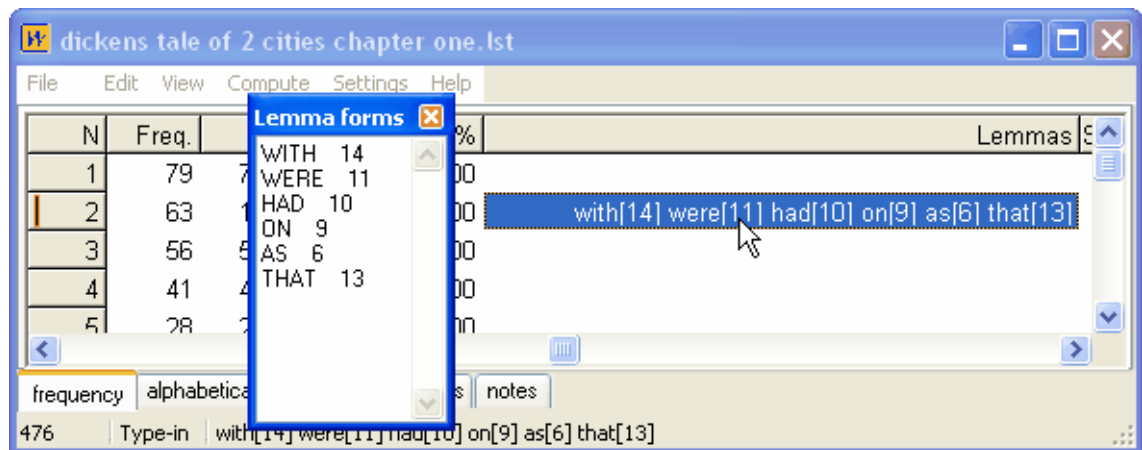
Sie können die Lemmata auch mittels einer [Textdatei](#) zusammenführen, wodurch der Vorgang des Übereinstimmens und Zusammenführens automatisch abläuft. Der eigentliche Vorgang geschieht dann, wenn Sie in WordList, Concord oder KeyWords die Menüoption *Lemmata übereinstimmen* () auswählen. Jeder Eintrag in Ihrer Lemmaliste wird darauf überprüft, ob er mit einem der Einträge in Ihrer Wortliste übereinstimmt. Wenn zum Beispiel die Wörter **bin**, **war** und **waren** vorkommen, werden sie als Lemmata von **sein** gespeichert. Wenn die Wörter **gehen** und **ging** gefunden werden, wird **ging** mit dem Eintrag **gehen** zusammengeführt.

Automatisch zusammenführen

Um den Lemmatisierungsvorgang zu beschleunigen, können Sie alle Einträge in der aktuellen Liste automatisch zusammenführen, die Ihre Kriterien erfüllen.

Nicht alle Lemmaformen werden angezeigt

Auf die Lemmaspalte doppelklicken, wie in diesem Screenshot,



und ein Fenster mit allen Lemmaformen wird geöffnet und die verschiedenen Komponenten werden angezeigt.

Siehe auch: [Automatisch zusammenführen](#), [Eine Textdatei für Lemmatisierung verwenden](#)

8.10 Indexlisten: Verwendungen

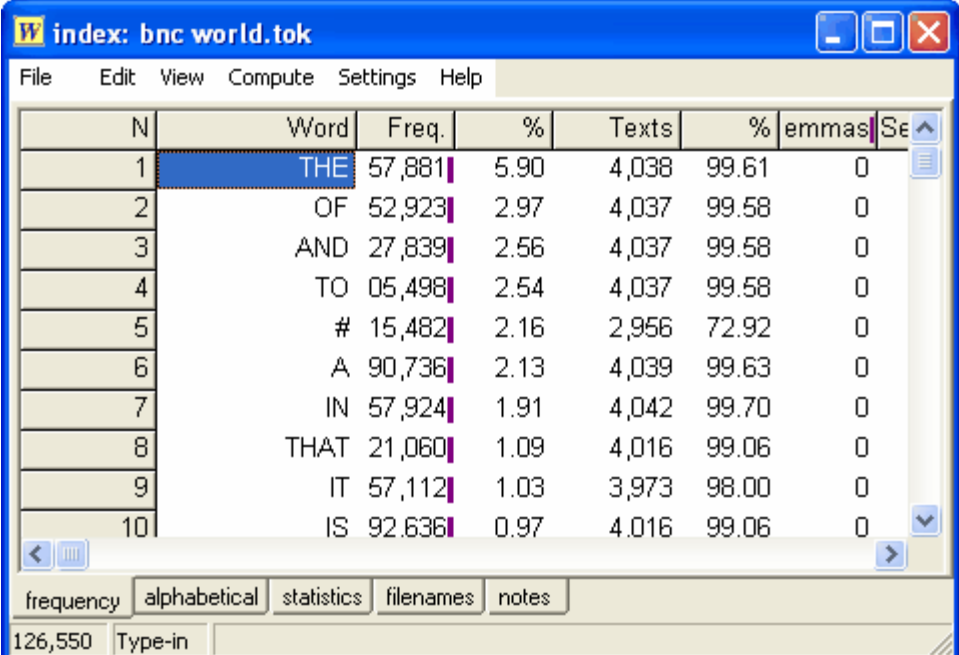
Wozu dient das?

1. Eine der Verwendungen für einen Index ist das Protokollieren der Positionen aller Wörter in Ihrer Textdatei, damit Sie danach sehen können, welches Wort in welchem Abschnitt des Texts vorkommt. Eine weitere Verwendung ist der schnellere Zugang zu diesen Wörtern, zum Beispiel bei der Erstellung einer Konkordanz. Wenn Sie im Index ein Wort oder mehrere Wörter auswählen und  anklicken, erhalten Sie eine schnelle Konkordanz.
2. Außerdem können Sie Indexe dazu verwenden, "[Transinformationswerte](#)" zu berechnen, bei denen Wort-Types miteinander in Beziehung gebracht werden.
3. Mit einem Index können Sie auch [Wortcluster](#) sehen.

Siehe auch: [Indexliste erstellen](#), [Indexlisten anzeigen](#), Inhalt der WordList-Hilfe.

8.11 Indexlisten: anzeigen

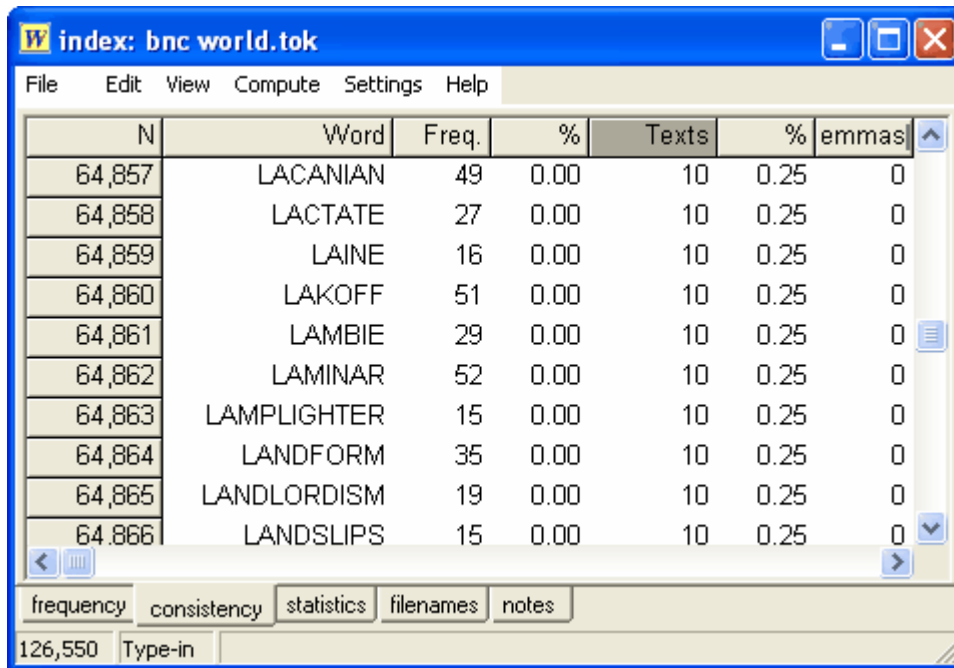
In WordList können Sie einen Index genauso wie jede andere Wortlistendatei öffnen, und zwar unter Datei | Öffnen. Sie können auch im Controller in der Registerkarte Vorherige Listen einen bereits vorhandenen Index auswählen und doppelklicken. Der Index wird dann in Form einer langen Wortliste angezeigt.



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	

Das obere Bild zeigt die häufigsten 10 Wörter im BNC World Corpus. Nummer 5 (#) steht für Elemente, die Nummern enthalten, wie z. B. €50,00. Diese häufigen Wörter sind auch sehr konsistent -- sie erscheinen in mindestens 99% der 4.054 Texte von BNC World.

In der unteren Ansicht sehen Sie Wörter, die nach der Anzahl der Texte sortiert wurden: alle diese Wörter erscheinen 10 Mal im Corpus, ihre Frequenzen sind jedoch unterschiedlich.



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

Sie können ein oder mehrere Wörter hervorheben oder mit der Option  markieren, und danach auf  klicken, um eine schnelle Konkordanz zu erstellen.

Siehe auch: [Indexliste erstellen](#), [WordList-Cluster](#), Inhalt der WordList-Hilfe.

8.12 einen WordList-Index erstellen

Um einen Index zu erstellen, wählen Sie zuerst im [Controller](#) im Menü *Einstellungen* | *Einstellungen anpassen* | *Index*. Hier müssen Sie den Dateinamen für den Index angeben.

WordList durchsucht die von Ihnen ausgewählten Texte und speichert Informationen darüber, welche Wörter an welchen Stellen jeweils vorkommen.

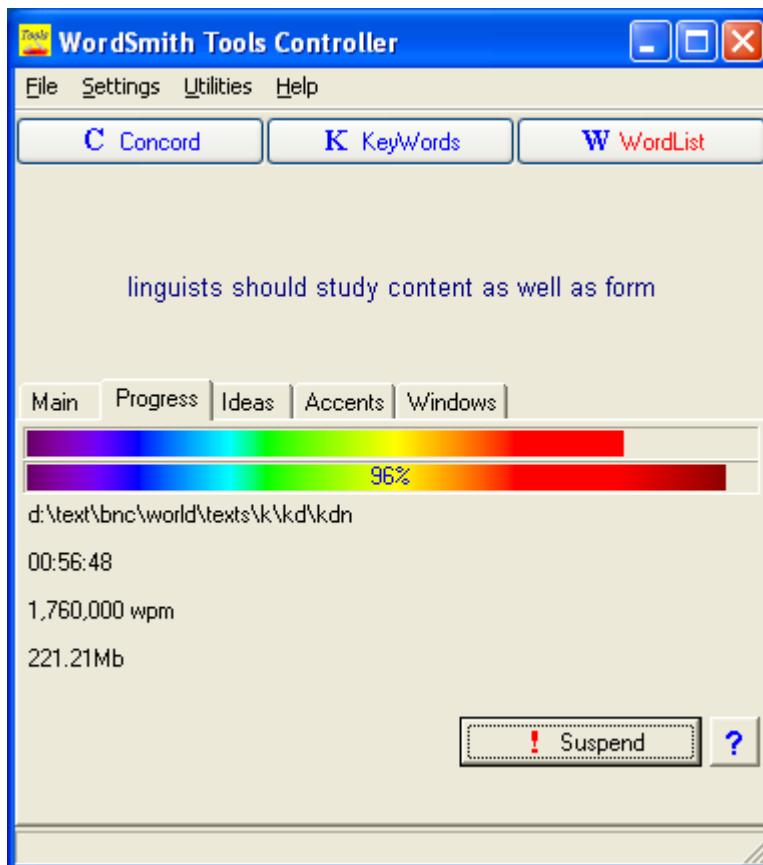
Indexdateien

Für jeden Index werden zwei Dateien erstellt:

eine **.tok**-Datei: eine große Datei, in der die Position jedes einzelnen Worts (Token) gespeichert wird.

eine **.typ**-Datei: speichert alle verschiedenen Wörter (Types), die in der Datei vorkommen.

Mithilfe eines Index können für jedes Wort die [Wortcluster](#)- und [Transinformations](#)-Werte berechnet werden. Im folgenden Screenshot sieht man die Statusanzeigen für einen Index des BNC World Corpus. Es hat eine Stunde gedauert, um 96% des Vorgangs abzuschließen: etwa 1,8 Millionen Wörter pro Minute. Die daraus resultierende Datei **BNC words.tok** ist 1,6 GB groß; **BNC words.typ** benötigt einen Speicherplatz von 26 MB.



Zu einem Index hinzufügen

Um neue Texte zu einem bereits bestehenden Index hinzuzufügen, wählen Sie Datei | Neu | Index. Der vorhandene Dateiname wird bereits für einen Index verwendet, also werden Sie gefragt, ob Sie neue Einträge hinzufügen wollen ('Ja') oder einen neuen Index erstellen wollten ('Nein').

Siehe auch: [Indexlisten verwenden](#), [Indexlisten anzeigen](#), Inhalt der WordList-Hilfe.

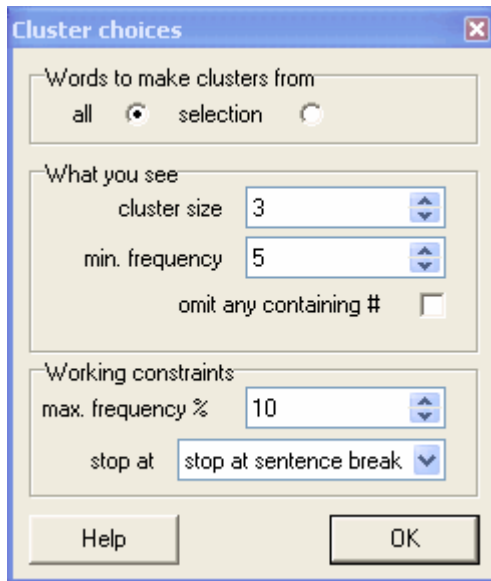
8.13 Indexcluster

WordList-Cluster

Eine Wortliste muss nicht nur aus einzelnen Wörtern bestehen. Sie können eine Wortliste erstellen mit Einträgen von zwei, drei, bis maximal acht Wörtern pro Zeile. Um in WordList die Cluster-Verarbeitung auszuführen, müssen Sie zuerst einen [Index erstellen](#).

Wie kann ich Cluster sehen?

[Öffnen](#) Sie den Index. Wählen Sie jetzt *Berechnen* | *Cluster*.



Wörter, mit denen man Cluster bilden kann

- "alle": alle Cluster, die Wörter über einer bestimmten Frequenz beinhalten (für einen großen Korpus wie den BNC World dauert das sehr lange), oder
- "Auswahl": Cluster nur für Wörter, die Sie ausgewählt haben (z. B. Sie haben BOOK und BOOKS markiert, und suchen nach Clustern wie **book a table, in my book**).

Um Wörter auszuwählen, die nicht nebeneinander stehen, drücken Sie Strg und klicken Sie die Nummer links an -- halten Sie Strg gedrückt und drücken Sie dann auf eine andere Nummer. Die erste angeklickte Nummer wird dann grün angezeigt und die anderen weiß. In dem Bild unten, basierend auf einem Index des BNC World Corpus, wurden die Wörter **world** und dann **life** durch Anklicken der Nummern 164 und 167 ausgewählt.

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	55,374	0.05	3,213	79.36

frequency alphabetical statistics filenames notes

343,617 Type-in LIFE

Dieser Vorgang nimmt Zeit in Anspruch. Im Fall des BNC World sind im Index die Positionen aller 100 Millionen Wörter gespeichert. Wenn man wie im oberen Beispiel nach 3-Wort-Clustern sucht, dauert es eine Minute, um alle 115.000 Fälle mit **world** und **life** zu verarbeiten und die 5.719 Cluster wie z. B. **the world bank** und **of real life** zu identifizieren. Chris Tribble berichtet, dass es mit seinem PC 36 Stunden dauerte, alle 3-Wort-Cluster des gesamten BNC zu berechnen. Er konnte seinen PC in dieser Zeit zwar anderweitig nutzen, dennoch möchte man so eine Berechnung nicht öfters durchführen.

Anzeige

Die "Clustergröße" muss zwischen zwei und acht Wörtern liegen.

Die "Mindestfrequenz" gibt an, wie häufig dieser Cluster mindestens vorkommen muss.

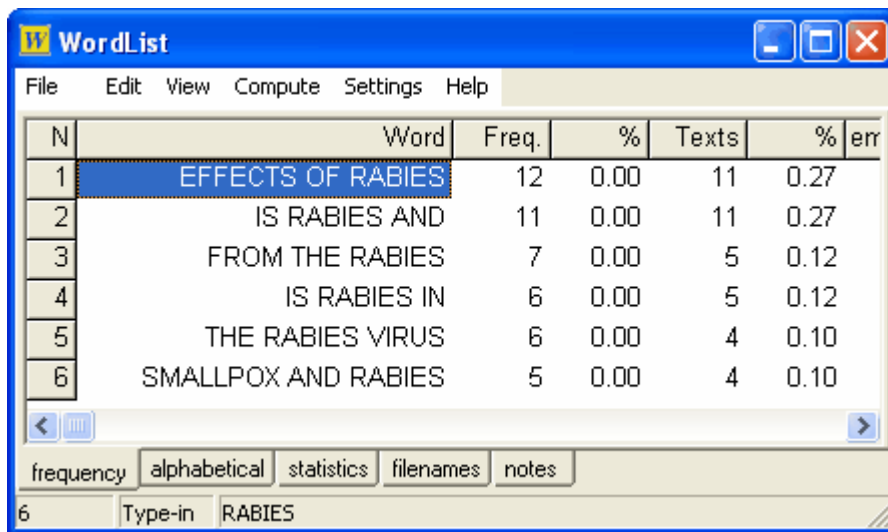
Hier hat der Benutzer 3-Wort-Cluster ausgewählt, die fünf Mal oder öfter vorkommen.

Einschränkungen

Die Einstellung "max. Frequenz %" beschleunigt den Vorgang. Dieser Wert steht für den maximalen Frequenzprozentwert, der bei der Berechnung von Clustern für ein gegebenes Wort noch berücksichtigt wird. Es gibt sehr viele Elemente, die äußerst häufig vorkommen, und Sie sind wahrscheinlich nicht an Clustern interessiert, die mit diesen Elementen *beginnen*. Zum Beispiel macht das Element **the** etwa 6% aller englischen Wortlisten aus, und Sie möchten sicherlich keine Cluster berechnen, die mit **the . . .** beginnen -- falls doch, können Sie den max. Prozentwert auf 0,5% oder 0,1% festlegen (wodurch im BNC World Corpus die 102 häufigsten Wörter ignoriert werden). Sie werden dennoch Cluster erhalten, die sehr häufige Elemente in der Mitte oder am Ende beinhalten, wie etwa **a** in der Phrase **book a table**, nicht jedoch **in my book**, das mit dem sehr häufigen Wort **in** beginnt. Je mehr Wörter Sie in der Berechnung mit einschließen, umso länger dauert dieser Vorgang.

Anhalten bei bietet, gleich wie [Concord: Cluster](#), eine Anzahl von Einschränkungen, wie z. B. Satzende bzw. andere Umbrüche, die durch Interpunktion markiert sind. Diese Einschränkungen dienen dazu Cluster auszuschließen, die über ein Satzende hinausgehen, da diese sehr wahrscheinlich keinen Sinn ergeben.

Wie sehen Cluster aus?

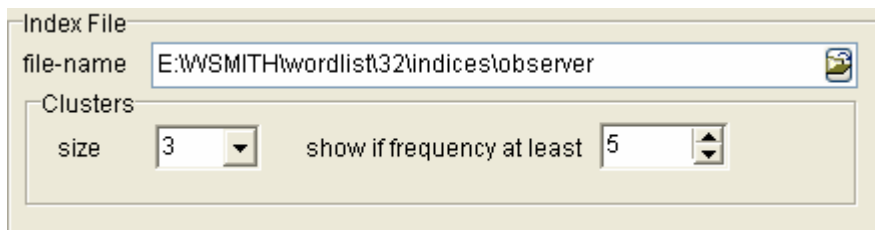


N	Word	Freq.	%	Texts	%	err
1	EFFECTS OF RABIES	12	0.00	11	0.27	
2	IS RABIES AND	11	0.00	11	0.27	
3	FROM THE RABIES	7	0.00	5	0.12	
4	IS RABIES IN	6	0.00	5	0.12	
5	THE RABIES VIRUS	6	0.00	4	0.10	
6	SMALLPOX AND RABIES	5	0.00	4	0.10	

frequency alphabetical statistics filenames notes

6 Type-in RABIES

Hier sehen Sie eine kleine Gruppe von 3-Wort-Clustern mit dem englischen Wort "rabies" (dt. 'Tollwut') aus dem BNC World Corpus. Einige davon sind sinnvolle Einheiten bestehend aus mehreren Wörtern. Alle Cluster werden angezeigt, die zumindest fünf Mal vorkommen: um diese Einstellung zu ändern, wählen Sie *Einstellungen anpassen* | *Index* im Controller und legen den Wert für "anzeigen ab Mindestfrequenz" wie folgt fest:



Index File

file-name E:\WSMITH\wordlist32\indices\observer

Clusters

size 3 show if frequency at least 5

Siehe auch: [Cluster in Concord](#)

8.14 Menüsuche

Über das Menü können Sie Zeichenfolgen innerhalb eines Eintrags suchen -- z. B. alle Wörter, die "vor" enthalten (durch die Eingabe ***vor*** -- die Sternchen bedeuten, dass das Element in der Mitte eines Worts auftreten kann, mit ***vor** wird *bevor* aber nicht *vorher* gefunden, während ***vor*** beide Wörter einschließt). Diese Suchen können wiederholt werden.

Mit dieser Funktion können Sie Wortteile finden und dadurch Ihre Wortliste bearbeiten, z. B. durch das Zusammenführen von zwei Wörtern zu einem Eintrag.

Sie können mit der Wildcard ***** nach Wortteilen (am Schluss oder in der Mitte) suchen.

Folglich ergibt ***ST*** Einträge wie *meistens*, *rasten*, etc.

***ST** ergibt Einträge wie *sonst*, *ist*, etc.

Sie können mit **F8** Ihre letzte Suche wiederholen.

Hotkeys für Suchfunktionen sind:

- F8** die letzte Suche wiederholen (in Kombination mit F10 oder F11 verwenden)
- F10** von der aktuellen Zeile abwärts suchen
- F11** von der aktuellen Zeile aufwärts suchen
- F12** Suche vom Anfang beginnen

Diese Funktion ist nützlich für die [Lemmatisierung](#) (Wörter zusammenführen, die zu einem

Eintrag gehören, z. B. *spielen/spiele/spielst/spielte/spielend*, etc.)
Siehe auch: [Eintrag durch Eingabe suchen](#)

8.15 Transinformationswerte

Wozu dient das?

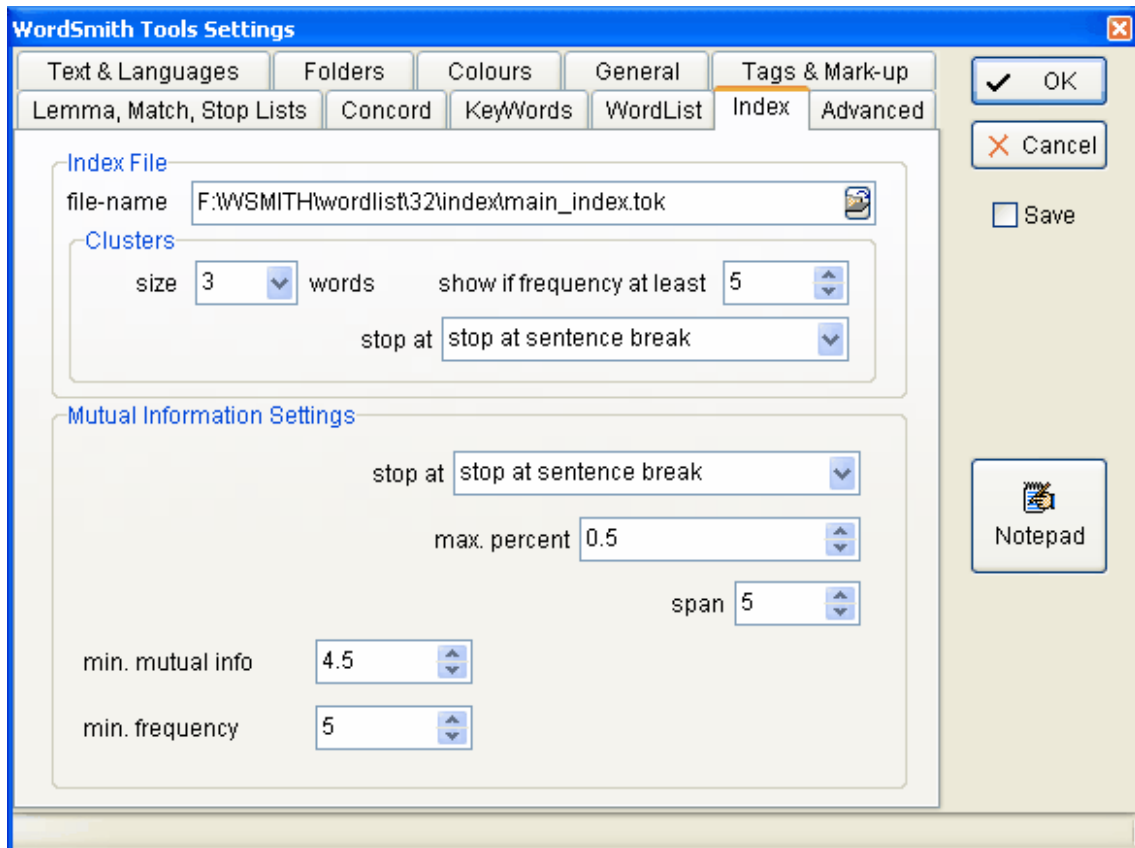
Bei der Transinformation (MI; steht für engl. "Mutual Information") wird die Beziehung zwischen zwei Wörtern berechnet. Zum Beispiel wenn das Wort *Problem* oft in Verbindung mit dem Wort *lösen* vorkommt, haben diese Wörter möglicherweise einen hohen Transinformationswert. Normalerweise scheint das Wort *das* sehr viel häufiger in der Nähe von *Problem* auf als *lösen*. Deshalb wird bei der Berechnung des Transinformationswerts nicht nur die Häufigkeit berücksichtigt, mit der ein Wort in der Nähe eines anderen auftritt, sondern auch wie häufig dieses Wort generell vorkommt. Da *das* sehr häufig unabhängig von *Problem* vorkommt, werden diese zwei Wörter nicht in Verbindung zueinander gesetzt und ihr MI-Wert ist niedrig.

Diese Beziehung ist bilateral: Im Fall von den englischen Wörtern *kith* und *kin*, wird nicht unterschieden zwischen der Tatsache, dass *kith* fast immer gemeinsam mit *kin* vorkommt, umgekehrt *kin* jedoch nicht nur mit gemeinsam mit dem Wort *kith* auftritt.

Es gibt viele verschiedene Formeln um zu berechnen, wie eng zwei Wörter einer Kollokation miteinander in Beziehung stehen. Der MI in WordSmith ("spezifische Transinformation") wird mit einer Formel berechnet, die von Gaussier, Lange und Meunier entwickelt wurde und in [Oakes](#), S. 174 beschrieben ist; hier basiert die Wahrscheinlichkeit auf der Gesamtanzahl der Token des Korpus. Andere Messungen dieser Beziehung werden auch berechnet; für weitere Informationen siehe [Transinformation: Anzeige](#).

Einstellungen

Die Einstellungen für Transinformation finden Sie im [Controller](#) unter *Einstellungen anpassen* | *Index* oder in einer Menüoption in **WordList**.



anhalten bei: Sie können den Kollokationsbereich selbst festlegen. Mit den Einstellungen in diesem Screenshot, würde in "**Ich schrieb den Brief. Kannst Du ihn abschicken?**" das Wort **abschicken** nicht als möglicher Kollokator von **Brief** gelten, weil zwischen diesen Wörtern ein Satzende ist.

max. Prozent jeden Token ignorieren, dessen Frequenz höher ist als der angegebene Prozentsatz. (Dadurch wird vermieden, dass die Transinformation für Wörter wie **der**, **von**, etc. berechnet wird, deren Frequenz wahrscheinlich über 1,0% liegt.)

Bereich: die Anzahl der Wörter zwischen Kollokator und Basis. Mit einem Bereich von 5, gelten für das Basiswort **schrieb** die Wörter **Ich**, **den**, **Brief**, **Kannst**, **Du** und **ihn** als mögliche Kollokatoren, wenn unter *anhalten bei* festgelegt ist *keine Grenzen*.

min. Transinformation: der Mindestwert des MI, ab dem das Ergebnis angezeigt wird. Ein sinnvoller Wert liegt bei 3,0. Bei einem Wert darunter ist es wahrscheinlich, dass die Beziehung zwischen Basis und Kollokator eher schwach ist.

Mindestfrequenz: die Mindestfrequenz für Elemente, deren Transinformation berechnet werden soll (Standard = 5). (Kommt ein Element nur ein- oder zweimal vor, ist die Transinformation wahrscheinlich nicht informativ.)

Siehe auch: [Transinformation: Anzeige](#), [Transinformation berechnen](#), [Indexliste erstellen](#), [Indexlisten anzeigen](#), Inhalt der WordList-Hilfe.

Siehe [Oakes](#) für weitere Informationen zu Transinformation.

8.16 Transinformation: Berechnen

In WordList oder in Concord

In Concord

Die Transinformation (MI für engl. "Mutual Information") wird nicht standardmäßig für eine Kollokatorenliste berechnet. Zur Berechnung des MI benötigen Sie eine Wortliste, um die relevanten Daten zu Verfügung zu stellen.

Angenommen Sie haben eine Konkordanz erstellt mit allen Dateien in `c:\wsmith4\text\shakespeare` und dem Suchwort *love*. Sie erhalten Kollokatoren wie *Romeo, hate, the, Juliet, Nurse*, etc. Alle diese Kollokatoren zeigen einen "Relationswert" (MI-Wert) von "??", da der Wert noch nicht berechnet wurde.

Wenn Sie den Wert noch nicht berechnet haben, erstellen Sie mit WordList eine Wortliste derselben Textdateien (oder eines anderen [Referenzkorpus](#), wenn Sie das bevorzugen). Stellen Sie sicher, dass Sie die Wortliste tatsächlich basierend auf dem [Referenzkorpus](#) erstellen möchten.

Wählen Sie nun das Menüelement  und Concord verwendet den Dateinamen des Referenzkorpus. Jeder der Kollokatoren wird nachgeschlagen und der MI-Wert anhand der Informationen im Referenzkorpus berechnet.

In WordList

Um die Transinformation (MI) zu berechnen, benötigen Sie einen [WordList-Index](#). Rufen Sie die alphabetische [Ansicht](#) der Liste auf.

Wenn Sie  anklicken, können Sie auswählen, ob Sie den MI-Wert für die ausgewählten (markierten) Einträge berechnen möchten, für alle Einträge, oder nur für Einträge zwischen zwei bestimmten Anfangsbuchstaben, z. B. von A bis D.

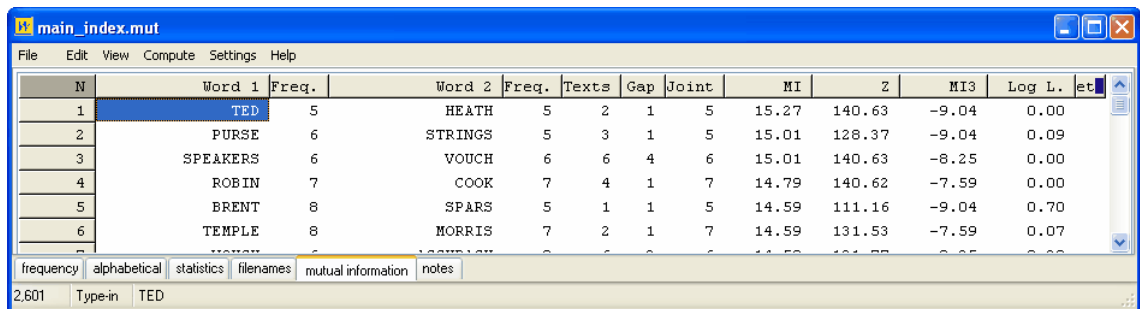
Wenn Sie nur ein paar Elemente für die MI-Berechnung auswählen möchten, können Sie diese zuerst markieren (mit ).

Den MI-Wert für jeden einzelnen Eintrag eines Index zu berechnen, dauert ziemlich lange: zumindest ein paar Stunden im Fall eines Index basierend auf 10 Millionen Wörtern. Es dauert etwa 24 Stunden, um den gesamten BNC, ca. 100 Millionen Wörter, zu verarbeiten, auch auf einem modernen, leistungsstarken PC. Vergessen Sie nicht, die Ergebnisse nach dem Vorgang zu speichern! Sie können immer Teile der Liste berechnen (z. B. A bis D) und später Ihre Transinformationsliste mit einer anderen Transinformationsliste (z. B. E bis H) [zusammenführen](#).

Siehe auch: [Kollokatoren](#), [Transinformationseinstellungen](#), [Transinformation: Anzeige](#), [Indexliste erstellen](#), [Indexlisten anzeigen](#), Inhalt der WordList-Hilfe.

8.17 Transinformation: Anzeige

Die Transinformationsanzeige beinhaltet mehrere Spalten:



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	

Wort 1: das linke Wort, gefolgt von Freq. (die Frequenz des Worts im gesamten Index).

Wort 2: das rechte Wort, gefolgt von Freq. (die Frequenz des Worts im gesamten Index).

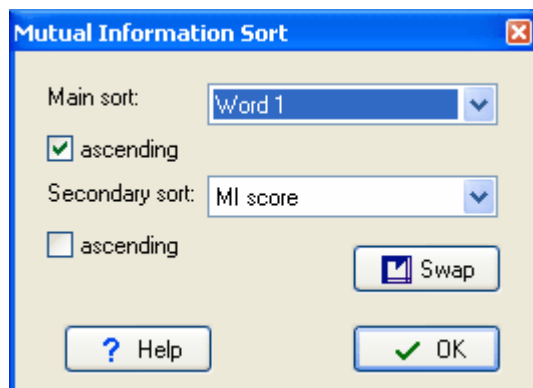
Texte: Die Anzahl der Texte, in denen dieses Wortpaar gefunden wurde (56 Mal im gesamten Index).

Spanne: die häufigste Spanne zwischen Wort 1 und Wort 2.

Gemeinsam: ihre gemeinsame Frequenz

In Zeile 2 in diesem Fenster erscheint das Wort PURSE sechsmal im gesamten Index, STRINGS fünfmal. Zusammen erscheinen sie fünfmal -- das bedeutet in diesem kleinen Korpus ist das Wort **strings** immer Teil des Phrase **purse strings**. Die Spanne ist 1, weil **strings** ein Wort nach **purse** kommt. Das Wortpaar **purse strings** kommt in drei Texten vor.

Wie immer können die Daten durch Anklicken der Spaltenüberschriften sortiert werden. In dem Beispiel oben ist "MI" die Primärsortierung und "Word 1" die Sekundärsortierung.



Sie können doppelt sortieren, mit Primär- und Sekundärsortierung, weil Sie möglicherweise sehen möchten, was mit der Liste passiert, wenn Sie die Liste nach MI-Wert, Z-Wert oder etwas anderem sortieren. Manchmal ist es auch notwendig, die Wörter alphabetisch zu belassen und nur innerhalb jedes Worts nach MI-Wert oder Z-Wert zu sortieren. Klicken Sie auf *Tauschen* um Primär- und Sekundärsortierung auszutauschen.

Vergleichen Sie dies mit der nach Z-Wert sortierten Anzeige ([Oakes](#), S. 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 (UK Premierminister in den 1970ern) immer noch an oberster Stelle und SPEAKERS ... VOUCH ist noch sichtbar, aber einige andere Wörter haben sich dazwischen geschoben.

Hier sehen Sie die nach MI3-Wert sortierte Anzeige ([Oakes](#), S. 172).

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

frequency alphabetical statistics filenames **mutual information** notes

2,601 Type-in TED

Wörter, die sehr viel häufiger vorkommen, finden sich nun an den oberen Positionen.

Zum Schluss noch eine Sortierung nach Log-Likelihood ([Dunning](#), 1993):

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

frequency alphabetical statistics filenames **mutual information** notes

2,601 Type-in THERE'LL

In diesem Fall sind Wörter in der Spalte "Word 2" sehr häufige Wörter, die zusammen mit den Wörtern in der Spalte "Word 1" grammatikalische Kollokationen bilden.

Siehe auch: [Transinformation](#), [Transinformation berechnen](#), [Indexliste erstellen](#), [Indexlisten anzeigen](#), Inhalt der WordList-Hilfe.

Vgl. [Oakes](#) für weitere Informationen zu Transinformation.

8.18 neu sortieren: Konsistenzliste

The frequency-ordered consistency display can be re-sorted by
alphabetical order (Word)

total frequencies overall (Total, the default)

by the *frequencies* in any given file (you see the file names).

Click on Word, Total or a filename to choose.

The sort can be either ascending or descending, the default being descending.

See also: [Sorting word-lists](#)

8.19 Statistik

Diese Statistik umfasst:

Anzahl der Dateien, auf denen die Wortliste basiert

Dateigröße (in Byte, d. h. Zeichen)

Anzahl der Wörter im Text (sog. *Tokens*)

Anzahl unterschiedlicher Wörter im Text (sog. *Types*)

[Type-Token-Relationen](#)

Anzahl der [Sätze](#) im Text

durchschnittliche Satzlänge (in Wörtern)

Standardabweichung von der Satzlänge (in Wörtern)

Anzahl der [Absätze](#) im Text

durchschnittliche Absatzlänge (in Wörtern)

Standardabweichung von der Absatzlänge (in Wörtern)

Anzahl der [Überschriften](#) im Text

durchschnittliche Länge der Überschriften (in Wörtern)

Anzahl der [Abschnitte](#) im Text

durchschnittliche Länge der Abschnitte (in Wörtern)

Standardabweichung von der Länge der Überschriften (in Wörtern)

Anzahl der Wörter, die nur aus einem Buchstaben bestehen

...

Anzahl der Wörter mit **n** Buchstaben (um diese anzusehen, scrollen Sie in dem Listenfeld nach unten)

(14 ist die [Standardeinstellung](#) für maximale Wortlänge) Sie können diesen Wert in den *WordList-Einstellungen* im Menü *Einstellungen* auf bis zu 50 Buchstaben festlegen. Längere Wörter werden abgekürzt und mit einem + am Ende des Worts gekennzeichnet.

Die Anzahl der Types (unterschiedliche Wörter) wird für jeden Text einzeln berechnet. Wenn Sie also eine Wortliste aus mehreren Texten erstellt haben, wird die Summe der Types der einzelnen Texte nicht der Anzahl der Types der gesamten Textsammlung entsprechen.

Siehe auch: [WordList-Anzeige](#) (mit einem Screenshot), [Statistikzusammenfassung](#), [Anfang und Ende von Textsegmenten](#).

8.20 Wörter aus Textliste importieren

Wozu dient das?

Sie möchten möglicherweise eine Wortliste basierend auf einigen Daten erstellen, Sie verfügen jedoch nur über die Daten in Form einer Liste und nicht über die Originaltexte.

Anforderungen

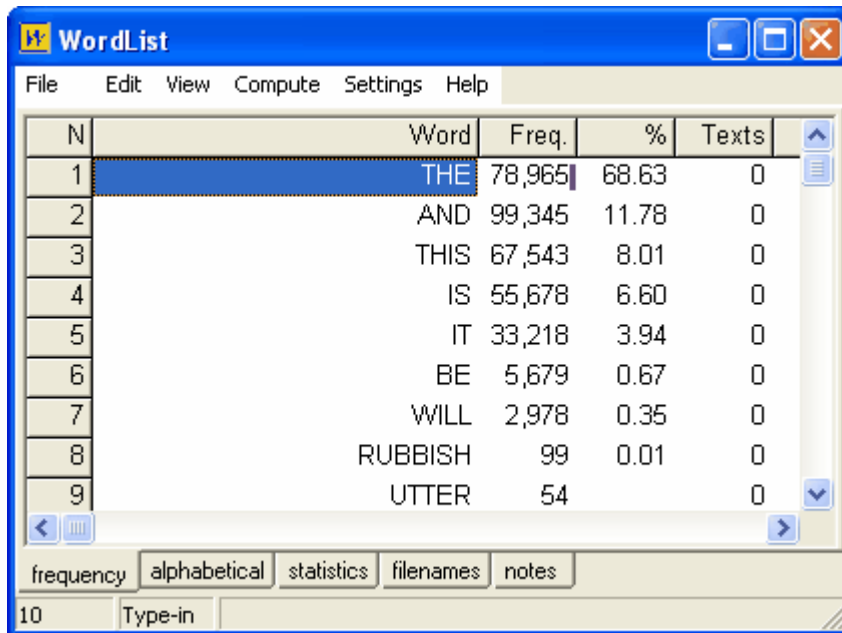
Ihre Datei kann in einer beliebigen [Sprache](#) verfasst sein (wählen Sie diese aus, bevor Sie die Liste erstellen) und in [Unicode](#) oder [ASCII](#) kodiert sein.

Sie muss jedoch in einem ähnlichen Format wie eine [Stoppliste](#) sein, außer dass nach jedem Wort ein Tabulator und die Frequenz als Ganzzahl stehen müssen (Dezimalstellen werden ignoriert). Verwenden Sie keine Punkte oder Kommas als 1000er-Trennzeichen, da diese dann als verschiedene Wörter interpretiert werden. Die Wörter müssen weder nach Frequenz noch alphabetisch geordnet sein.

Beispiel

```
; Meine Wortliste für Testzwecke.
THIS 67543
IT 33218
WILL 2978
BE 5679
COMPLETE 45
AND 99345
UTTER 54
RUBBISH 99
THE 578965
IS 55678
```

Sie sollen solche Ergebnisse erhalten:



The screenshot shows the WordList application window. It has a menu bar with File, Edit, View, Compute, Settings, and Help. Below the menu is a table with the following data:

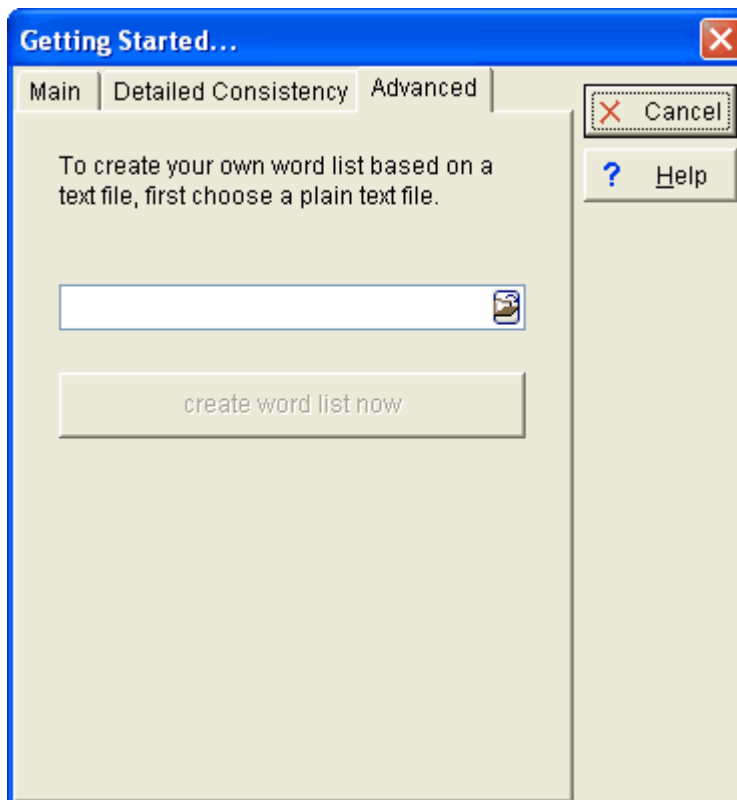
N	Word	Freq.	%	Texts
1	THE	78,965	68.63	0
2	AND	99,345	11.78	0
3	THIS	67,543	8.01	0
4	IS	55,678	6.60	0
5	IT	33,218	3.94	0
6	BE	5,679	0.67	0
7	WILL	2,978	0.35	0
8	RUBBISH	99	0.01	0
9	UTTER	54		0

At the bottom of the window, there are tabs for frequency, alphabetical, statistics, filenames, and notes. The frequency tab is selected. Below the tabs is a text input field with the number 10 and the label 'Type-in'.

Statistik wird auf die einfachste mögliche Art berechnet: die Wortlängen (plus Durchschnitt und Standardabweichung) und die Anzahl der Types und Tokens. Da die meisten Verfahren die Gesamtanzahl der Wörter (Tokens) und die Anzahl der verschiedenen Wörter (Types) benötigen, sollten Sie diese Wortliste in KeyWords, etc. verwenden können.

Wie mache ich das?

Wenn Sie die Menüoption *Neu* () in WordList wählen, wird ein Fenster mit drei Registerkarten geöffnet: eine *Haupt*-Registerkarte für die gängigsten Zwecke,



eine für [Detaillierte Konsistenz](#) und eine weitere (*Erweitert*) für die Erstellung einer Wortliste anhand einer Standardtext-Datei.

Wählen Sie Ihre .txt-Datei und klicken Sie auf *Wortliste jetzt erstellen*.

8.21 Type-Token-Relationen

Wenn ein Text aus 1.000 Wörtern besteht, spricht man von 1.000 "Token". Viele dieser Wörter kommen jedoch öfter vor, also enthält der Text vielleicht nur 400 unterschiedliche Wörter. Diese unterschiedlichen Wörter werden "Types" genannt.

Die Relation zwischen Types und Token wäre in diesem Beispiel 40%.

Aber diese Type-Token-Relation (TTR) kann sehr unterschiedlich sein, abhängig von der Länge des Texts (oder der Größe des Textkorpus), der untersucht wird. Ein Artikel mit 1.000 Wörtern hat eine TTR von 40%; ein kürzerer Artikel erreicht vielleicht 70%; ein Text mit 4 Millionen Wörtern ergibt wahrscheinlich eine Type-Token-Relation von etwa 2%, und so weiter. Diese Type-Token-Information ist in den meisten Fällen bedeutungslos, obwohl sie bei einer WordList-Statistik angezeigt wird. Die konventionelle TTR ist natürlich informativ, wenn Sie mit einem Korpus arbeiten, der aus vielen gleich großen Textsegmenten besteht (z. B. die Korpora LOB und Brown). Wenn Sie allerdings mit Texten arbeiten, die im Alltag auftauchen (speziell wenn Ihr Forschungsschwerpunkt auf dem Text im Gegensatz zur Sprache liegt), dann sind diese Text bezüglich ihrer Länge normalerweise sehr unterschiedlich, und dann ist eine konventionelle TTR nicht besonders hilfreich.

WordList nutzt aus diesem Grund eine andere Methode zur Berechnung dieser Relation. Die standardisierte Type-Token-Relation (STTR) wird immer nach n Wörtern berechnet, während WordList die Dateien verarbeitet. Der [Standard](#) liegt bei $n = 1.000$. Die Relation wird also für die ersten 1.000 Wörter berechnet, dann erneut für die darauffolgenden 1.000 Wörter, und so weiter, bis das Ende des Textkorpus erreicht ist. Ein laufender Durchschnitt wird berechnet, was

bedeutet, dass Sie eine durchschnittliche Type-Token-Relation auf Basis von aufeinanderfolgenden Textchunks zu jeweils 1.000 Wörter erhalten. (Texte mit weniger als 1.000 Wörtern (bzw. dem für n festgelegten Wert) werden eine standardisierte Type-Token-Relation von 0 ergeben.)

N-Wert festlegen

Legen Sie den Wert für n in den [Einstellungen für Minimum und Maximum](#) auf eine beliebige Zahl zwischen 100 und 20.000 fest.

Was wird bei der STTR tatsächlich gezählt?

Anmerkung: Bei der Berechnung der Relation werden a) alle [unterschiedlichen Formen](#) als ein Wort gezählt (also **sagen** und **sage** sind zwei Types); b) nur die Wörter verwendet, die nicht in der [Stoppliste](#) sind; c) nur die Wörter verwendet, die innerhalb der von Ihnen festgelegten Wortlänge liegen; d) Ihre Einstellungen bezüglich [Zahlen](#) und [Bindestrichen](#) berücksichtigt. Die angezeigte Zahl ist ein Prozentwert der neuen Types pro n Token. Auf diese Weise können Sie Type-Token-Relationen bei Texten mit unterschiedlicher Textlänge vergleichen. Diese Methode steht der Methode von [Tuldava](#) (1995:131-50) gegenüber, der sich auf einen Wert aus einer dreistufigen Akkumulierung verlässt. Die Methode, mit der WordSmith die STTR berechnet, wurde von mir selbst entwickelt, ist jedoch ähnlich einer Methode, die von dem Mathematiker [David Malvern](#) in Zusammenarbeit mit Brian Richards (University of Reading) entwickelt wurde.

Weitere Erklärung

TTR und STTR sind jeweils ziemlich grobe Messungen, obwohl oft angenommen wird, dass dadurch "lexikalische Dichte" ausgedrückt werden kann. Angenommen Sie haben einen Text, der 1.000 Wörter lang von **ELEFANT, LÖWE, TIGER**, etc. handelt; die nächsten 1.000 Wörter von **MADONNA, ELVIS**, etc. und wiederum die nächsten 1.000 Wörter von **WOLKE, REGEN, SONNENSCHNEIEN**. Wenn der n-Wert bei STTR auf 1.000 festgelegt wird, und Sie einen STTR-Wert von 48% erhalten, enthält die Statistik keine Informationen darüber, dass im Text Veränderungen von Afrika zu Musik und schließlich zum Wetter stattgefunden haben. Liegt die Textgrenze (n-Wert) zwischen Afrika und Musik bei 650 Wörtern statt 1.000, schlägt sich dieser Unterschied nicht in der Statistik nieder. *Wodurch* würde dann ein Unterschied entstehen? In einem Text, der von Wolken handelt und von einer Person geschrieben wurde, die zwischen verschiedenen Wolkenarten unterscheidet, würde auch Wörter vorkommen wie **NEBEL, DUNST, KUMULUS, KUMULONIMBUS**. Der STTR-Wert wäre also höher als bei einem Text von einem Kind, das häufig das Wort **WOLKE** verwendet, und dieses mit Adjektiven wie **HOCH, NIEDRIG, SCHWER, DUNKEL, DÜNN, SEHR DÜNN** beschreiben würde... und dazu die Adjektive **DUNKEL, DÜNN**, etc. häufig wiederholt...

(Anm.: Shakespeare ist dafür bekannt, dass er, wenn man von diesen Messungen ausgeht, einen ziemlich begrenzten Wortschatz verwendet hat!)

8.22 Groß- und Kleinschreibung beachten

Normalerweise erstellen Sie eine Wortliste, bei der Groß- und Kleinschreibung nicht beachtet wird, weil die Großschreibung in den meisten Sprachen nicht nur die Eigennamen, sondern auch den Satzanfänge, Überschriften, Substantive, etc. markiert. Beim Erstellen einer Wortliste, die zwischen haupt, Haupt und HAUPT unterscheidet, aktivieren Sie "Groß- und Kleinschreibung beachten" (*Einstellungen anpassen* | *WordList* | *Groß- und Kleinschreibung beachten* im WordSmith [Controller](#)).

In Ihrer Liste werden zuerst wahrscheinlich alle Einträge in GROSSBUCHSTABEN erscheinen. Drücken Sie **Strg+L** oder wählen Sie die Menüoption [Layout](#) () , um diese Einstellung zu verändern.

8.23 Einstellungen für Minimum und Maximum

Diese umfassen:

Mindestwortlänge

Standard: 1 Buchstabe. Beim Erstellen einer Wortliste können Sie die Mindestwortlänge festlegen, z. B. um alle Wörter mit weniger als 3 Buchstaben auszuschließen.

Maximale Wortlänge

Standard: 49 Buchstaben. Sie können Wörter mit bis zu 50 Zeichen zulassen. Wenn ein Wort diese Länge überschreitet und die Option "Abkürzen mit +" aktiviert ist, wird das Wort auf die maximale Zeichenanzahl gekürzt und WordList fügt am Ende des gekürzten Worts das Symbol + hinzu. Wenn Sie die Option "Abkürzen mit +" nicht aktiviert haben, wird das Wort nicht in Ihre Wortliste aufgenommen. Wenn Sie eine Wortliste mit Wörtern erstellen möchten, die alle aus vier Buchstaben bestehen, können Sie Mindest- und maximale Länge auf 4 festlegen und die Option "Abkürzen mit +" nicht aktivieren.

Mindestfrequenz

Standard: 1. Standardmäßig werden alle Wörter gespeichert, auch diejenigen, die nur einmal vorkommen. Wenn Sie nur nach häufigen Wörtern suchen, können Sie irgendeine Zahl bis 32.000 wählen.

maximale Frequenz

Das Standardmaximum liegt bei 2.147.483.647 (2 Gigabyte). Um ein Wort zu finden, das so häufig vorkommt, muss viel Text analysiert werden! Sie können die maximale Frequenz auf 500 und die Mindestfrequenz auf 50 festlegen: so erstellen Sie eine Wortliste, die nur gebräuchliche Wörter enthält.

Type-Token-Mittelwert (Standard 1.000)

Ermöglicht eine reibungslose Berechnung der Type-Token-Relation für Wortlisten. Wählen Sie eine Zahl zwischen 10 und 20.000. Für weitere Informationen siehe [WordList Type-Token-Relationen](#).

Siehe auch: [Texteigenschaften](#), [Stopplisten](#), [Standardeinstellungen vornehmen](#)

8.24 Sortierreihenfolge

Wie mache ich das?

Sie können eine Liste einfach sortieren, indem Sie auf die oberste Zeile klicken. Sortieren Sie erneut, um von aufsteigender Sortierung zu absteigender Sortierung zu wechseln.

Mit einer geöffneten Wortliste kann das Frequenzfenster nicht sortiert werden, aber Sie können das alphabetische Fenster auf viele verschiedene Arten neu sortieren (wählen Sie in den Registerkarten am unteren Rand von WordList die Registerkarte aus).

Um einer der speziellen Sortierungen auszuwählen, die unten beschrieben werden, drücken Sie F6 oder Strg+F6 oder Umschalt+Strg+F6. Oder wählen Sie die entsprechende Menüoption.

Alphabetische Wortsortierung

Viele Sprachen haben ihre eigene spezielle Sortierreihenfolge. Sie sollten also vor dem Sortieren überprüfen, ob die ausgewählte [Sprache](#) mit den zu sortierenden Wörtern übereinstimmt. Spanisch beispielsweise wird folgendermaßen sortiert: **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**.

Wortsortierung umkehren

Mit dieser Funktion können Sie Wörter nach Suffixen sortieren. Die Sortierung wird von den

Wortenden bestimmt, nicht den Wortanfängen. So finden Sie z. B. alle Wörter mit der Endung *-ung* direkt untereinander.

Wortlängensortierung

Mit dieser Funktion können Sie Wörter nach Länge sortieren (1 Buchstabe, 2 Buchstaben, etc. bis zu Wörter mit 50 Buchstaben). Innerhalb einer Gruppe gleich langer Wörter ist die Sekundärsortierung alphabetisch.

Konsistenzsortierung

Klicken Sie die Spaltenüberschrift "Texte" an, um Wörter entsprechend der [Konsistenz](#) zu sortieren.

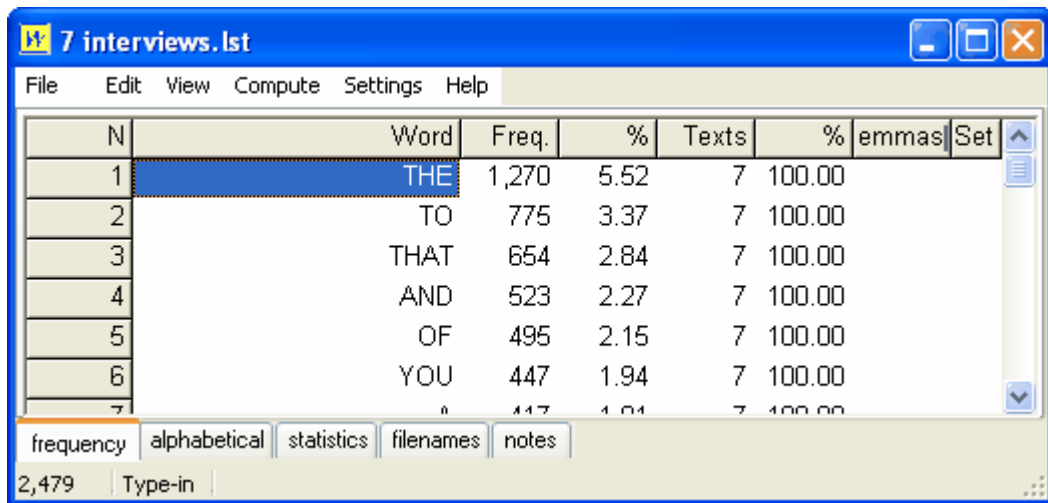
Siehe auch: [Concord-Sortierung](#), [KeyWords-Sortierung](#), [Einträge bearbeiten](#); [Akzentzeichen](#); [Sprachen auswählen](#)

8.25 WordList-Anzeige

Each WordList display shows

- the word
- its frequency
- its frequency as a percent of the running words in the text(s) the word list was made from
- the number of texts each word appeared in
- that number as a percentage of the whole corpus of texts

The Frequency display might look like this:

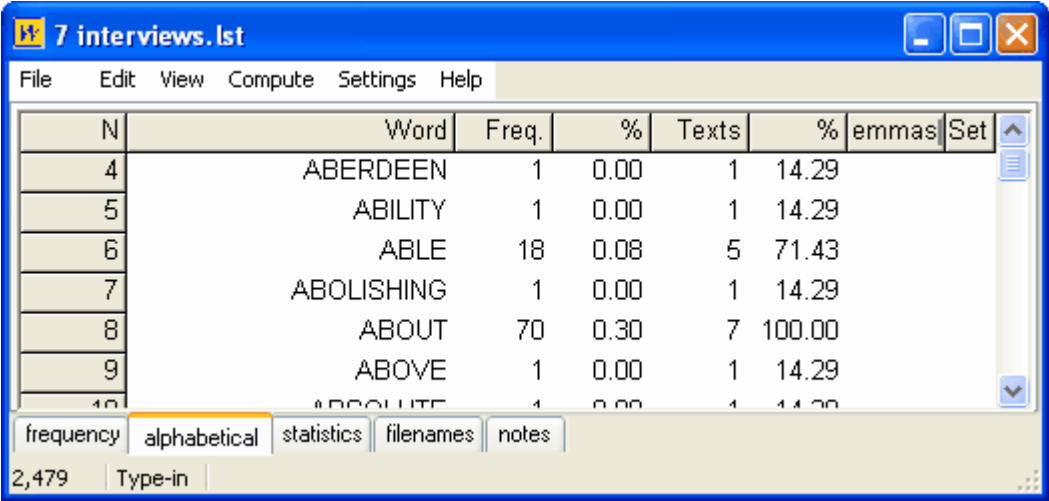


N	Word	Freq.	%	Texts	%	emmas	Set
1	THE	1,270	5.52	7	100.00		
2	TO	775	3.37	7	100.00		
3	THAT	654	2.84	7	100.00		
4	AND	523	2.27	7	100.00		
5	OF	495	2.15	7	100.00		
6	YOU	447	1.94	7	100.00		
7	a	417	1.81	7	100.00		

frequency alphabetical statistics filenames notes

2,479 Type-in

Here you see the top 6 words in a word list based on 7 interviews. There are 2,479 words altogether but in the screenshot we can only see the first few. The Freq. column shows how often each word cropped up (**THE** appeared 1,270 times in the 7 texts), and the % column tells us that 1,270 represents 5.52% of the running words in the 7 texts. The Texts column shows that **THE** comes in 7 texts, that is 100% of the texts used for the word list.



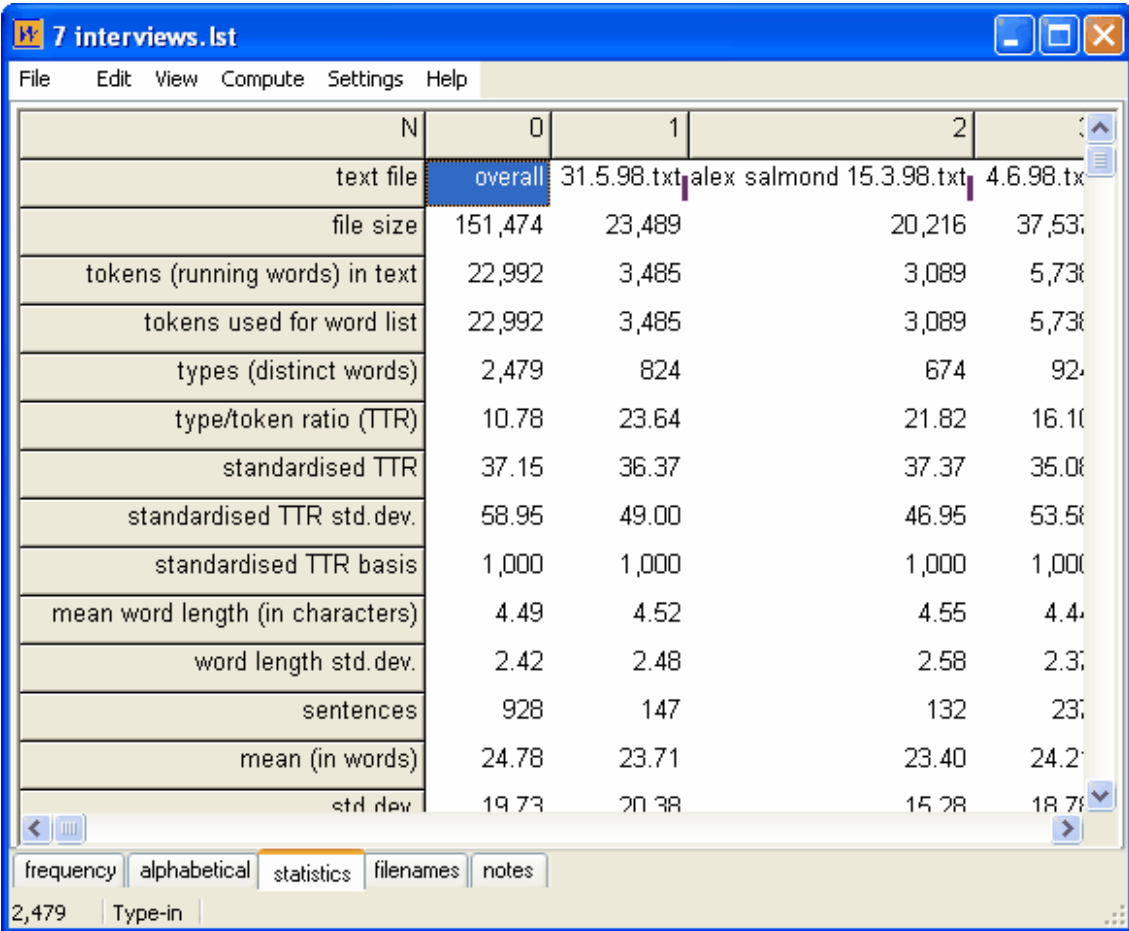
N	Word	Freq.	%	Texts	%	emmas	Set
4	ABERDEEN	1	0.00	1	14.29		
5	ABILITY	1	0.00	1	14.29		
6	ABLE	18	0.08	5	71.43		
7	ABOLISHING	1	0.00	1	14.29		
8	ABOUT	70	0.30	7	100.00		
9	ABOVE	1	0.00	1	14.29		
10	ABOLISHING	1	0.00	1	14.29		

frequency alphabetical statistics filenames notes

2,479 Type-in

The Alphabetical listing also shows us some of the words but now they're in alphabetical order. **ABLE** comes 18 times altogether, and in 5 of the 7 texts. **ABOUT**, on the other hand, comes in all 7 texts.

Now let's examine the statistics.



N	0	1	2	3
text file	overall	31.5.98.txt	alex salmond	15.3.98.txt
file size	151,474	23,489	20,216	37,536
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	2.36
sentences	928	147	132	236
mean (in words)	24.78	23.71	23.40	24.21
std dev	19.73	20.38	15.28	18.79

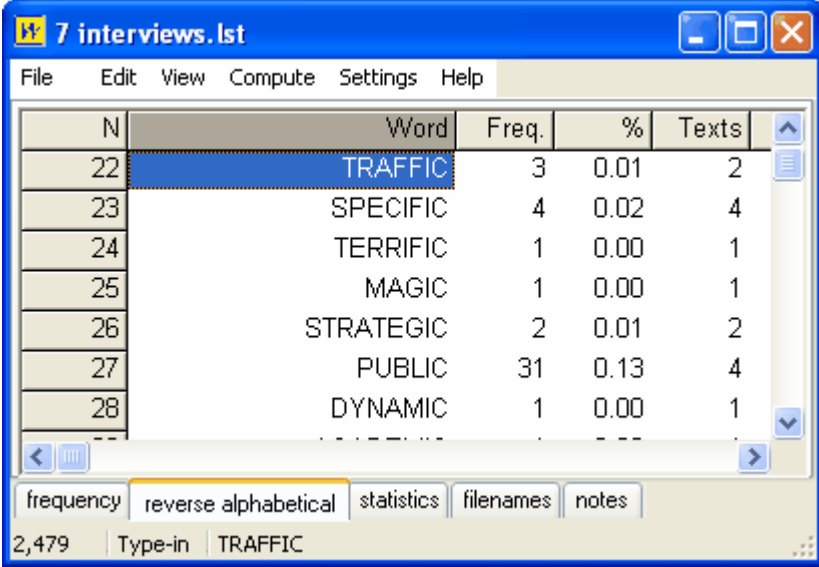
frequency alphabetical statistics filenames notes

2,479 Type-in

In all 7 texts, there are 2,749 word types (as pointed out above). The total running words is

22,992. Each word is about 4.49 characters in length. There are 928 sentences altogether, on average 24.78 words in length. In the text of the interview with Alex Salmond, there are only 674 different word types and that interview is only just over 3,000 words in length. This is explained in more detail in the [Statistics](#) page.

Finally, here is a screenshot of the same word list sorted "reverse alphabetically". In the part which we can see, all the words end in **-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

frequency reverse alphabetical statistics filenames notes

2,479 Type-in TRAFFIC

To do a reverse alphabetical sort, I had the Alphabetical window visible, then chose *Edit | Reverse Word sort* in the menu. To revert to an ordinary alphabetical sort, press F6.

See also : [Consistency](#), [Lematisation](#)

8.26 WordSmith Controller: WordList-Einstellungen

These are found in the main [Controller](#) under *Adjust Settings | WordList*.

This is because some of the choices -- e.g. [Minimum & Maximum Settings](#) -- may affect other Tools.

There are 2 sets : What you Get and What you See.

WHAT YOU GET

Word Length & Frequencies

See [Minimum & Maximum Settings](#).

Standardised Type/Token

See [WordList Type/Token Information](#).

Tags

If you want to include tags in a word list, you need to set up a [Tag File](#) first. Then ensure that "Include Tags" is checked, so that you will see the tags in the word list. Tags are not counted in any statistics based on a running word count or number of tokens or types. What you will see for each is its frequency, that frequency as a percentage of the running words excluding tags, and the number of texts it is in.

WHAT YOU SEE

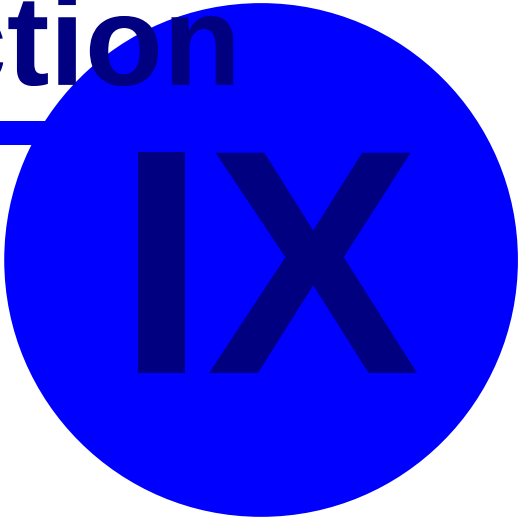
Case Sensitivity

Normally, you'll make a case-insensitive word list. If you wish to make a word list which distinguishes between the, The and THE, activate [case sensitivity](#).

See also: [Using Index Lists](#), [Viewing Index Lists](#), WordList Help Contents, [Computing word list clusters](#).

Viewer und Aligner

Section



IX

9 Viewer und Aligner

9.1 Zweck

Dies ist ein Programm zum Anzeigen von Texten und anderen Dateien, in dem interessante Wörter markiert werden. Standardmäßig wird ein Text im [Standardtext](#)-Format angezeigt, aber [SGML- oder HTML](#)-Text kann auf viele verschiedene Arten angezeigt werden. Weiters können Akzentzeichen als numerische Codes angezeigt werden, z. B. um die Codes für [Akzentzeichen](#) zu untersuchen. Es gibt zahlreiche [Einstellungen](#) und [Optionen](#), die Sie verändern können.

Viewer kann dazu verwendet werden, Texte zu "skimmen", d. h. Texte mit einer festgelegten Geschwindigkeit durchzulesen.

Sie können Viewer auch dazu verwenden, Ihre Textdatei zu [bearbeiten](#) (z. B. wenn Symbole wie £ zum Windows-Zeichensatz konvertiert wurden) und Sätze oder Absätze zu nummerieren, indem Sie den Text [als Textdatei speichern](#).

Außerdem dient Viewer dazu, durch [Alignment](#) eine synchronisierte Version zweier Texte zu erstellen, wobei abwechselnd Sätze oder Absätze von beiden Texten in eine Datei gespeichert werden.

Viewer ist ein eigenes Utility, das mit den anderen Tools verknüpft ist: Viewer wird aufgerufen, wenn Sie zum Quelltext zurückkehren müssen, z. B. um [Schlüsselwörter](#) im Kontext anzuzeigen. Siehe auch: [Viewer-Einstellungen](#), [Viewer-Optionen](#)

9.2 Alignment mit Viewer

Diese Funktion synchronisiert die Sätze zweier Dateien. Übersetzer müssen die Unterschiede zwischen Original und Übersetzung untersuchen. Andere Linguisten möchten möglicherweise die Unterschiede zwischen zwei Versionen desselben Texts in einer Sprache untersuchen.

Menschen, die [verschiedene Sprachen](#) studieren, können dieses Werkzeug für zweisprachiges Lesen verwenden, um so die Unterschiede z. B. im Satzbau festzustellen.

Es hilft Ihnen dabei, einen neuen Text zu erstellen, der aus den synchronisierten Sätzen zweier Dateien besteht. Auf diese Weise können Sie die Übersetzung mit dem Original vergleichen.

Beispiel

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, ...* (aus *Stifters Bergkristall*, übersetzt von Harry Steinbauer, in *German Stories - Deutsche Novellen*, Bantam Books 1961)

Übersetzung: *The boy communicated this thought to his sister and she followed him. But the road down the neck could not be found either. Though the sun shone clearly, ...*

Text nach Alignment:

<G1> Der Knabe sagte diesen Gedanken dem Schwesterchen, und diese folgte.

<E1> The boy communicated this thought to his sister and she followed him.

<G2> Allein auch der Weg auf den Hals hinab war nicht zu finden.

<E2> But the road down the neck could not be found either.

<G3> So klar die Sonne schien, ...

<E3> Though the sun shone clearly, ...

Ein synchronisierter Text wie dieser hilft dabei, Einfügungen und Auslassungen, Normalisierungen, Stilwechsel oder bevorzugten Satzbau zu erkennen. In diesem Fall hat sich der Übersetzer dafür entschieden, starke Ähnlichkeiten zu vermeiden.

Siehe auch: [Schaltflächen](#)

9.3 Alignment und Verschieben

You will see one text in one colour and the other in another. (Colours can be changed in the *Settings* menu.)

You may well want to alter sentence ordering. The translator may have used three sentences where the original had only one.

adjusting by dragging with the mouse

To alter sentence order, just drag a sentence to the next one above of the same colour.

Finally you will want to [save \(F2\) the results](#).

See also: Viewer contents

9.4 bearbeiten

Obwohl **Viewer** kein vollwertiges Textverarbeitungsprogramm ist, wurden einige Funktionen zum Bearbeiten von Texten integriert, um damit häufige Formatierungsprobleme zu beheben.

Bearbeiten (): öffnet ein Fenster, in dem Sie den gesamten aktuellen Satz bzw. Absatz bearbeiten können.

Überflüssige Leerstellen entfernen: überprüft jeden Satz des Texts und entfernt alle überflüssigen Leerstellen -- wenn zwei oder mehr Leerstellen direkt hintereinander stehen, werden diese auf eine Leerstelle reduziert.

Zeilen suchen, die mit einem Kleinbuchstaben beginnen: findet Sätze oder Absätze, die nicht mit einem Großbuchstaben oder einer Zahl beginnen. Diese möchten Sie wahrscheinlich mit dem vorangehenden Satz oder Absatz [verbinden](#). Dieses Problem tritt häufig auf, wenn der Text als "Nur Text mit Zeilenumbrüchen" gespeichert wurde (wo sich am Ende jeder Zeile ein harter Zeilenumbruch befindet, unabhängig davon, ob der Absatz dort endet.)

[Kurze Zeilen suchen](#)

[Tags einfügen](#)

Danach sollten Sie Ihren Text speichern (F2).

Sie können auch:

- eine neue Datei zur Ansicht öffnen (Sie können bis zu 10 Textdateien innerhalb von **Viewer** öffnen)

- eine Textdatei in die [Zwischenablage](#) kopieren (auswählen, danach Strg+Einfüg drücken)

- die gesamte aktuelle Textdatei oder einen Teil davon drucken

- den aktuellen [Zeichensatz](#) ändern

- nach Wörtern oder Phrasen suchen (F12 drücken)

- skimmen** (automatisch durch Bildschirmseiten des Texts "blättern". Die veränderte Form des Cursors zeigt an, dass Sie sich im Skim-Modus befinden. Die Geschwindigkeit können Sie anhand einer kleinen horizontalen Bildlaufleiste verändern, und Sie können den Vorgang jederzeit durch Klicken auf den Text oder Drücken einer beliebigen Taste beenden.)

9.5 Sätze und Absätze nummerieren

You can use the **Viewer** to make a copy of your text with all the sentences and/or paragraphs tagged with **<S>** and **<P>**.

To do this, simply read in the text file in, choose *Edit | Insert Tags*, then [save it as a text file](#).

To choose between sentence or paragraph mode, see [Settings](#).

See also: Viewer contents

9.6 Optionen

Modus: Satz/Absatz

Dieser Modus wechselt zwischen Satzmodus (Standard) und Absatzmodus. Sie können also auswählen, ob bei Ihren Textdateien pro Zeile ein Satz oder ein Absatz angezeigt werden soll. Sie können auch mit Alignment einen Text nach Sätzen oder Absätzen synchronisieren. Die anderen Funktionen (z. B. [verbinden](#), [trennen](#)) sind nicht von diesem Modus abhängig.

Wenn Sie den Text erneut speichern, bevorzugen Sie wahrscheinlich den Absatzmodus; hier wird am Ende jedes Absatzes ein harter Zeilenumbruch eingefügt. Im Satzmodus wird am Ende jedes Satzes ein harter Zeilenumbruch eingefügt, wodurch der Text dann aus einer langen Liste von Sätzen besteht.

Anzeige: Normal/Akzentzeichen, etc. als Codes

in normalen Textformat, oder mit Akzentzeichen als Codes anzeigen, z. B. für **café au lait** wird **caf<233> au lait** angezeigt.

Tags ignorieren: einblenden/ausblenden

Ihre [zu ignorierenden Tags](#) können eingeblendet oder ausgeblendet werden.

Format: Standardtext/HTML/SGML

Die Standardeinstellung ist [Standardtext](#), aber wenn Ihre Texte in [HTML, XML oder SGML](#) formatiert sind, können Sie diese Einstellung ändern, und alle [Tags](#) werden in der linken Spalte angezeigt.

Header: einblenden/ausblenden

Wenn Ihr Text nicht nur [Tags](#), sondern auch einen Header enthält, kann dieser auch eingeblendet oder ausgeblendet werden.

9.7 Sätze verbinden und trennen

Joining

The easiest way to join two sentences is simply to drag the one you want to move onto its neighbour above. Or you can mark them with , then press . You will be told if you're in danger of jumbling up your text!

Splitting in two

To split a sentence, press . You will get a list of the words. Click on the word which should **end** the sentence, then press OK.

example

"It
is
good!"
Mary
wanted
more.

This will insert the following words (*Mary wanted more.*) into a new line below.

See also: Viewer contents

9.8 Einstellungen

Diese Einstellungen sind übliche Einstellungen, die Sie in den meisten Tools finden:

[Farben](#)

[Schriftart](#)

[Drucken](#)

[Texteigenschaften](#)

[Alle Einstellungen überprüfen](#)

9.9 technische Aspekte

Wann wird ein Satz nicht als Satz erkannt?

Es gibt keine perfekte mechanische Methode, Satzgrenzen zu bestimmen. Zum Beispiel, eine Überschrift endet oft nicht mit einem Punkt, wird aber normalerweise nicht als Teil des darauffolgenden Satzes betrachtet. Auch ein Satz endet oft nicht mit einem Punkt, wenn danach eine Liste von Elementen folgt.

Der von **Viewer** verwendete Algorithmus ist: Ein Satz endet bei einem Punkt, Frage- oder Ausrufezeichen (?!), direkt gefolgt von einem oder mehreren [Worttrennzeichen](#) und danach einem Großbuchstaben (A-Z; auch Akzentbuchstaben), einer Ziffer oder einem Währungszeichen. Dasselbe Verfahren wird auch in **WordList** angewandt, obwohl **WordList** versucht, zwischen Sätzen und Überschriften zu unterscheiden, wodurch sich oft eine unterschiedlich hohe Anzahl der Sätze in den beiden Tools ergibt.

Hier ein Beispiel aus *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 betrachtet fälschlicherweise - Joe! als eigenen Satz, jedoch *"Wo-ho!" said the coachman.* nicht als zwei getrennte Sätze: Das Programm würde zwei Sätze erkennen, wenn das Wort nach *ho!* mit einem Großbuchstaben beginnen würde (z. B. in *Wild Bill, the coachman, said.*).

Viewer erkennt aus diesem Grund manchmal Satzgrenzen, die nicht mit den von *Ihnen* identifizierten Satzgrenzen übereinstimmen. (*I saw Mr. Smith.* wird als zwei Sätze erkannt; zahlreiche Überschriften sind zu einem Satz zusammengefügt.) Aus diesem Grund können Sie mit der Funktion *Kurze Sätze suchen* nach eigenartigen Ein-Wort-Sätzen [suchen](#).

Wie lang ist ein Satz?

Der Speicherplatz für jeden Satz oder Absatz liegt bei 10.000 Zeichen. **Viewer** kann bis zu 16.368 Sätze oder Absätze anzeigen. (Wenn der Durchschnitt bei 10 Wörtern pro Satz liegt, sind das insgesamt 160.000 Wörter.)

Kodierung: Datenträger, Akzentzeichen, etc.

Um beste Ergebnisse zu erzielen, verwenden Sie [ASCII](#)- oder [ANSI](#)-Versionen Ihrer beiden Texte.

Zum Speichern des durch Alignment synchronisierten Texts wird ein Windows-[Zeichensatz](#) verwendet.

Siehe auch: **Viewer: Inhalt**

9.10 falsch übereingestimmte Übersetzungen

Viewer can help find cases where alignment has slipped (one sentence having been translated as two or three). One method is to use the menu item *Match by Capitals*. This searches for matching proper nouns in the two versions: if say Paris is mentioned in sentences 25 of the source text and not in sentence 25 of the translation but in sentence 27, it is very likely that some slippage has occurred.

Viewer will search forwards from the current text sentence on, and will tell you where there's a mis-match. You should then search back from that point to find where the sentences start to diverge. It may be useful to sample every 10 or every 20 to speed up the search for slippage.

When you find the problem, [un-join](#) or [join](#) and/or edit the text as appropriate, then save it.

See also: [The technical side...](#), [Finding unusual sentences](#), **Viewer contents**

9.11 Fehlerbehebung

Der Satz oder Absatz ist nicht vollständig sichtbar.

Klicken Sie auf , um die Zeilen automatisch an die Fenstergröße anzupassen. So wird die Zeilenhöhe entsprechend der aktuell markierten Datenspalte angepasst.

Die Textdatei ist nicht vollständig sichtbar.

Klicken Sie auf , um die Anzeige zu aktualisieren. Viewer wird die Textdatei neu einlesen (bis zu höchstens 16.368 Sätzen).

Die Farben sind nicht schön.

Verändern Sie Farben im Hauptmenü unter *Einstellungen* | *Einstellungen anpassen*. Die Farben für jede Sprache im bilingualen Fenster sind diesselben wie die Farben für Primär- und Sekundärsortierung in **Concord**.

Siehe auch: Viewer: Inhalt

9.12 ungewöhnliche Sätze

Da **Viewer** Satzgrenzen anhand von Punkten, Fragezeichen und Ausrufezeichen erkennt, wird ein String wie "Hallo! Paul! Komm her!" in drei sehr kurze Sätze unterteilt. Unter Umständen möchten Sie, dass dieser String als ein Satz erkannt wird, z. B. wenn ein Übersetzer ihn so übersetzt hat: "Oi, Paulo, venha cá!".

Es kann hilfreich sein, ungewöhnlich kurze Sätze zu suchen und zu überprüfen, ob diese Sätze richtig verarbeitet wurden.

Mit dieser Funktion werden auch Zeilen gefunden, die mit einem Kleinbuchstaben beginnen: wenn ein Satz oder Absatz nicht mit einem Großbuchstaben oder einer Zahl beginnt, möchten Sie diesen wahrscheinlich mit dem vorangehenden Satz oder Absatz verbinden. Dieses Problem tritt häufig auf, wenn der Text als "Nur Text mit Zeilenumbrüchen" gespeichert wurde (wo sich am Ende jeder Zeile ein harter Zeilenumbruch befindet, unabhängig davon, ob der Absatz dort endet.)

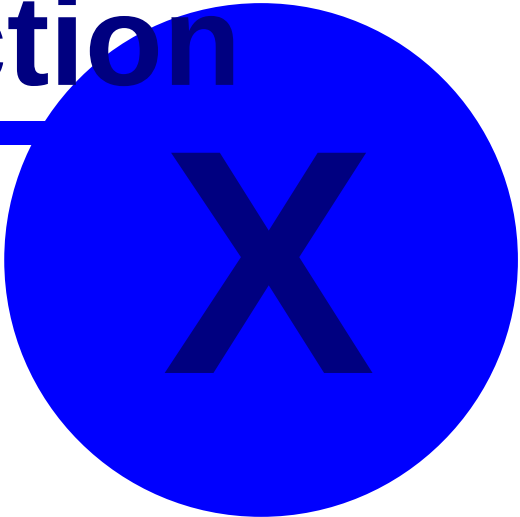
suchen

F8 oder *Kurze Sätze suchen*. **Viewer** sucht dann den nächsten problematischen Satz oder Absatz, und Sie können diesen durch Ziehen mit dem vorangehenden Satz oder Absatz [verbinden](#).

Siehe auch: [Technische Aspekte](#), [Falsch übereingestimmte Übersetzung finden](#), Viewer: Inhalt

Utility-Programme

Section



X

10 Utility-Programme

10.1 Daten aus früheren Versionen konvertieren

10.1.1 Daten aus früheren Versionen konvertieren

As Oxford WordSmith Tools develops, it has become necessary to store more data along with any given word-list, concordance etc. For example, data about which [language\(s\)](#) were selected for a concordance, [notes](#) now stored with every type of results file, etc. Therefore it has been necessary to supply a tool to convert data from the formats used in WS 1.0 to 3.0 to the new format for the current version.



This is the Data Converting tool.

If you try to open a file made with a previous version you should be offered a chance to convert it first.

10.2 WebGetter

10.2.1 Übersicht

Wozu dient das?

Das dient dazu, Ihren eigenen Textkorpus aufzubauen, indem Sie Webseiten mithilfe einer Suchmaschine herunterladen.

Wie mache ich das?

Geben Sie ein Wort oder eine Phrase ein, und klicken Sie auf Suchen oder drücken Sie die Eingabetaste.

Wie funktioniert das?

WebGetter besucht die Suchmaschine, die in dem zweiten Feld angegeben wurde, und lädt in etwa die ersten 100 Quellen herunter. Das Programm verwendet also die Suchmaschine dazu, eine Liste hilfreicher Referenzen zusammenzustellen. Dann sendet es einen Robot zu jeder Webadresse, und lädt die Webseite herunter (nicht aus dem Cache der Suchmaschine, sondern von der originalen Website). Es werden oft mehrere Robots gleichzeitig gesendet -- dadurch behindert eine langsame Übertragung nicht alle anderen Übertragungen.

Nachdem eine Webseite übertragen wurde, überprüft der WebGetter-Robot den Text auf die festgelegten Anforderungen (in [Einstellungen](#)). Wenn die Seite groß genug ist, wird eine Datei mit einem der Webadresse sehr ähnlichen Namen auf Ihrem Datenträger gespeichert.

Wenn alle Referenzen durchsucht wurden, sucht WebGetter mit der Suchmaschine nach weiteren Referenzen.

Siehe auch: [Einstellungen](#), [Anzeige](#), [Einschränkungen](#)

10.2.2 Einstellungen

Es gibt folgende Einstellungen:

- an welchem Ort die Texte gespeichert werden. Das festgelegte Verzeichnis wird als *Stammverzeichnis* verwendet. Wenn Sie das Verzeichnis `c:\temp` festlegen, und nach "besteiorl" suchen, werden die Ergebnisse in `c:\temp\besteiorl` gespeichert. Wenn Sie eine weitere Suche nach "Oxford WordSmith Tools" durchführen, werden die Ergebnisse unter `c:\temp\wordSmithTools` gespeichert.
- Time-out: die Anzahl der Sekunden, nach denen der WebGetter-Robot die Anforderung einer Webseite abbricht, die nicht antwortet. Empfohlener Wert: 20 Sekunden.
- max. gleichzeitig: WebGetter versendet mehrere Robots gleichzeitig, wobei jeder Robot eine andere Webseite anfordert. Empfohlener Wert: 20; das bedeutet maximal 20 Webseiten werden gleichzeitig übertragen.
- Sprache: Sie legen die Sprache fest, die Sie benötigen.
- Mindestdateilänge (empfohlen 20KB): die Mindestgröße für jede Textdatei, die aus dem Internet übertragen wird. Kleine Dateien enthalten möglicherweise nur Links zu ein paar Bildern und sonst nichts.
- Mindestwortanzahl (empfohlen 300): nach jedem Download durchsucht WebGetter die übertragenen Textdateien und zählt die Wörter. Wenn die Anzahl darunter liegt, wird die Datei nicht gespeichert.
- benötigte Wörter: Sie können wahlweise einige Wörter angeben, die in jedem Download vorkommen müssen; entweder alle Wörter oder mindestens eines der Wörter.

Suchmaschinen

Laden Sie eine Auswahl an Suchmaschinen, indem Sie auf Suchmaschinen klicken. Sie erhalten die neuesten Informationen über jede Suchmaschine unter www.lexically.net/downloads/searchengines.htm.

Erweiterte Optionen

Wenn Sie in einer Umgebung mit einem "Proxy Server" arbeiten, wird dieser von WebGetter automatisch erkannt und verwendet, es sei denn, Sie deaktivieren das entsprechende Kontrollkästchen. Wenn Sie sich nicht sicher sind, fragen Sie Ihren Netzwerkadministrator.

Der Raster mit Einstellungen

Dieser Raster enthält:

Name ignorieren	Der oben angezeigte Name, in der Liste der Suchmaschinen Websites, die beim Übertragen nicht besucht werden (im Gegensatz zum Anfordern einer Liste). Wenn WebGetter von Google eine Seite bekommt, ist nur die Google-Liste relevant, und nicht weitere Google Webseiten.
URL	Die URL, wo sich die Suchmaschine befindet.
Suchbegriff	Die Suchwort-Syntax
Max.	Wieviele Treffer bei jedem Kontakt überprüft werden
Nächste	
Sprache	Die gesuchte Sprache
Andere	

Das Suchwort wird mehr oder weniger so festgelegt, wie Sie es selbst bei einer Suche mit derselben Suchmaschine festlegen würden. Es werden nur wenige erweiterte Einstellungen für jede Suchmaschine verwendet; Sie können Ihre eigenen Einstellungen durch Eingabe in die Suchbegriffsspalte des Rasters ausprobieren. Finden Sie heraus, welche aktuellen Einstellungen eine Suchmaschine verwendet, indem Sie diese ausprobieren und dann den Suchbegriff

entsprechend anpassen. Einige Suchmaschinen möchten Cookies auf Ihrem PC hinterlegen, und dadurch können Übertragungsprobleme auftreten.

Sie sehen die Adresszeile in der Registerkarte Erweitert; WebGetter sendet an die Suchmaschine Informationen über Suchwort, maximal angezeigte Trefferanzahl pro Kontakt, Sprache der Treffer und die Suche weiterer Ergebnisse.

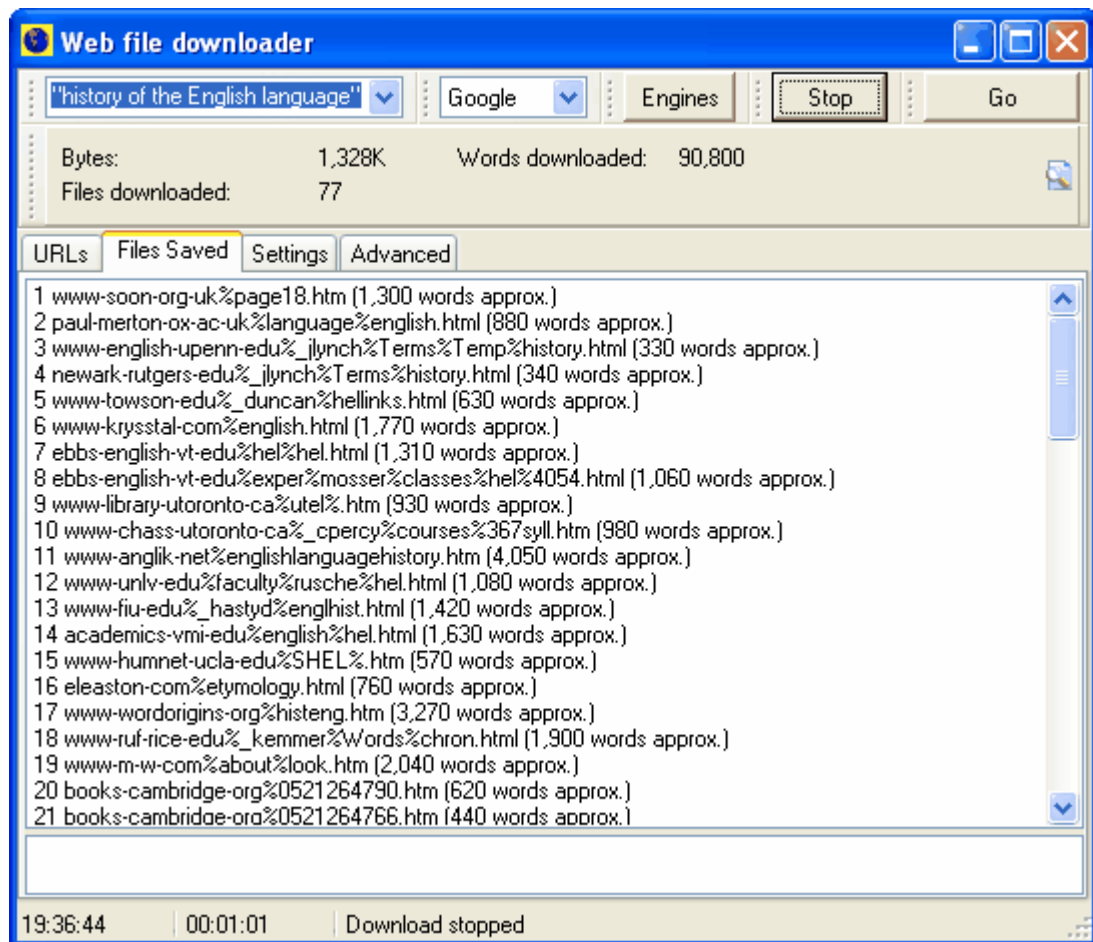
Siehe auch: [Anzeige](#), [Einschränkungen](#)

10.2.3 Anzeige

Während WebGetter arbeitet, werden die besuchten URLs angezeigt. Wenn die URLs grau angezeigt werden, war der beinhaltete Text zu kurz, um sinnvoll genutzt werden zu können, oder sie wurden noch nicht kontaktiert. Wenn die URLs dunkelblau angezeigt werden, wurden sie bereits auf dem Datenträger gespeichert. Oben wird angezeigt, wie viele Byte bereits eingelesen wurden, und jedes Mal, wenn eine den Anforderungen entsprechende Datei gespeichert wird, steigt die Anzahl der Dateien und Wörter. Unten wird die aktuelle Zeit und die abgelaufene Zeit angezeigt.

Es gibt eine Registerkarte, über die Sie auf eine Liste erfolgreich übertragener Dateien zugreifen können.

Hier sehen Sie eine Teilliste aus einer Suche des Ausdrucks "history of the English language" (*in* Anführungszeichen) mit einer Breitbandverbindung nach einer Minute und einer Sekunde.



Wie Sie sehen können, wurden etwa 1,3MB Webseiten untersucht, und darin wurden 90.000 Wörter (1,1MB) in Texten, die den Standardeinstellungen entsprachen (mindestens 10KB groß und 300 Wörter lang), gefunden und gespeichert. In dieser Zeit kam es ein paar Mal zu Time-Outs; das passierte wahrscheinlich deshalb, weil 20 Sekunden für manche komplizierte Sites oder langsame Server nicht ausreichend sind.

Siehe auch: [Einstellungen](#), [Einschränkungen](#)

10.2.4 Einschränkungen

Die Ergebnisse hängen ab von der Suchmaschine und den verwendeten Suchbegriffen. Das Internet ist ein riesiges schwarzes Brett; viele Informationen darin sind bloß Werbung oder Katalogpreise, etc. Die Ergebnisse für Suchbegriffe werden von den Suchmaschinen gesammelt, indem die Wörter untersucht werden, die der Autor der Webseite eingefügt hat. Es gibt keine Garantie dafür, dass die Webseiten tatsächlich von dem angegebenen Suchbegriff handeln, obwohl sie natürlich irgendeinen Zusammenhang aufweisen sollten.

Verwenden Sie die [Einstellungen](#), um Ihre Suche einzugrenzen.

Siehe auch: [Anzeige](#)

10.3 Languages Chooser

10.3.1 Übersicht

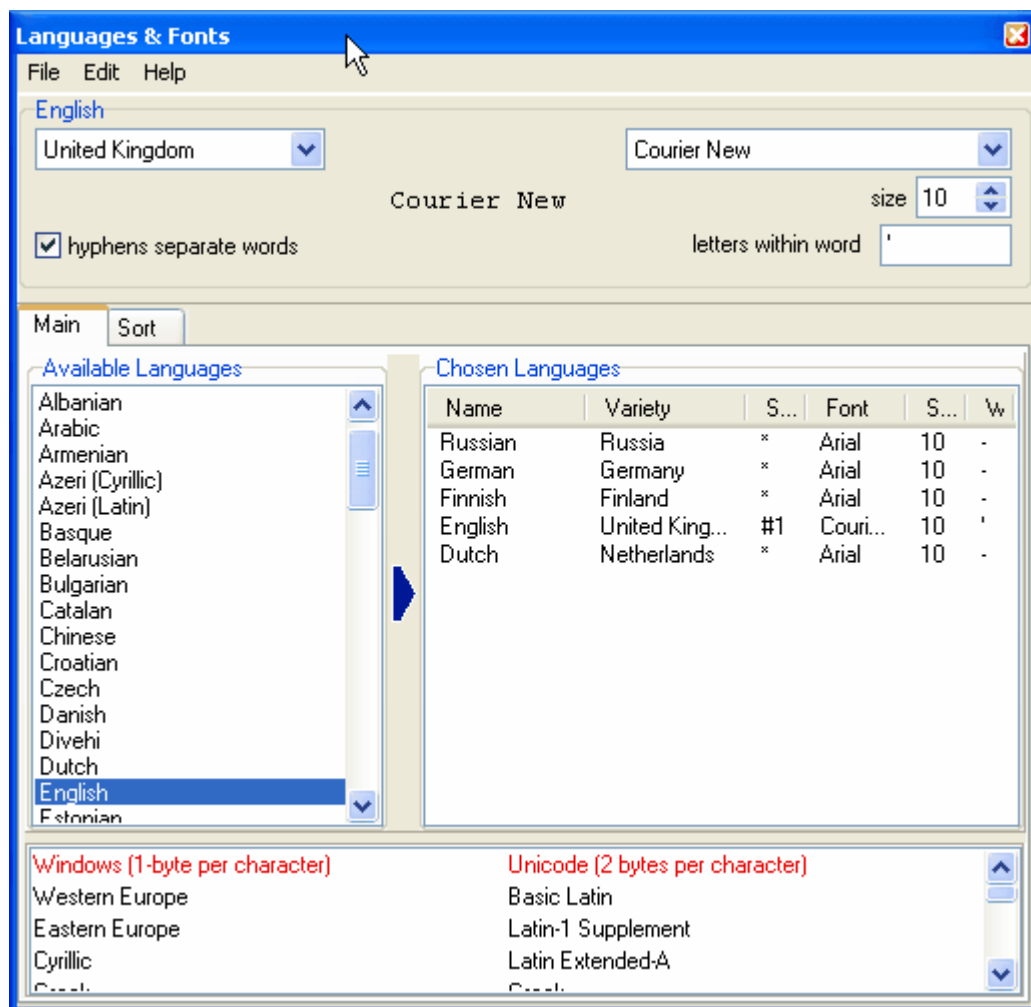


Der Languages Chooser ist ein Tool, mit dem man die Sprachen auswählen kann, die man verarbeiten möchte. Diese Einstellungen müssen Sie wahrscheinlich nur einmal vornehmen, und zwar dann, wenn Sie Oxford WordSmith Tools zum ersten Mal verwenden.

Wie finde ich das?

Der Languages Chooser ist über das Hauptmenü im WordSmith Controller aufrufbar: *Utilities | Languages Chooser*.

Was Sie dort sehen, sieht folgendermaßen aus:



Fünf Sprachen wurden bereits ausgewählt.

Am unteren Rand sehen Sie, welche Zeichen mit der aktuellen Schriftart dargestellt werden können (im Sinne von Windows ANSI- oder Unicode-Text). Unter Windows kann die Schriftart

"Courier New", die auf diesem PC gefunden wurde, die Zeichen für Westeuropäisch, Osteuropäisch, Kyrillisch, etc. anzeigen, sowie einige Bereiche des Unicode-Standards.

Siehe auch: [Sprache](#), [Schriftart](#), [Sortierreihenfolge](#), [Andere Sprachen](#), [Auswahl speichern](#)

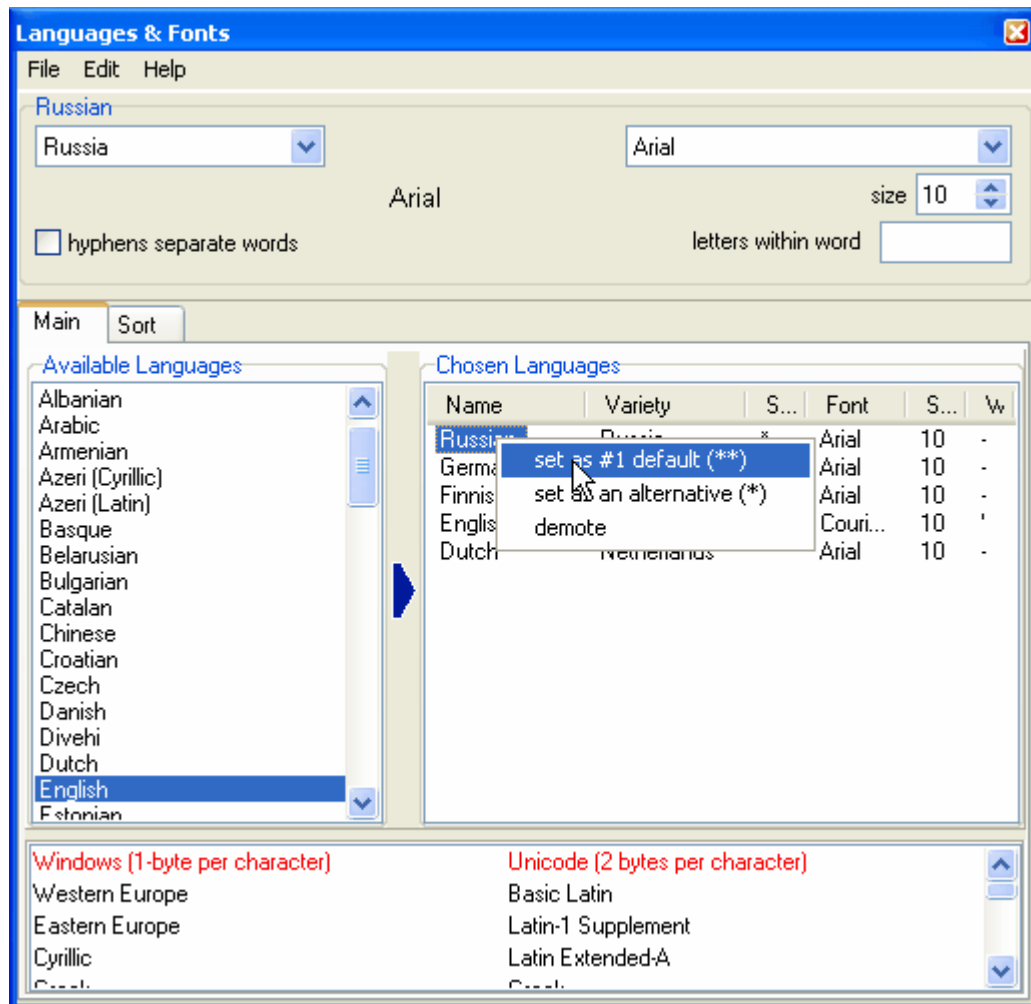
10.3.2 Sprache

Wie finde ich das?

Der Languages Chooser ist über das Hauptmenü im WordSmith Controller aufrufbar: *Utilities | Languages Chooser*.

Was kann ich dort einstellen?

Die Liste der Sprachen auf der linken Seite zeigt alle Sprachen an, die von dem PC, auf dem Sie arbeiten, unterstützt werden. Wenn manche davon in grau angezeigt werden, werden diese Sprachen von Ihrer Windows-Version "unterstützt", wurden aber auf bei Ihrem Betriebssystem noch nicht installiert. (Um einen umfangreicheren mehrsprachigen Support zu installieren, benötigen Sie die originale Windows CD-ROM, oder Sie können Hilfe im Internet finden.)



Auf der rechten Seite befinden sich die aktuell ausgewählten Sprachen für die Nutzung von WordSmith. Die Standardsprache sollte mit #1 markiert sein, und die anderen Sprachen, die Sie

eventuell verwenden möchten, mit *. Für jede ausgewählte Sprache können Sie die Symbole angeben, die innerhalb eines Worts vorkommen können, z. B. einen Bindestrich im Deutschen, weil es keinen Sinn ergibt, Wörter wie "Ich-Erzähler" als die Wörter "Ich" und "Erzähler" zu bearbeiten.

Um den Status der ausgewählten Sprache zu verändern, klicken Sie mit der rechten Maustaste auf die Sprache. Dieser Benutzer ist dabei, Russisch als Standardsprache auszuwählen. Um nicht gewollte Sprachen zu löschen, klicken Sie mit der rechten Maustaste auf die Sprache und wählen Sie "zurückstufen". Um eine Sprache hinzuzufügen, ziehen Sie zuerst die Sprache aus dem linken Fenster in das rechte und wählen dann Land und Ihre bevorzugte Schriftart für diese Sprache aus.

Jedes Mal, wenn Sie die Sprache wechseln, verändern Sie auch die Liste der verfügbaren [Schriftarten](#) und die Reihenfolge der [sortierten Wörter](#). Das Fenster am unteren Rand zeigt an, welche Zeichen von Unicode oder 1-Byte-Formaten in der hervorgehobenen Sprache unterstützt werden.

Einige Sprachen markieren [Worttrennzeichen](#) nicht.

Siehe auch: [Andere Sprachen](#), [Auswahl speichern](#)

10.3.3 Schriftart

Das Fenster "Schriftart" zeigt die Schriftarten an, die für die jeweilige Sprache vorhanden sind, abhängig davon, welche Schriftarten auf Ihrem Rechner installiert sind. Sie brauchen eine Schriftart, mit der die Zeichen, die Sie benötigen, angezeigt werden können: im Internet können Sie zahlreiche spezielle Schriftarten finden. [Unicode](#)-Schriftarten können eine enorme Anzahl verschiedener Zeichen umfassen, jedoch muss der Text auch im Unicode-Format gespeichert werden. Wenn Sie die Schriftart verändern, ändert sich auch die Liste der verfügbaren Zeichen.

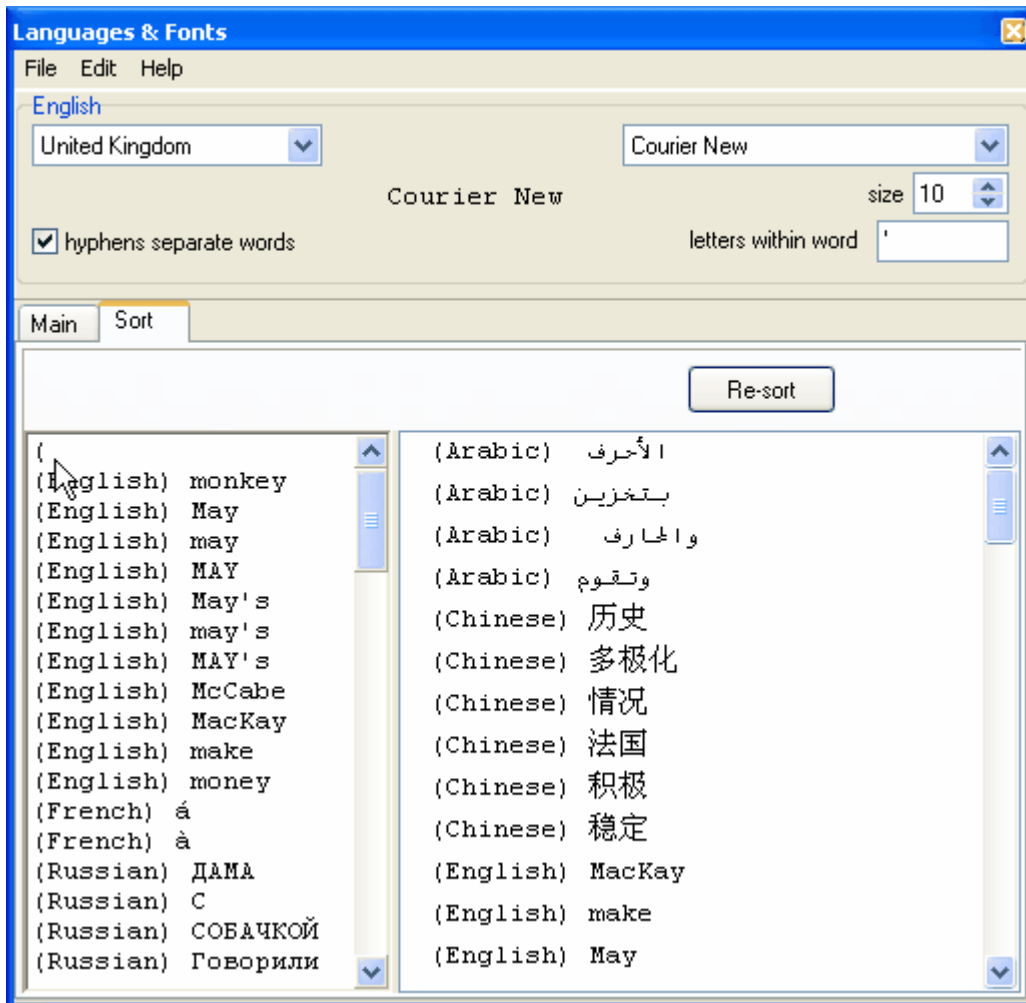
Klicken Sie hier, um mehr über [Unicode](#) zu erfahren.

Siehe auch: [Sprache](#), [Sortierreihenfolge](#), [Andere Sprachen](#), [Auswahl speichern](#)

10.3.4 Sortierreihenfolge

Die Sortierung wird an die ausgewählte Sprache angepasst. (Spanisch, Dänisch, etc. weisen Unterschiede zur deutschen Sortierung auf.)

Die Anzeige



- Sie sehen unter "Neu sortieren" zwei Fenster. Das linke Fenster enthält einige Wörter in verschiedenen Sprachen, zu denen Sie Ihre eigene Sprache hinzufügen können. Der Cursor im Screenshot zeigt auf das Feld, in dem der Benutzer gerade die Eingabe vornimmt. Es wurde bereits eine Klammer eingegeben. Wenn Sie das Zeichen mit Ihrer Tastatur nicht eingeben können, kopieren Sie es aus einem Ihrer Texte.
- Das Fenster auf der rechten Seite zeigt an, wie diese Wörter in der von Ihnen ausgewählten Sprache sortiert werden.

Siehe auch: [Sprache](#), [Schriftart](#), [Andere Sprachen](#), [Auswahl speichern](#)

10.3.5 Andere Sprachen

Wenn Sie mit einer Sprache arbeiten, die nicht in der Liste vorhanden ist, wählen Sie *Bearbeiten* und erstellen einen neuen Sprachnamen auf Basis einer der vorhandenen Sprachen. Wählen Sie eine Schriftart aus, die alle Zeichen und Symbole, die Sie verwenden möchten, anzeigen kann. Die Sortierreihenfolge ist gleich wie bei der Sprache, auf der Ihre neue Sprache basiert.

Siehe auch: [Sprache](#), [Schriftart](#), [Sortierreihenfolge](#), [Auswahl speichern](#)

10.3.6 Auswahl speichern

Speichern Sie Ihre Ergebnisse bevor Sie das Fenster schließen, damit Oxford WordSmith Tools Ihre Einstellungen bezüglich Schriftart und Sprache #1 bzw. untergeordneter Standardsprachen bis zum nächsten Öffnen beibehält und Sie die Einstellungen nicht erneut vornehmen müssen. Die Ergebnisse werden in `\wsmith4\language_choices.ini` gespeichert.

Siehe auch: [Sprache](#), [Schriftart](#), [Sortierreihenfolge](#), [Andere Sprachen](#)

10.4 Minimal Pairs

10.4.1 Zweck



A program for finding possible typos and pairs of words which are minimally different from each other (minimal pairs). For example, you may have a word list which contains ALEADY 5 and ALREADY 461, that is, your texts contain 5 instances where there is a possible misprint and 461 which are correct. This program helps to find possible variants and typos and anagrams.

See also : [requirements](#), [choosing your files](#), [output](#), [rules and settings](#), [running the program](#).

10.4.2 Anforderungen

Sie benötigen eine Wortliste im Textformat. Jede Zeile sollte ein Wort und seine Frequenz enthalten, und dazwischen einen Tabstopp, z. B.

```
THE          75.432
WAS          9.895
```

oder

```
1          THE          75.432
2          WAS          9.895
```

Sie können so eine Liste mit [WordList](#) erstellen. Zum Beispiel, markieren Sie die beiden Spalten, die Wörter und Frequenz enthalten, und klicken Sie auf die Schaltfläche ".txt"; danach

- deaktivieren Sie das Kontrollkästchen "jede Zeile nummerieren"
- Zeilen speichern = "alle" (wenn jedoch 0-xxx angezeigt wird, 0 durch 1 ersetzen)
- Spalten speichern = "alle markierten"

Siehe auch: [Zweck](#), [Dateien auswählen](#), [Output](#), [Regeln und Einstellungen](#), [Programm verwenden](#).

10.4.3 Dateien auswählen

- Choose your input word list (which must be in plain text format) by clicking the button at the right of the edit space and finding the word list .txt file.
- If it has numbered lines, check the ".txt is pre-numbered" box.
- If it has a header (WS3 will by default produce 3 lines of header information) make sure you have set the "Header lines to skip" box to the right number.

- You must specify where to save your results. The results will show all the typos and minimal pairs which the program finds.
- Choose also,
 - whether to number the list of results
 - whether to show the frequencies of possible typos
 - whether to show the rule which generated the result.

See also : [aim](#), [requirements](#), [output](#), [rules and settings](#), [running the program](#).

10.4.4 Output

Ein mögliches Beispiel für den Output ist

418 OBWOHLE (7) OBWOHL(37975)

Die Zeilen sind hier nummeriert, und die Zahlen in Klammer bedeuten, dass OBWOHLE sieben Mal und OBWOHL 37.975 Mal aufscheint.

Ein Beispiel mit Kleinschreibung unter Verwendung eines holländischen medizinischen Texts:

136 aplasie (1) aplasia(1)[L]

137 apyogene (1) apyogeen(1)[S]

138 arachnoideales (1) arachnoidales(1)[I]

In Zeile 136 ist ein Buchstabe falsch, in Zeile 137 sind zwei Buchstaben vertauscht und in Zeile 138 wurde ein falscher Buchstabe hinzugefügt.

Ein Beispiel aus der Zeitung "The Guardian" auf der Suche nach Anagrammen:

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]

Siehe auch: [Zweck](#), [Anforderungen](#), [Dateien auswählen](#), [Regeln und Einstellungen](#), [Programm verwenden](#).

10.4.5 Regeln und Einstellungen

Regeln

Einfügungen (abxcd vs. abcd)

Mit dieser Regel wird nach einem zusätzlichen Buchstaben gesucht, der in ein Wort eingefügt wurde, z. B. HOWWEVER.

Vertauschte Buchstaben (abcd vs. acbd)

Mit dieser Regel wird nach vertauschten Buchstaben gesucht, z. B. HOVEWER.

1 Buchstabe unterschiedlich (abcd vs. abxd)

Mit dieser Regel wird nach Wörtern gesucht, die sich nur um einen Buchstaben unterscheiden, z. B. HOWEXER.

Anagramme auch (abcd vs. adbc)

Mit dieser Regel werden dieselben Buchstaben in unterschiedlicher Reihenfolge gesucht, z. B. HWWVROEE.

Einstellungen:

Buchstaben am Wortende ignorieren:

Mit dieser Regel können Sie beliebige Buchstaben auswählen, die Sie am Ende des Worts ignorieren möchten, z. B. wenn Sie "s" festlegen, wird bei der Suche nach ELEPHANT das Wort ELEPHANTS nicht berücksichtigt.

Mindestübereinstimmung des Wortanfangs

Mit dieser Einstellung (Standard = 1) ist es Ihnen möglich festzulegen, nur Minimalpaare zu suchen, die mit demselben Buchstaben beginnen. Zum Beispiel müssen bei ALEADY nur alle Wörter mit dem Anfangsbuchstaben A auf Ähnlichkeit überprüft werden. Wenn dieser Wert auf 1 eingestellt wurde, wird BLEADY nicht als Minimalpaar erkannt. Um alle Wörter zu prüfen, stellen Sie den Wert auf 0. Das Programm benötigt dadurch die 26-fache Zeit!

Mindestwortlänge

Diese Einstellung legt die Mindestwortlänge fest, ab der die Möglichkeit eines Tippfehlers berechnet wird. Der Standard ist 5, was bedeutet, dass ein Wort mit vier Buchstaben einfach ignoriert wird. Dadurch wird die Verarbeitung beschleunigt, da die meisten Tippfehler eher in langen Wörtern vorkommen.

nur Wörter, die anfangen mit...

Wenn Sie diese Option auswählen, wird das Programm die folgenden Einstellungen ignorieren (max. Wortfrequenz). Hier können Sie eine Reihe verschiedener Buchstaben eingeben, z. B. F, G, H, und dann wird das Programm alle Wörter, die mit F, G oder H beginnen, unabhängig von deren Frequenz auf Minimal Pairs überprüfen, basierend auf den oben festgelegten Regeln und Einstellungen.

max. Wortfrequenz

(wird ignoriert, wenn "nur Wörter, die anfangen mit" aktiviert ist) Wie häufig kann ein Tippfehler vorkommen? Das hängt davon ab, auf wie viel Text Ihre Wortliste basiert. Der Standard ist 10, es wird also bei jedem Wort, das 11 Mal vorkommt, angenommen, dass es richtig eingegeben wurde.

Werkseinstellungen (die Standardwerte werden wiederhergestellt)

Aktuelle Einstellungen speichern (speichert Ihre aktuelle Auswahl für Datei und Regeln)

Gespeicherte Einstellungen laden (stellt Ihre zuletzt gespeicherte Auswahl wieder her)

Siehe auch: [Zweck](#), [Anforderungen](#), [Dateien auswählen](#), [Output](#), [Programm verwenden](#).

10.4.6 Programm verwenden

- Klicken Sie auf "Berechnen".
Sie sollten dann den Quelltext sehen, in dem ein paar Zeilen angezeigt werden. Einige der Zeilen und Spalten werden möglicherweise grau und andere weiß angezeigt: verschieben Sie die Spalten- und Zeilennummern, bis die richtigen Daten weiß und alle Überschriften oder Zeilennummern grau angezeigt werden.

Wenn Sie in der Mitte anhalten möchten, klicken Sie auf "Anhalten".

Die Statusleiste am unteren Rand des Fensters zeigt an, wie viele Wörter in der Wortliste gefunden wurden, wie viel Zeit verstrichen ist und die geschätzte verbleibende Dauer, um den gesamten Vorgang durchzuführen.

Sie können "Ergebnisse" anklicken, um die Ergebnisdatei anzuzeigen, wenn Sie den Vorgang abgeschlossen haben.

Zum Schluss auf "Beenden" klicken.

Siehe auch: [Zweck](#), [Anforderungen](#), [Dateien auswählen](#), [Output](#), [Regeln und Einstellungen](#)

10.5 Datei-Utilities

10.5.1 Index



Dieses Unterprogramm bietet mehrere allgemeine Funktionen:

[Zwei Dateien vergleichen](#)

[Datei-Chunker](#)

[Duplikate finden](#)

[Umbenennen](#)

Löcher finden: für "[Löcher](#)" in Textdateien

[Splitter](#)

[Joiner](#)

10.5.2 Splitter

10.5.2.1 Zweck von Splitter

Dieses Unterprogramm dient dazu, große Textdateien in viele kleine Dateien zu splitten. **Splitter** benötigt dazu folgende Angaben:

Trennzeichen des Textendes

Das Symbol, das als Markierung des Textendes gilt: z. B. **[FF]** oder **<ende der geschichte>** oder **</Text>** oder **!#** oder **[FF*]** oder **[FF?????]**

Einschränkungen:

- 1 Diese Markierung des Textendes muss am Anfang einer Zeile in der großen Originaldatei stehen.
- 2 Dabei werden Groß- und Kleinschreibung beachtet: **</Text>** erkennt die Markierung **</text>** nicht.
- 3 Das erste Zeichen eines Textende-Trennzeichens kann keine [Wildcard](#) sein, wie z. B. **#**, ***** oder **?**.
- 4 ***** und **#** können nur einmal in jedem Textende-Trennzeichen vorkommen.

Splitter erstellt jedes Mal eine neue Datei, wenn die von Ihnen festgelegte Markierung des Textendes gefunden wird.

Zielverzeichnis

In diesem Verzeichnis möchten Sie die kleinen Dateien speichern. (Wenn dieses Verzeichnis auf dem Netzwerk liegt, dann benötigen Sie dafür Schreibrechte.)

Erforderliche Größen

Die Mindest- und Höchstanzahl der Zeilen, die Ihre kleinen Dateien beinhalten dürfen (Standard = 2 und 30.000). Nur Dateien innerhalb dieser Grenzen werden gespeichert. Diese Funktion ist sehr hilfreich beim Extrahieren von Dateien aus sehr großen CD-ROM-Dateien. Die Standardeinstellung ist 2, um sehr kleine Textdateien zu vermeiden, z. B. *Kurznachrichten* aus der Zeitung. Wenn Sie jedoch an sehr kurzen Texten interessiert sind, legen Sie hier den Wert 1 fest.

Eine "Zeile" bedeutet von einer [Zeilenschaltung](#) bis zur nächsten.

Erste Zeile in Klammern setzen

Mit dieser Option legen Sie fest, ob die erste Zeile jeder neuen Textdatei in spitze Klammern (<

>) gesetzt wird oder nicht. Oft beinhaltet die erste Zeile nach einer Textende-Markierung eine Art Header. Wenn Sie nicht < am Anfang und > am Ende der ersten Zeile einfügen möchten, lassen Sie das Kontrollkästchen deaktiviert.

Titelzeile

Wenn Sie wissen, welche Zeile der Texte immer den Titel für den betreffenden Untertext enthält, legen Sie den Zähler auf diese Nummer fest, ansonsten belassen Sie ihn bei 0.

Siehe auch: [Joiner](#), [Dateinamen](#), [Wildcards](#), [Schaltflächen](#), Text Converter: Index.

10.5.2.2 Splitter: Dateinamen

Splitter erstellt viele kleine Dateien basierend auf einer oder mehreren großen Dateien.

Die neuen [Dateinamen](#) werden folgendermaßen erstellt:

Ein Verzeichnis mit dem Namen der Quelldatei wird erstellt. Gibt es zu viele Dateien für ein Verzeichnis, werden Unterverzeichnisse erstellt.

Erkennt Splitter eine Überschrift, enthält der Name jeder Datei den Titel, eine Nummer und .txt.

Gibt es keine Überschrift, enthält der Dateiname nur die Nummer und die .txt-Dateierweiterung.

So wie eine große Datei mit dem Namen HALLO.DAT in einige kleine Dateien gesplittet:

\HALLODAT\1.txt

\HALLODAT\2.txt

...

\HALLODAT\1\512.txt

etc.

Tipps

1. Splitter nummeriert jedes Mal ab 1 aufwärts.
2. Beachten Sie, dass die kleinen Dateien häufig viel mehr Speicherplatz benötigen als die Originaldatei. Das kommt daher, dass das Betriebssystem des Datenträgers eine Mindestdateigröße vorschreibt. Eine Textdatei mit nur einem Zeichen nimmt diesen Mindestspeicherplatz ein, der möglicherweise viele tausend Byte groß ist. Dennoch ist es empfehlenswert, die Textdateien so zu verwalten, dass jede Datei einen eigenen Text beinhaltet, und zwar mit Hilfe von Splitter. Beim Erstellen von Wortlisten und Schlüsselwortlisten arbeiten Sie jedoch besser mit [Batches](#).
3. CD-ROM-Dateien, die auf die Festplatte gespeichert werden, können schreibgeschützt sein. Das Attribut "Schreibgeschützt" können Sie mit Hilfe des Text Converters ändern.

10.5.2.3 Splitter: Wildcards

Die Raute, #, dient als Platzhalter für eine beliebige **Zahl**, also [FF#] findet [FF3] oder [FF9987], aber nicht [FF] oder [FF 9] (weil zwischen FF und 9 eine Leerstelle ist) oder [FFhallo].

* Das Sternchen * steht für einen beliebigen **String**, also [FF*] findet alle der oben angeführten Beispiele. * wird als letztes Zeichen im Textendesymbol verwendet. Damit würde auch [FF alles bis zum nächsten Zeilenumbruch gefunden.

^ Das Steuerzeichen ^ steht für einen beliebigen **Buchstaben**, also [FF^] findet [FFZQ], aber keines der anderen Beispiele.

? Das Fragezeichen steht für ein beliebiges **Zeichen** (inkludiert Leerstellen, Interpunktionszeichen, Buchstaben), also [FF??] findet von den Beispielen oben nur [FF 9], aber keines der anderen Beispiele.

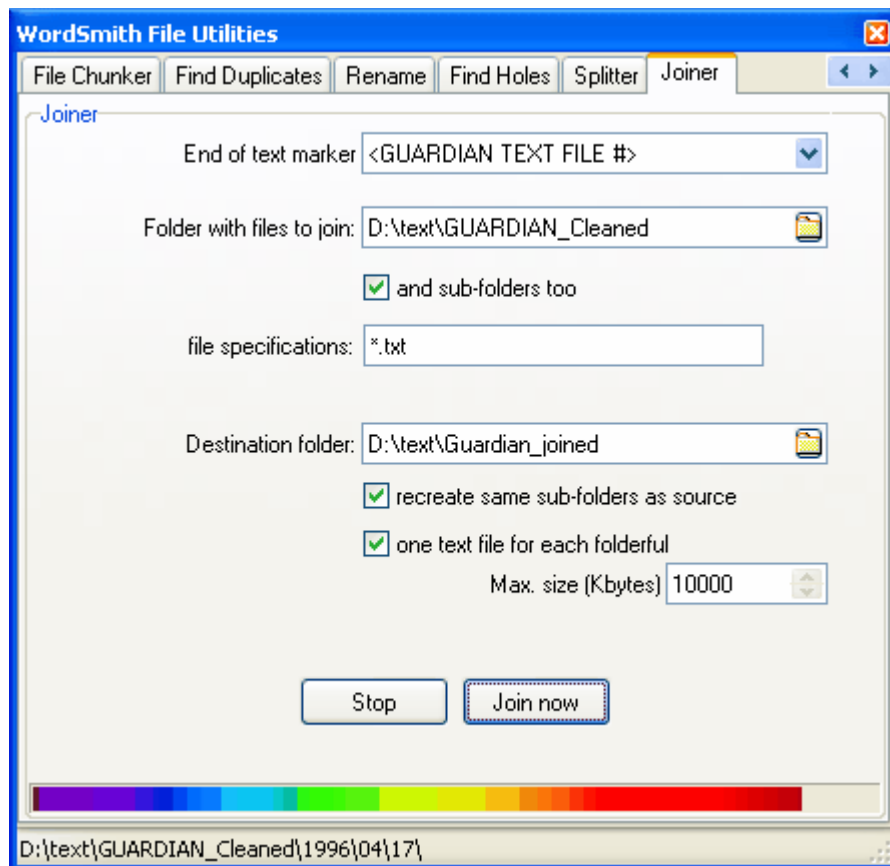
Um die Zeichen #, ^, ? oder * selbst zu suchen, setzen Sie das betreffende Zeichen in Anführungszeichen, z. B. "?" "#" "^" "*".

Siehe auch: [Einstellungen](#)

10.5.3 Textdateien zusammenführen

Dieses Unterprogramm dient dazu, kleine Textdateien zu größeren Dateien zusammenzuführen. Möglicherweise sind Sie nicht an den einzelnen Texten interessiert, wollen aber die Strukturen von zahlreichen Texten analysieren.

Wenn Sie die Funktion **Joiner** anwählen, wird folgendes Dialogfeld angezeigt:



Markierung des Textendes

Verschiedene Symbole können das Ende eines Texts markieren: z. B. **[FF]** or **<end of story>** oder **</Text>** oder **!#** oder **[FF*]** oder **[FF?????]**. Diese Markierung des Textendes finden Sie am Anfang einer Zeile in der großen Originaldatei. Das Zeichen # wird bei der Bearbeitung von der Textnummer ersetzt.

Verzeichnis mit zusammenzuführenden Dateien

Dort befinden sich die kleinen Dateien, die Sie zusammenführen möchten, im Augenblick. Sie werden nicht gelöscht, sondern in einem Zielverzeichnis zusammengeführt.

auch Unterverzeichnisse

Aktivieren Sie dieses Kontrollkästchen, wenn Sie auch Dateien in Unterverzeichnissen zusammenführen möchten.

Dateitypen

Geben Sie alle Dateitypen an, die Sie zusammenführen möchten, z. B. ***.*** oder ***.txt** oder ***.txt; *.ctx**.

Zielverzeichnis

In dieses Verzeichnis möchten Sie die kleinen Dateien kopieren und zusammenführen. (Liegt dieses Verzeichnis auf dem Netzwerk, dann benötigen Sie dafür Schreibrechte.)

Unterverzeichnisse wie in Quelle

Ist dieses Kontrollkästchen aktiviert, wird dieselbe Verzeichnisstruktur wie im Quellverzeichnis hergestellt. In diesem Beispiel werden alle Unterverzeichnisse von **d:\text** **\guardian_cleaned** unter **d:\text\guardian_joined** neu erstellt.

ein Text pro Verzeichnis

Ist dieses Kontrollkästchen aktiviert, werden alle Quelldateien eines Verzeichnisses zu einer Datei im Zielverzeichnis zusammengeführt.

Max. Größe (KB)

Die maximale Größe der Zieldatei(en) in Kilobyte. 1.000 bedeutet, dass Ihre Zieldateien jeweils fast 1 Megabyte groß sein werden. Das entspricht etwa 150.000 Wörtern ohne Tags und auf Englisch. Diese Option ist nur verfügbar, wenn *ein Text pro Verzeichnis* nicht aktiviert ist.

Schaltfläche "Stopp"

Unterbricht den Vorgang.

Siehe auch: Splitter, Text Converter: Index.

10.5.4 zwei Dateien vergleichen

The point of it

The idea is to be able to check whether 2 files are similar or not. You may often make copies of files and a few weeks later cannot remember what they were. Or you have used [File Chunker](#) to copy a big file to floppies and want to be sure the copy is identical to the original.

This program checks whether

- a) they are the same size
- b) they have the same contents
(it goes through both, byte by byte, checking whether they match)
- c) they have the same attributes
(file attributes can be "read only" [you cannot alter the file], "system" [a file which Windows thinks is central to your operating system], "hidden" [one which is so important that Bill Gates may be reluctant to even let you know it exists on your disk])
- d) they have the same time & date.

How to do it

Specify your 2 files and simply press "Compare".

See also : [file chunker](#), [find duplicates](#), [rename](#)

10.5.5 Datei-Chunker

Wozu dient das?

Zweck dieses Utilities ist es, eine große Datei in kleine Dateien aufzuteilen, damit man diese Dateien auf einer Diskette oder einer CD-ROM speichern kann. Es wäre sonst unmöglich, eine 5MB große Datei auf 3 oder 4 Disketten zu speichern und auf einen anderen PC zu kopieren.

An dem anderen PC müssen die Chunks dann natürlich wieder zu einer Datei zusammengesetzt werden.

Wie mache ich das? eine Datei kopieren

1. Wählen Sie die "zu teilende Datei" aus (die Sie auf einen anderen PC kopieren möchten).
2. Wählen Sie "Laufwerk und Verzeichnis aus", in dem Sie die Chunks speichern möchten. Wenn Sie A: wählen, müssen Sie für jeden Chunk eine formatierte Diskette einlegen.
3. Geben Sie die "Größe jedes Chunks" ein (Standard = 1.400K, der Speicherplatz auf einer Diskette)
4. Wählen Sie aus, ob Sie bei diesem Vorgang die "Datei komprimieren" möchten.
5. Klicken Sie auf "Kopieren".

Wie mache ich das? eine Datei wiederherstellen

1. Wählen Sie den ersten Chunk aus ("Chunk").
2. Wählen Sie das Verzeichnis, in dem Sie die Datei "wiederherstellen" möchten.
3. Wählen Sie aus, ob die Chunks nach dem Wiederherstellen gelöscht werden sollen ("Chunks löschen").
4. Klicken Sie auf "Wiederherstellen".

Siehe auch: [zwei Dateien vergleichen](#), [Duplikate finden](#), [umbenennen](#)

10.5.6 Duplikate finden

The point of it

The idea is to be able to check whether you have files with the same name in different folders. You may often make copies of files and a few weeks later cannot remember where they were.

This program only checks whether the files it is comparing have the same name. (You could use [Compare 2 Files](#) to see whether they are in fact identical.) It handles lots of folders, the point being to locate unnecessarily duplicated files or confusing reuse of the same filenames.

How to do it

Specify your Folder 1 and simply press "Search". *Find Duplicates* will go through that folder and any sub-folders and will report any duplicates found.

Or you can specify 2 different folders (e.g. on different drives) and the process compares

one set with the other.

See also : [compare two files](#), [file chunker](#), [rename](#)

10.5.7 umbenennen

Wozu dient das?

Mit dieser Funktion können viele Dateien in einem Vorgang in einem oder mehreren Verzeichnissen umbenannt werden. Möglicherweise haben Ihre Dateien besonders lange Namen, und stellen deshalb ein Problem für bestimmte Anwendungen dar. Oder aber es ist einfach zu aufwändig, die Dateien alle einzeln umzubenennen.

Wie mache ich das?

Geben Sie Ihr Verzeichnis an, die Dateitypen, die Sie umbenennen möchten, und ob Unterverzeichnisse auch verarbeitet werden sollen.

Die Standardeinstellung ist *Alle Dateien *.*.*

Geben Sie eine "Maske für den neuen Namen" ein und eine Anfangszahl.

Zum Beispiel, mit der Maske **SUN** und der Anfangszahl 0 wird die erste gefundene Datei, die ursprünglich **Langer_und_komplizierter_Name.txt** hieß, zu **SUN0.txt** umbenannt. Die nächste Datei wird zu **SUN1.txt** umbenannt, und so weiter. (Wenn die nächste Datei **Langer_und_komplizierter_Name.htm** heißt, wird sie umbenannt zu **SUN2.htm**)

Wenn Sie "Dateien finden" anklicken, sehen Sie eine Liste aller Dateien, die diese Kriterien erfüllen. Wenn Sie nun auf "Umbenennen" klicken, wird jede Datei entsprechend Ihrer Einstellungen neu benannt.

Siehe auch: [zwei Dateien vergleichen](#), [Datei-Chunker](#), [Duplikate finden](#)

10.6 Text Converter

10.6.1 Zweck

This program does a "Search & Replace", on virtually any number of files.

It is very useful for going through large numbers of texts and re-formatting them as you prefer, e. g. taking out unnecessary spaces, ensuring only paragraphs have <Enter> at their ends, changing accented characters, ensuring you have Windows £ symbols, etc.

converting text

For a simple search-and-replace you can type in the search item and a replacement; for more complex conversions, use a [Conversion File](#) so that **Text Converter** knows which symbols or strings to convert. It operates under Windows and saves using the Windows [character set](#), but will convert text using DOS or Windows character sets. You can use it to make your text files suitable for use with your Internet browser.

It does a "search and replace" much as in word-processors, but it can do this on lots of text files, one after the other. As it does so, it can also replace up to **any number of** strings, not just one.

Once the conversion file is prepared and [Settings](#) specified, the **Text Converter** will read each source file and either create a new version or replace the old one, depending on the [over-write setting](#).

You will be able to see the details of how many instances of each string were found and replaced overall.

filtering files

And/or you may need to make sure texts which meet certain criteria are [put into the right folders](#).

Tip

The easiest way to ensure your text files are the way you want, especially if you have a very large number to convert, is to copy a few into a temporary folder and try out your conversion file with the Text Converter. You may find you've failed to specify some necessary conversions. Once you're sure everything is the way you want it, delete the temporary files.

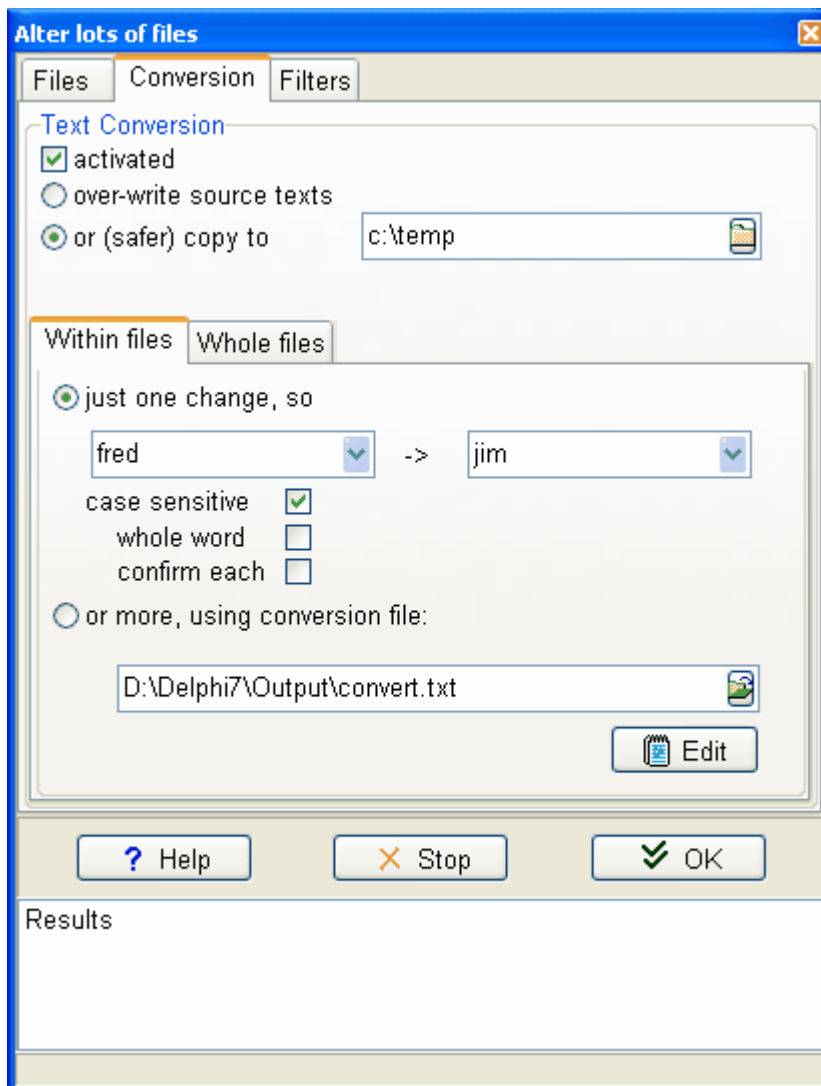
See also: Text Converter Contents, [The buttons](#)

10.6.2 Text Converter: Einstellungen

1. Choose *Files* (the top left tab). Decide whether you want the program to process sub-folders of the one you choose. There is no limit to the number of files Text Converter can process in one operation.
2. Click on the *Conversion* tab, and:
3. Decide whether you want to make copies of the text files, or to over-write the originals. Obviously you must be confident of the changes to choose to over-write; copying however may mean a problem of storage space.
4. Specify what to convert, that is the search-words and what you want them to be replaced with. For a quick conversion you can simply type in a word you want to change and its replacement (e.g. *Just one change* so that **responsible** becomes **responsible**) or you can choose your own pre-prepared [Conversion File](#).
5. Or in the Whole Files section you can choose simply to update legacy files by choosing
Dos to Windows,
Unix to Windows,
MS Word .doc to .txt or
into Unicode.
6. If you might want some files not to be converted, or simply don't want any conversions but instead to place files in appropriate sub-folders, choose the [Filters](#) tab.

If you choose *Over-write Source texts*, Text Converter will work more quickly and use less disk space, but of course you should be quite sure your conversion file codes are right before starting!

Note that **some space on your hard disk will be used even if you plan to over-write**. The conversion process does its work, then if all is well the original file is deleted, and the new version copied. There has to be enough room in the destination folder for the largest of your new files; it is much quicker for it to be on the same drive as the source texts. If it isn't, your permission will be asked to use the same drive.



Press **OK** to start; you will see a list of results, as in the screenshot below.
If you want to stop **Text Converter** at any time, click on the Cancel button or press Escape.

```
D:\Temp\test.txt converted = c:\temp\test.txt
1 done
```

Right-click to see the source or the converted result file.

See also: Text Converter Contents.

10.6.3 Text Converter: Syntax

The syntax for a [Conversion File](#) is:

Only lines beginning / or " are used. Others are ignored completely.
Every string for conversion is of the form "**A**" -> "**B**". That is, the original string, the one

you're searching for, enclosed in double quotes, is followed by a space, a hyphen, the > symbol, and the replacement string.

Control Codes

Control codes can be symbolised like this: {CHR(xxx)} where xxx is the number of the code.

Examples: {CHR(13)} is a carriage-return, {CHR(10)} is a line-feed, {CHR(9)} is a tab, {CHR(12)} is a printer form-feed. To represent <Enter> which comes at the end of paragraphs and sometimes at the end of each line, you'd type {CHR(13)}{CHR(10)} which is carriage-return followed immediately by line-feed.

Use {CHR(34)} if you need to refer to double inverted commas.

Wildcards (*,?,# and ~)

* You can use the asterisk as a wildcard. Thus "<*>" -> "" will delete any string in < > brackets from your text. "<head */head>" will delete any string starting "<head " and ending "/head>", even if there are hundreds of characters between them. The default search distance is 1,000 characters, with a maximum of 25,000. (The text is read chunk by chunk into a 30,000 character buffer, so the maximum will work fine at the start of the text; after this only 1,000 characters of search-space are guaranteed.) As deleting a lot of text can get rid of more text than you expect if the text is not properly marked up in the first place, you will probably need to over-ride the default search distance by specifying it in brackets, e.g. "<head*(100)/head>". The asterisk may not be the first or last symbol between the double quotation marks in the search-string.

The asterisk also retains up to 1,000 characters. "<div*(100)>" remembers all the characters up to > and can use them in the replacement: Thus "<div*(100)>" -> "[section *]" will produce [section 1 They Meet Again] if the original has <div1 They Meet Again>. "<div*>" will do the same thing but would allow up to 1,000 characters' search for the >.

Use # to symbolise any number. "<div#>" will find <div1>, <div2>, <div468>, etc. If # is in the replacement too, the exact same number will be used in the replacement. Thus "<div#>" -> "[section #]" will produce [section 468] if the original has <div468>.

? The question mark stands for any single character, except a space. Up to ten ?s can be used in the replacement string to reproduce the character referred to by the ?s in the search-string.

~ The tilde means except. ~"<p>" "<*>" -> "" means delete everything in between angle brackets, except a case of <p>.

Use {CHR(42)} if you need to refer to *, {CHR(35)} for #, {CHR(63)} for ? and {CHR(126)} for ~.

Whole word, case Insensitive, Confirm, redundant Spaces

/C stops to confirm you wish to go ahead before each change.

/W does a whole word search (ensuring the alteration only happens if there's a [word separator](#) on either side) (/W "the" finds the but not other or then or bathe).

/I does a case insensitive search (/I "restaurant" -> "hotel" replaces restaurant with hotel and RESTAURANT with HOTEL and Restaurant with Hotel, i.e. respecting case as far as possible). You can combine these, e.g.

/IWC "the" -> "this"

/S cuts out all redundant spaces. That is, it will reduce any sequence of two or more spaces to one, and it also removes some common formatting problems such as a lone space after a carriage-return or before punctuation marks such as .,; and). /S can be used on a line of its own or in combination with other searches.

Additions (/A, /T and {v})

/A means add text. /A "Ulan" START inserts Ulan at the start, /A "Bator" END inserts Bator at the end of the text. See \wsmith4\convert.txt to see one in use.

`/T` means add title. So `/T "<title>*</title>" -> ""` looks for `<title> ... </title>` and if it's found, inserts the wording given into the file. This will make your browser show the title at the top of the screen.

`{v=""}` means remember this and use it in another line of the conversion file when you find `{v}`. `"26 Dec." -> "Boxing Day" {v="Xmas"}` stores the reference `Xmas` and `"1 May" -> "Mayday" {v="after Easter"}` stores `after Easter` for use in a later line, such as `"/celebration/" -> "{v}"`. Assuming that your text has a mention of `26 Dec.` and `1 May`, this example, on finding `/celebration/` in the text, will put `Xmas` if the most recent mention in the text was `26 Dec.` and `after Easter` if the most recent mention was `1 May`.

See `\wsmith4\convert.txt` to see examples in use.

See also: Text Converter Contents.

10.6.4 Text Converter: verschieben, wenn

Mithilfe dieser Funktion werden alle Dateien, die das von Ihnen angegebene Wort bzw. die von Ihnen angegebene Phrase beinhalten, in ein neues Verzeichnis verschoben.

Wozu dient das?

Angenommen Sie verfügen über eine Reihe von Dateien, manche davon mit Dialogen zwischen Werther und Lotte, andere mit Texten über die Chinesische Mauer oder die Anatomie von Fliegen. Sie möchten die Dateien mit den Werther-Lotte-Dialogen finden und in ein Verzeichnis mit dem Namen *Leiden* verschieben.

Wie mache ich das?

1. Klicken Sie auf die Registerkarte *Filter* (oben).
2. Aktivieren Sie das Kontrollkästchen *Aktiviert*.
3. Geben Sie das Wort oder die Phrase an, die in dem Text vorkommen muss. Bei dieser Funktion werden Groß- und Kleinschreibung beachtet.
4. Wählen Sie aus, an welcher Stelle im Text das Wort steht:
 - beliebige Stelle,
 - an einer beliebigen Stelle vor einem anderen Wort/einer anderen Phrase, oder
 - zwischen zwei verschiedenen Wörtern oder Phrasen.
5. Geben Sie an, was mit den Dateien geschehen soll, die diese Bedingungen erfüllen:
 - nichts
 - in ein bestimmtes Verzeichnis kopieren, oder
 - in ein bestimmtes Verzeichnis verschieben, oder
 - Datei löschen (Vorsicht!).

Sie können auch basierend auf dem unter 3. festgelegten Wort (bzw. Phrase) ein Unterverzeichnis anlegen.

Sie haben auch die Möglichkeit, die Erweiterung `.txt` hinzuzufügen (nützlich bei Dateien, die keine Erweiterung haben, z. B. BNC-Dateien)

Siehe auch: Text Converter: Inhalt.

10.6.5 Konvertierungsdatei für Text Converter

Erstellen Sie Ihre Konvertierungsdatei für Text Converter mit einem [Standardtext](#)-Editor wie dem Windows Texteditor.

Sie können die Datei basierend auf `\wsmith4\convert.txt` erstellen.

Enthalten die Originaldateien [Akzentzeichen](#), verwenden Sie für die Konvertierungsdatei den

DOS-Editor, wenn die Dateien ursprünglich unter DOS geschrieben wurden, bzw. einen Windows Editor, wenn die Dateien ursprünglich in einem Windows Textverarbeitungsprogramm geschrieben wurden. Einige Windows Textverarbeitungsprogramme können beide Formate bearbeiten.

Es können unbegrenzt viele Zeilen konvertiert werden, und jede Zeile kann zwei Strings enthalten, die durch Anführungszeichen (" ") getrennt werden müssen und jeweils bis zu 80 Zeichen lang sein können.

Der Text Converter führt alle Veränderungen in der Reihenfolge aus, wie sie in der Konvertierungsdatei festgelegt wurden. Beachten Sie, dass eine Veränderung Einfluss auf darauf folgende Veränderungen haben kann.

Veränderungen, bei denen die Originaldatei vergrößert wird

Durch die meisten Veränderungen wird die Größe der Originaldatei verringert. Text Converter führt auch Veränderungen durch, bei denen die Originaldatei vergrößert wird -- solange genügend Speicherplatz vorhanden ist!

Tipp

Um Zeilenschaltungen am Ende einer Zeile, nicht jedoch am Ende eines Absatzes zu entfernen, untersuchen Sie die Absatzenden auf Elemente, die nur am Ende eines Absatzes vorkommen. Wenn Absätze zum Beispiel mit zwei Zeilenschaltungen enden, verwenden Sie die folgenden Zeilen in Ihrer Konvertierungsdatei:

```
"{CHR(13)}{CHR(10)}{CHR(13)}{CHR(10)}" -> "{%}"
```

(mit dieser Zeile werden zwei Zeilenschaltungen durch {%} ersetzt.) (Es könnte jede andere eindeutige Kombination sein. Der Ersetzungsvorgang ist etwas schneller, wenn das gesuchte Element gleich viele Zeichen enthält wie das Element, durch das es ersetzt wird; in diesem Fall 4 Zeichen.)

```
"{CHR(13)}{CHR(10)}" -> " "
```

(mit dieser Zeile werden alle Zeilenschaltungen durch ein Leerzeichen ersetzt, um die Trennung von Wörtern aufrechtzuerhalten)

```
"{%}" -> "{CHR(13)}{CHR(10)}{CHR(13)}{CHR(10)}"
```

(mit dieser Zeile wird die Kombination {%} durch zwei aufeinander folgende Zeilenschaltungen ersetzt, wodurch die ursprüngliche Absatzstruktur wiederhergestellt wird).

```
/S
```

(mit dieser Zeile werden alle überflüssigen Leerstellen entfernt)

Siehe auch: [Vorlage Konvertierungsdatei](#), [Syntax](#), Text Converter: Inhalt.

10.6.6 Text Converter: Vorlage Konvertierungsdatei

Sie können den gesamten folgenden Text oder einen Teil davon in Ihre [Zwischenablage](#) kopieren und im Windows Editor einfügen.

```
[ Kommentarzeile -- hier können Sie einfügen, was Sie möchten, denn
diese Zeile wird ignoriert ]
```

```
[ zuerst eine Rechtschreibkorrektur ]
```

```
"responsable" -> "responsible"
```

```
[ nun werden die spitzen Klammern < > durch [ ] ersetzt, und die
geschwungenen Klammern { } durch ( ) ersetzt ]
```

```
"<" -> "["
```

```
">" -> "]"
```

```
"}" -> ")"
```

```
"{" -> "("
```

```
/S
```

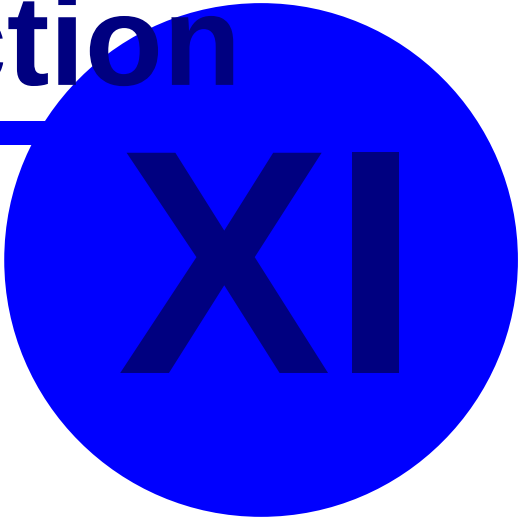
```
[ dadurch werden alle überflüssigen Leerstellen entfernt ]
```

Die Datei `\wsmith4\convert.txt` ist eine Konvertierungsdateivorlage, die für die Konvertierung von Textdateien aus dem British National Corpus verwendet wird.

Siehe auch: Text Converter: Inhalt.

Tags und Markup

Section



XI

11 Tags und Markup

11.1 Übersicht

Wozu dient Markup?

Markup-Text ist Text, in dem zusätzliche Informationen in Form von *Tags* eingebaut sind, z. B. "Wir<pronoun> mögen<verb> Spaghetti<noun>.<end of sentence>". Sie können nicht nur Wörter, sondern auch Tags in eine Konkordanz einbeziehen.

Sie möchten möglicherweise die zusätzlichen Informationen *anzeigen* oder *ignorieren*, sodass Sie nur den Text sehen ("Wir mögen Spaghetti"). **Oxford WordSmith Tools** wurde so gestaltet, dass Sie auswählen können was angezeigt und was ignoriert wird.

Sie möchten möglicherweise [HTML oder SGML](#) Tags oder Entity-Referenzen *übersetzen*: Wenn Ihr Text **É** enthält, möchten Sie stattdessen wahrscheinlich eher **É** sehen. Sie können möglicherweise innerhalb der Textdateien *auswählen*, z. B. einen Header auslassen oder anstelle des gesamten Texts nur die Konklusion anzeigen.

Und Sie möchten mit **Oxford WordSmith Tools** vielleicht nur Dateien bearbeiten, die bestimmte Kriterien erfüllen, z. B. mit "sex=f" im Header einer Textdatei, wo also die Rede von einer Frau gehalten wird.

Wählen Sie Tags aus, können Sie den Effekt davon in der Option Texte Auswählen ansehen, indem Sie die Schaltfläche [anzeigen](#) anklicken. Alle beizubehaltenden Tags werden angezeigt und alle ignorierten Tags werden durch Leerstellen ersetzt.

Siehe auch: [Tag-Handling](#), [Eine Tagdatei erstellen](#), [Nächstes Tags in Concord anzeigen](#), [Audio und Video in Concord](#), [Tag-Konkordanz erstellen](#), [Tagtypen](#), [Tags anzeigen](#), [Tags als Selektoren verwenden](#), [Tags in WordList](#)

11.2 Tag-Arten

Sie müssen festlegen, wie jeder Tagtyp beginnt und endet, und Sie sollten die Tags konsistent verwenden. Beschränken Sie sich auf Zeichen, die sonst nicht in Ihren Texten vorkommen.

acht spezielle Markierungen

Acht Markierungsarten sind entscheidend für Wortlisten: diejenigen, die [Anfang und Ende von Überschriften](#), [Abschnitten](#), [Sätzen](#) und [Absätzen](#) markieren. Geben Sie diese Tags in den entsprechenden Feldern an, wenn Sie die [Texteigenschaften](#) einstellen.

Tags innerhalb von zwei Trennzeichen

Diese Tags werden oft dazu verwendet, die Wortart jedes Worts zu bestimmen; sie werden in [HTML, XML, SGML](#) auch als "Schalter" verwendet, z.B. **<H1>** um eine Überschrift 1 zu beginnen und **</H1>** um sie zu beenden. Sie sollten bei allen Tags jeweils dieselben Zeichen für Anfang und Ende wählen, normalerweise in einer bestimmten Art von Klammer (wie der British National Corpus, der [SGML](#)-Tags verwendet): **<Noun>**, **<Verb>**, **<Pronoun>**.

Entity-Referenzen

[HTML, XML und SGML](#) verwenden so genannte Entity-Referenzen für Symbole, die nicht Teil des Standardalphabets sind, z.B. **é**; **té** das für **été** steht.

Legen Sie diese zwei Markup-Arten fest, indem Sie 'Einstellungen | Taglisten' oder 'Einstellungen | Texteigenschaften | Tags auswählen'. Sie sehen dann ein Dialogfeld mit 'Text ignorieren' und die Schaltfläche 'Durchsuchen'.

Die Option [Tags ignorieren](#) ermöglicht es Ihnen Tags festzulegen, die in der Konkordanz oder Wortliste unterdrückt werden sollen.

Die Option [Tags, die inkludiert werden](#) ermöglicht es Ihnen eine Tagdatei mit Tags festzulegen, die in der Konkordanz oder Wortliste angezeigt werden sollen.

Mit der Option [zu übersetzende Entities](#) können Sie Entity-Referenzen festlegen, die Sie während der Verarbeitung konvertieren möchten, wie z. B. **´**.

Multimedia-Markierungen

In Textdateien können Tags als Referenz für Audio- oder Videodateien dienen, die wiedergegeben werden können. Zum Beispiel, ein Text enthält etwa Folgendes: **bla bla bla ... bla bla** etc. Eine Konkordanz für **bla bla** könnte das Tag aufnehmen, damit Sie die Quelle (eine .mp3-Datei) wiedergeben können. Siehe [Multimedia-Tags definieren](#).

Siehe auch: [Tag-Übersicht](#), [Tag-Handling](#), [Eine Tagdatei erstellen](#), [Nächstes Tags in Concord anzeigen](#), [Tag-Konkordanz erstellen](#), [Tags anzeigen](#), [Tags als Selektoren verwenden](#), [Audio und Video in Concord](#)

11.3 Umgang mit Tags

alle Tags ignorieren

Legen Sie alle öffnenden und schließenden Symbole fest unter Einstellungen Anpassen | Tags und Mark-up | Mark-up ignorieren und solche Tags werden in Wortlisten und Konkordanzen unterdrückt (als wären sie in den originalen Textdateien nicht vorhanden).

Beispiel: **<*>** So werden alle Wörter zwischen dem **<** Symbol und dem nächsten **>** Symbol ignoriert (bis zu 1.000 Zeichen entfernt).

manche Tags ignorieren und andere beibehalten

Wenn Sie manche Tags ignorieren und andere beibehalten möchten, müssen Sie eine [Tagdatei](#) erstellen, in der alle Tags aufgelistet sind, die Sie beibehalten möchten. Diese Tags erscheinen dann in Ihren Wortlisten und Konkordanzen.

Diese Datei können in Oxford WordSmith Tools unter Einstellungen in der Menüoption Tags festlegen. Solche Tags werden dann in Ihre Wortlisten, Konkordanzen, etc. aufgenommen, als wären sie normale Wörter oder Suffixe.

Beispiel: Angenommen Sie haben **<*>** unter "Tags ignorieren" festgelegt, aber **<title>**, **<body>** und **<conclusion>** als beizubehaltende Tags in Ihrer Tagdatei festgelegt, dann wird WordSmith **<title>**, **<body>** oder **<conclusion>** in Ihre Daten aufnehmen, aber **<introduction>**, **<heading>**, **<threat>**, etc. ignorieren.

Tags beibehalten ist nur aktiv, wenn ein [Dateiname](#) angezeigt wird und Sie die Schaltflächen *Laden* oder *Entfernen* angeklickt haben. Wenn Sie *Laden* anklicken, wird angezeigt, welche Tags aus der Tagdatei eingelesen wurden.

Wenn Sie den Dateinamen in Ihren [Standardeinstellungen](#) (**wshell.ini**) festlegen und **autoload tagfile=YES** inkludieren, werden die Tags automatisch beim Starten von WordSmith geladen.

Entity-Referenzen zu anderen Zeichen konvertieren

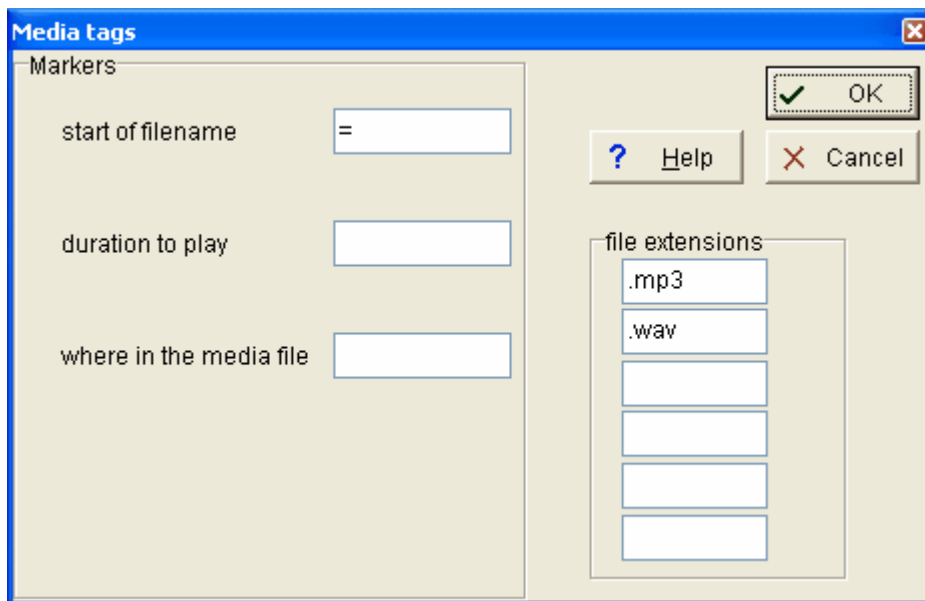
Wenn Sie [SGML oder HTML](#)-Dateien verwenden, möchten Sie möglicherweise Symbole und Sonderzeichen konvertieren. Zum Beispiel wird in SGML, XML und HTML **—** anstelle eines Gedankenstrichs verwendet. Dazu müssen Sie zuerst eine [Tagdatei](#) erstellen, die alle Strings enthält, die Sie konvertieren möchten. Danach wählen Sie unter *Einstellungen anpassen* | *Tags und Markup* | *Entity-Datei (zu übersetzende Entities)* Ihre Entity-Datei aus. WordSmith wird dann alle Entity-Referenzen in dieser Datei durch die entsprechenden Zeichen ersetzen.

Wenn Sie den Dateinamen in Ihren [Standardeinstellungen](#) (`wshe11.ini`) festlegen und **autoload tags to translate=YES** inkludieren, werden die Entities automatisch beim Starten von WordSmith geladen.

Siehe auch: [Tag-Übersicht](#), [Eine Tagdatei erstellen](#), [Nächstes Tags in Concord anzeigen](#), [Tag-Konkordanz erstellen](#), [Tagtypen](#), [Tags anzeigen](#), [Tags als Selektoren verwenden](#), [Standardeinstellungen speichern](#)

11.4 Multimedia-Tags

In diesem Screenshot kann man sehen, wie Multimedia-Tags definiert werden. Um einen Verweis wie z. B. `<a href=http://gandalf.hit.uib.no/c/1/32401-1.mp3>` im Quelltext zu finden,



genügt das Zeichen "=", um den Anfang des Dateinamens zu definieren. In diesem Fall muss nach dem "=" eine Webadresse stehen. Für einen Text, der Tags wie z. B. `<sound$$C:\mysounds\talk.wav>` enthält, müssen Sie mit \$\$ den Anfang des Dateinamens angeben. Die Dateierweiterungen definieren Dateitypen, die Ihr Computer wiedergeben kann. Für diese Funktion ist es natürlich erforderlich, dass Ihr Computer Video- bzw. Audioformate lesen und wiedergeben kann. Windows verwendet Dateierweiterungen für das Erkennen des Formats bzw. des entsprechenden Wiedergabemodus.

Wiedergabedauer und Wiedergabelänge

Wiedergabedauer und Anfangspunkt der Wiedergabe werden in Sekunden gemessen.

Ohne angegebenen Wert wird die gesamte Audio- bzw. Videodatei wiedergegeben.

In Ihren Textdateien sind möglicherweise die Positionen für Dauer und Anfangspunkt markiert, z. B. `<sound$$C:\mysounds\talk.wav start:150 play:5>`. Ist dies der Fall, sollten Sie die Wiedergabedauer mit **play:** und den Anfangspunkt mit **start:** festlegen.

Wenn es keine Markierungen für Dauer und Anfangspunkt gibt, wird die erste Zahl als Anfangspunkt interpretiert und die zweite als Wiedergabedauer, z. B. ein Tag wie dieses: `<sound$$C:\mysounds\talk.wav 15 5>` in Ihrer Textdatei bedeutet "Wiedergabe der Datei c:\mysounds\talk.wav ab Sekunde 15 für eine Dauer von 5 Sekunden".

Standardeinstellungen

Die Standardeinstellungen sind: .mp3- und .wav-Dateien wiedergeben. Wenn Sie dieses Formular ausgefüllt haben, [speichern Sie Ihre Standardeinstellungen](#) für die nächste Sitzung.

Siehe auch: [Audio und Video in Concord](#), [Tag-Übersicht](#), [Eine Tagdatei erstellen](#), [Tag-Handling](#), [Tag-Konkordanz erstellen](#),
[Nächstes Tags in Concord anzeigen](#), [Tags anzeigen](#), [Tagtypen](#)

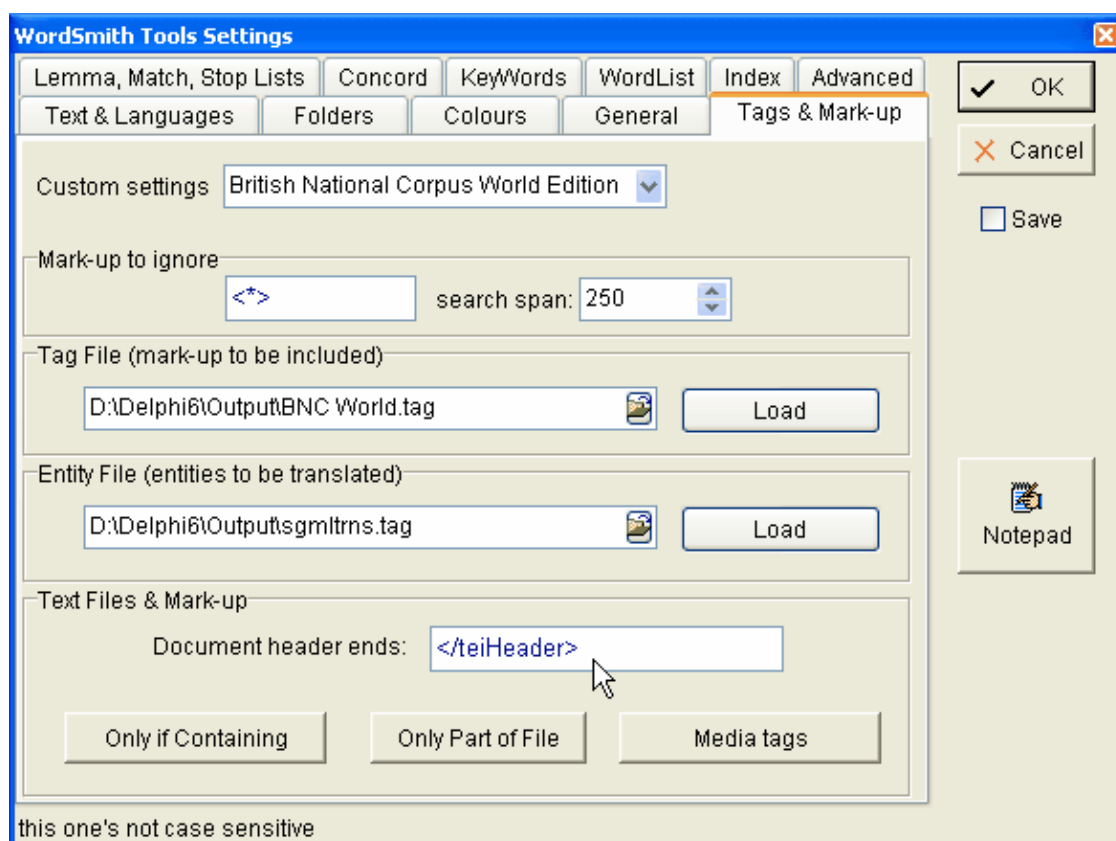
11.5 Tags als Selektoren

Standardeinstellungen

Die Standardeinstellungen sind: *Alle* Abschnitte in *allen* Texten einbeziehen, die unter [Texte auswählen](#) ausgewählt wurden, aber alle Tags mit spitzen Klammern ausschneiden.

Sie können mit **WordSmith** auch Tags dafür verwenden, einen Textabschnitt auszuwählen und den Rest zu ignorieren. Diese Funktion heißt "innerhalb von Texten auswählen". Sie können auch *zwischen* Texten wählen: das bedeutet, **WordSmith** "liest" den Anfang jedes Texts und prüft, ob er bestimmten Kriterien entspricht.

Diese Funktion finden Sie unter *Einstellungen | Einstellungen anpassen | Tags | Texte nach Inhalt filtern oder Nur Teil der Datei*.



Benutzerdefinierte Einstellungen

Es gibt in diesem Feld viele Möglichkeiten, die Ihnen bei Ihrer Auswahl in den unteren Feldern helfen. Wählen Sie den *British National Corpus World Edition* (wie in dem Screenshot), wird automatisch `</teiHeader>` in das untere Feld "Dokument-Header endet" eingefügt.

Mark-up ignorieren

Wenn Sie alle ungewollten Tags ausschneiden möchten, z. B. bei [HTML](#)-Dateien, geben Sie in *Mark-up ignorieren* z. B. `< >` oder `[ ]` oder `< >; [ ]` ein. Der "Suchbereich" gibt an, wie weit WordSmith nach einem Endtag wie z. B. `>` suchen soll, nach dem ein Anfangstag wie z. B. `<` gefunden wurde. (Grund dafür ist, dass diese Symbole z. B. auch in der Mathematik vorkommen.)

Tagdatei, Entity-Datei

Siehe [Eine Tagdatei erstellen](#).

Dokument-Header

Wenn Sie einfach nur den Dokument-Header ausschneiden möchten (jeder BNC-Text enthält am Anfang einen Header mit Copyright-Hinweisen), müssen Sie nur das entsprechende Tag festlegen, wie oben im `</teiHeader>`-Beispiel beschrieben. (Wenn Sie oben Benutzerdefinierte Einstellungen wählen, werden Sie automatisch die passenden Einstellungen erhalten.)

Für komplexere Suchen, können Sie dafür auch die oben gezeigten Schaltflächen [Texte nach Inhalt filtern](#) oder [Nur Teil der Datei](#) verwenden.

Reihenfolge der ausgewählten Optionen

Wenn Sie sowohl die Texte nach Inhalt filtern und nur Teile der Dateien suchen, wird WordSmith zuerst jede Textdatei darauf überprüfen, ob Sie Ihre Kriterien erfüllt, bevor Ihre Konkordanz, Wortliste, etc. erstellt wird. WordSmith wird

1. die [Dateien auswählen](#), die alle Wörter enthalten, die Sie festgelegt haben;
2. alle Abschnitte ausschneiden, die Sie unter "[Abschnitte entfernen](#)" festgelegt haben;
3. bei "[beizubehaltenden Abschnitten](#)" alles löschen, was nicht innerhalb dieser Abschnitte liegt;
4. den [Anfang jeder Zeile](#) entfernen, wenn erforderlich;
5. alle Entity-Referenzen verarbeiten, die Sie [übersetzen](#) möchten;
6. alle Tags [ignorieren](#), die nicht beibehalten werden (siehe Abschnitt "Mark-up ignorieren" in dem oberen Screenshot).

Siehe auch: [Tag-Übersicht](#), [Eine Tagdatei erstellen](#), [Tag-Handling](#), [Tag-Konkordanz erstellen](#), [Nächstes Tags in Concord anzeigen](#), [Tags anzeigen](#), [Tagtypen](#)

11.6 Texte nach Inhalt filtern

Wozu dient das?

Möglicherweise möchten Sie eine große Textsammlung nach Reden von Männern über 60, nach Werbematerial oder nach Unterrichtsdialogen filtern. Mit dieser Funktion durchsucht WordSmith jeden einzelnen Text nach bestimmten Wörtern oder Zeichen, sodass Sie am Schluss die richtigen Textdateien erhalten und alle irrelevanten Textdateien ignoriert werden.

Angenommen Sie haben eine große Textsammlung (z. B. den British National Corpus, BNC) und Sie wissen nicht mehr, welche der Dateien gesprochene Texte von Männern über 60 enthalten.

Da der BNC das Tag `stext` für gesprochenen Text verwendet, `sex=m` für männliche Sprecher, und `age=5` für Sprecher ab 60, können Sie mit WordSmith so Ihre Textsammlung filtern. WordSmith durchsucht dann den gesamten Inhalt jeder Textdatei (nicht nur Tags oder Überschriften, sondern den Inhalt der ersten [zwei Megabyte](#) der Datei) und überprüft, ob die Datei Ihren Anforderungen entspricht.

Sie können bis zu zwölf Tags festlegen, jedes bis zu einer Länge von 80 Zeichen. Dabei wird Groß- und Kleinschreibung berücksichtigt (d. h. Sie werden keine Treffer haben, wenn Sie z.B. aus Versehen `Age=5` eingeben)

Horizontal stehen die Optionen für Kombinationen, die mit "oder" verknüpft sind. Vertikal sind die Kombinationen mit "und" verknüpft. Die untere Gruppe steht für Kombinationen, die mit "nicht" verknüpft sind.

Nachdem die Textdateien verarbeitet wurden, können Sie in dem [Fenster "Textdatei auswählen"](#) die Dateien sehen, die Ihren Anforderungen entsprechen und die Liste für den späteren Gebrauch unter [Favoriten](#) speichern.

Beispiele:

Sie suchen Dateien, die sowohl das Wort **Katzen** als auch das Wort **Hunde** enthalten: Geben Sie **Katzen** in das erste Feld ein, und **Hunde** in das Feld darunter.

Sie suchen Texte, die das Wort **Blumen** enthalten, sowie entweder das Wort **Rosen** oder das Wort **Veilchen**: Geben Sie **Rosen** und **Veilchen** nebeneinander in ersten zwei Felder ein.

Geben Sie dann **Blumen** in der nächsten Zeile in dem Feld ganz links ein.

Sie suchen nach englischen Texten, in denen entweder das Wort **book** oder das Wort **hotel** vorkommen, aber weder das Wort **publish** noch das Wort **Booker Prize** vorkommen:

Geben Sie dazu in dem ersten Feld **book** ein, in dem Feld rechts daneben **hotel**, und in den ersten zwei Feldern in der untersten Zeile **publish*** und **Booker*** ein.

Siehe auch: [Tags als Selektoren](#), [Innerhalb von Texten auswählen](#), [Textdateien mit Text Converter filtern](#).

11.7 innerhalb von Texten auswählen

Anfang jeder Zeile/jedes Absatzes entfernen

In einigen Korpora (z. B. LOB) wird jede Zeile am Anfang mit einer fixen Anzahl von Kodierungen markiert, die Informationen bezüglich der Zeile enthalten. Mit dieser Funktion möchten Sie diese Kodierungen entfernen (also nach jeder Zeilenschaltung). Wählen Sie die Anzahl der Zeichen aus, die entfernt werden sollen (max. = 100; Standard = 0) Verwenden Sie - **1**, wenn Sie alles bis zum ersten alphabetischen Buchstaben am Anfang jeder Zeile entfernen möchten, und - **2** um alles bis zum ersten Tabstopp zu entfernen.

Abschnitte ausschneiden

Wenn Sie Textdateien mit Headern in [SGML, XML oder HTML](#) verwenden (z. B. der British National Corpus), können Sie die Header einfach aus Ihren Wortlisten, Konkordanzen, etc. ausschneiden, wie im Beispiel [Dokument-Header](#) beschrieben wird.

Für eine komplexere Auswahl können Sie hier festlegen, was ausgeschnitten werden soll, und zwar von welcher Stelle (z. B. **<HEAD>**) bis zu welcher Stelle (z. B. **</HEAD>**). Sie können bis zu drei verschiedene, getrennte Abschnitte zum Ausschneiden auswählen (**<HEAD>** bis **</HEAD>** oder **<BODY>** bis **</BODY>**). Mit dieser Funktion schneiden Sie alle Abschnitte mit diesen Markierungen innerhalb des gesamten Texts aus.

Abschnitte beibehalten (Kontexte)

Sie möchten einen Abschnitt eines Texts auswählen und den Rest entfernen. Geben Sie ein Tag an, das den gewünschten Beginn des Abschnitts markiert, und ein Tag, mit dem das Ende des Abschnitts markiert wird, z. B. **<Intro>** bis **<Body>**

(hiermit werden nur die Einleitungen der Texte analysiert), oder **Mary:** bis **Peter:** (hiermit werden alle Beiträge von Mary analysiert und der Rest ignoriert).

Natürlich müssen Sie sich vergewissern, dass die Abschnitte mit im Text einzigartigen Symbolen definiert sind, wie z. B. **<** oder **>**. Zum Beispiel im Falle **Mary:** und **Peter:** müssen Sie sicherstellen, dass bei jedem Beitrag von Mary direkt nach ihrem Namen ein Doppelpunkt steht, und dass alle Beiträge von Mary direkt vor **Peter:** enden. Bei dieser

Funktion werden Groß- und Kleinschreibung beachtet (**MARY**: würde also nicht gefunden werden).

Wenn Sie in dieser Funktion **<H1>** bis **</H1>** angeben, werden Sie in einem [HTML](#)-Text alle Hauptüberschriften erhalten (egal wie viele), aber ansonsten nichts.

Sie können zwei verschiedene Abschnitte beibehalten, z. B. **<Intro>** bis **</Intro>** für die Einleitung und **<Conclusion>** bis **</Conclusion>** für den Schlussteil. Die Endmarkierung muss nicht denselben Text wie die Anfangsmarkierung beinhalten -- Sie können z. B. **<INTRO>** bis **</BODY>** beibehalten, und dadurch den Schlussteil ausschneiden, falls dieser nach **</BODY>** kommt.

Siehe auch: [Tags als Selektoren](#), [Texte nach Inhalt filtern](#).

11.8 eine Tagdatei erstellen

Tag-Syntax

Bei jedem Tag werden Groß- und Kleinschreibung beachtet.

Tags beginnen üblicherweise mit **<** und enden mit **>**, aber prinzipiell kann jedes Symbol das erste und letzte Zeichen eines Tags sein.

Sie können die folgenden Zeichen verwenden:

- * bedeutet eine beliebige Zeichenfolge;
- ? bedeutet ein beliebiges Zeichen;
- # bedeutet eine beliebige Ziffer.

Verwenden Sie keine eckigen Klammern ("**[**" und "**]**"), um Kommentare in Ihre Tagdatei einzufügen, da **[** und **]** als mögliche Tagsymbole sehr nützlich sind. Sie können **#** als Platzhalter für eine Zahl verwenden (z. B. **<h#>** findet **<h5>**, **<h1>**, etc.). Das Fragezeichen steht für ein beliebiges Zeichen (**<?>** findet **<s>**, **<p>**, etc.) und das Sternchen steht für eine beliebige Anzahl beliebiger Zeichen (z. B. **<u*>** findet **<u who=Fred>**, **<u who=Mariana>**, etc.). Sie können auch -- gleich wie für [Stopplisten](#) -- eine Tagdatei vorbereiten.

Verwenden Sie den **Windows Editor** oder einen anderen Standardtext-Editor, um eine neue **Tag-Datei (*.tag)** zu erstellen. Schreiben Sie in jede Zeile einen Eintrag.

Es kann eine beliebige Anzahl an vordefinierten Tags gespeichert werden. Je mehr Sie jedoch verwenden, umso mehr Daten muss WordSmith verarbeiten, und dazu benötigt das Programm mehr Zeit und Arbeitsspeicher...

Tags beibehalten (Tags, die inkludiert werden)

Eine Tagdatei für beizubehaltende Tags enthält eine einfache Liste aller Tags, die in den Ergebnissen angezeigt werden sollen. Sie erhalten eine Taglisten-Vorlage für BNC-Tags (z. B. **bnc world.tag**) mit Ihren Installationsdateien (in Ihrem **\wsmith4**-Verzeichnis): Sie können eine neue Tagdatei erstellen, indem Sie eine dieser Dateien öffnen, verändern und unter einem anderen Dateinamen speichern.

Tags werden standardmäßig in einer Standard-Tagfarbe (grau) dargestellt, aber Sie können Hintergrund ("**background**") und Schriftfarbe ("**foreground**") der Tags verändern, indem Sie festlegen:

/colour="foreground on background"

z. B. **<noun> /colour="yellow on red"**

Verfügbare Farben:

'Black', 'White', 'Cream',
'Red', 'Maroon',
'Yellow',

'Navy','Blue','Light Blue','Sky Blue',
'Green','Olive','Dollar Green','Grey-Green','Lime',
'Purple','Light Purple',
'Grey','Silver','Light Grey','Dark Grey','Medium Grey'.

Bei den Farbbezeichnungen wird Groß- und Kleinschreibung nicht beachtet (obwohl sie Tags sind). Beachten Sie bitte bei den Wörtern "grey" und "colour" die britische Schreibweise.

Wenn Sie nach dem Tag "/play media" eingeben und das Tag wird in Ihren Textdateien gefunden, haben Sie die Möglichkeit, die [Audio- oder Videodatei](#) abzuspielen. Zum Beispiel mit dem festgelegten Tag

<sound *> /colour="blue on yellow" /play media

und dem folgenden Tag im Text

<sound c:\windows\Mozarts Zauberflöte.wav>

oder

<sound http://www.political_speeches.com/Mao_Ze_Dung.mp3>

können Sie auswählen, ob [die .wav-Datei oder die .mp3-Datei abgespielt](#) werden soll.

Zum Schluss können Sie eine Beschreibung hinzufügen, mit /description "Beschreibung":

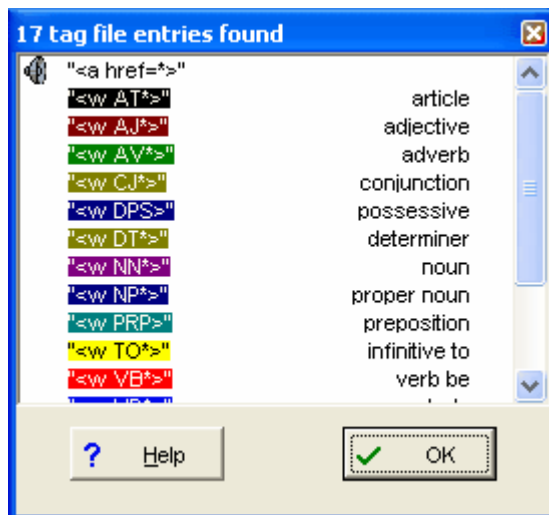
<w NN*> /description "noun" /colour="Cream on Purple"

<ABSTRACT> /description "Abschnitt"

<INTRODUCTION> /description "Abschnitt"

<SECTION 1> /description "Abschnitt"

Hier sehen Sie ein Beispiel davon, was angezeigt wird, wenn Sie eine Tagdatei auswählen und auf "Laden" klicken. Das erste Tag ist ein "play media"-Tag, wie man auch an dem Symbol erkennen kann. Für "noun" sehen Sie auch "cream on purple".



Entity-Datei (zu übersetzende Entities)

Eine Tagdatei mit zu übersetzenden Entity-Referenzen verwendet die folgende Syntax: Entity-Referenz + Leerstelle + Übersetzung. Zum Beispiel:

É É

é é

Eine Vorlage für eine solche Datei ist bei Ihren Installationsdateien inkludiert (**\wsmith4\sgmltrns.tag**): Sie können eine neue Datei erstellen, indem Sie diese Dateien öffnen, verändern und unter einem anderen Dateinamen speichern.

Siehe auch: [Tag-Übersicht](#), [Tag-Handling](#), [Nächstes Tags in Concord anzeigen](#), [Tag-Konkordanz erstellen](#), [Tagtypen](#), [Tags anzeigen](#), [Tags als Selektoren verwenden](#)

11.9 Anfang und Ende von Textsegmenten

WordSmith versucht vier Arten von Textsegmenten zu erkennen: Sätze, Absätze, Überschriften und Abschnitte. Bei der Verarbeitung werden Groß- und Kleinschreibung beachtet. Sie können in den Texten die Tags **<Enter>** und **<Tab>** verwenden, um das Ende eines Absatzes oder einen Tabstopp zu markieren. Für Satzenden ist **auto** eine weitere Option.

Sätze

Zum Beispiel **<s>** kann den Anfang eines Satzes markieren und **</s>** das Ende. Wenn Sie die Auswahl bei **auto** belassen, werden Satzenden anhand der Interpunktionszeichen **!/?** gefolgt von einem Großbuchstaben bestimmt.

Absätze

Zum Beispiel **<p *>** oder **<p>** markiert den Anfang eines Absatzes und **</p>** das Ende.

Überschriften

Zum Beispiel **<head>** markiert den Anfang einer Überschrift und **</head>** das Ende. Beachten Sie bitte, dass der British National Corpus Sätze innerhalb von Überschriften markiert; z. B.

<head>

<s n="2"><w NN1>Introduction

</head>

im Text **HXL**. Es ist erscheint ungewöhnlich, das einzelne Wort **Introduction** als Satz zu werten, weshalb WordSmith innerhalb von Überschriften keine Satz-Tags verwendet.

Abschnitte

Zum Beispiel **<section *>** markiert den Anfang eines Abschnitts und **</section>** das Ende.

Jedes dieser Tags wird gezählt, vorzugsweise wenn das Endtag, wie z. B. **</s>**, **</p>**, etc., vorhanden ist. Wenn im gesamten Text keine schließenden **</p>**-Tags vorkommen, werden Absätze jedes Mal gezählt, wenn das öffnende Tag gefunden wird.

Siehe auch: [Tag-Übersicht](#), [Tag-Handling](#), [Nächstes Tags in Concord anzeigen](#), [Tag-Konkordanz erstellen](#), [Tagtypen](#), [Tags anzeigen](#), [Tags als Selektoren verwenden](#)

Referenz

Section



XII

12 Referenz

12.1 32-Bit-Version

Diese Version von **Oxford WordSmith Tools** wurde vollkommen neu geschrieben (verglichen mit den früheren 16-Bit-Versionen), was viele Veränderungen "unter der Oberfläche" mit sich bringt. Einige Veränderungen, die Sie bemerken werden:

- lange [Dateinamen](#)
- verbesserter Umgang mit [Tags und Entities](#) einschließlich [Tag-Konkordanz](#)
- frühere Dateien können weiter benützt werden, sie sollten jedoch im 32-Bit-Format neu gespeichert werden. Wenn eine Datei noch im alten Format ist, haben Sie die Möglichkeit diese Datei zu "aktualisieren".
- [Bearbeitung von Zip-Dateien](#)
- verbesserter Export von Daten in Microsoft Word und [Excel](#)
- Bearbeitung von Unicode-Text, wodurch mehr [Sprachen](#) verarbeitet werden können
- Möglichkeit, beim Einlesen [Daten zu verändern](#), z. B. für sprachspezifische Lemmatisierung
- keine Begrenzung auf 16.000 Zeilen. (Die theoretische Obergrenze für Listen liegt bei über 134 Millionen Zeilen.)

Siehe auch: [Kontaktadressen](#).

12.2 Bibliografie

- 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", *Language*, V. 67, No. 4, pp. 763-89.

12.3 Bindestriche

Der Bindestrich verbindet zwei zusammengehörende Wörter miteinander. Die Bezeichnung "Preis-Leistungs-Verhältnis" kann abhängig von den Texteinstellungen als ein Wort oder drei Wörter betrachtet werden.

12.4 Bugs

Alle Computerprogramme haben so genannte Bugs (= Programmfehler). Möglicherweise haben Sie schon einmal eine Meldung über eine "Allgemeine Schutzverletzung" erhalten, wenn Sie teure Bildbearbeitungs- oder Textverarbeitungsprogramme verwenden. Wenn Sie diese Meldung in einem WordSmith-Programm erhalten, erscheint ein Bericht, in dem "allgemeine" Informationen über Ihren PC und "Details" über den aufgetretenen Fehler angezeigt werden. Diese Information hilft bei der Problemlösung und wird in einer kleinen Textdatei mit dem Namen **wshell.elf**, **concord.elf**, **wordlist.elf**, etc. gespeichert. Wenn Sie das Programm schließen, werden Sie gefragt, ob Sie diese Datei per E-mail an mich schicken möchten.

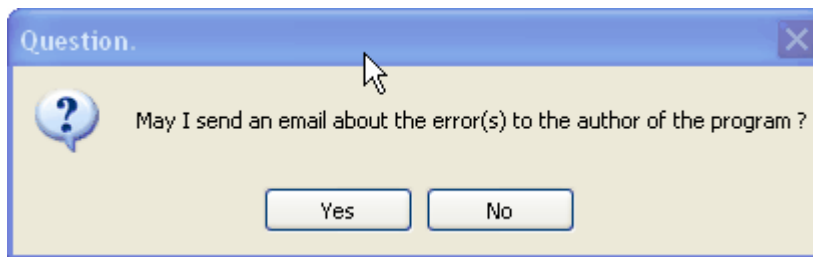
Wenn so etwas passiert, sehen Sie zuerst einen Dialog, der so ähnlich wie dieser aussieht:



Wenn Sie auf OK klicken, müssen Sie möglicherweise das Programm beenden, es sei denn, WordSmith funktioniert trotz des Problems weiterhin.

Normalerweise wird das betroffene Programm mit dem Problem fertig bzw. Sie können, ohne dass der WordSmith Tools [Controller](#) geschlossen werden muss, das Programm sofort neu öffnen, die [gespeicherten Ergebnisse](#) von der Festplatte wiederherstellen und mit Ihrer Textanalyse fortfahren. Wenn das nicht funktioniert, versuchen Sie Oxford WordSmith Tools vollständig zu beenden, oder Sie beenden Windows und starten den Computer neu.

Wenn Sie WordSmith beenden, sollte die folgende Meldung erscheinen:



Wenn Sie *Ja* wählen, wird Ihr E-Mail-Programm gestartet und in der E-Mail erscheint ein Text mit Informationen, die mir helfen, diesen Programmfehler zu beheben. **Die E-Mail wird nur dann verschickt, wenn Sie in Ihrem E-Mail-Programm auf Senden klicken. Sie wird nur an mich verschickt, und ich leite die E-Mail an niemanden weiter. Wenn Sie Bedenken bezüglich des Inhalts haben, lesen Sie sich den Text einfach durch. Er enthält Informationen über das Betriebssystem, den verfügbaren Arbeitsspeicher bzw. Festplattenspeicherplatz, die WordSmith-Version sowie einige technische Details über Vorgänge, die stattgefunden haben, während das Problem aufgetreten ist.**

Fehlermeldungen

Diese Fehlermeldungen warnen vor Problemen, die während eines Vorgangs auftreten können, z. B. wenn nicht genug Speicherplatz auf Ihrer Festplatte vorhanden ist, oder Sie einen ungültigen [Dateinamen](#) oder eine Zahl mit Komma eingeben.

Siehe auch: [Protokollierung](#), [Fehlerbehebung](#).

12.5 Danksagung

Oxford WordSmith Tools hat sich über die Jahre weiterentwickelt. Ursprünglich entstand jedes Tool aus dem Umstand heraus, dass ich als angewandter Linguist ein Tool für einen speziellen Arbeitsauftrag benötigt habe. Frühe Versionen wurden für DOS geschrieben, dann tauchte Windows™ auf.

Einzig mit dem Tool **Concord** war es ein bisschen anders. Concord entwickelte sich aus *MicroConcord* heraus, das Tim Johns und ich für DOS geschrieben hatten und 1993 von der Oxford University Press veröffentlicht wurde. **Concord** hat in dieser Windows-Version nun viele zusätzliche Funktionen und der gesamte Code musste neu geschrieben werden, aber die grundlegenden Funktionen des Entwurfs waren bereits in *MicroConcord* vorhanden.

Die erste veröffentlichte Version wurde größtenteils in Borland™ Pascal geschrieben, nur die zeitkritischen Abschnitte in Assembler. Daraufhin wurden die Programme in Delphi™ 16-Bit konvertiert; dies ist eine komplette 32-Bit-Version in Delphi 6, weiterhin mit den zeitkritischen Abschnitten in Assembler.

Ich danke allen Generationen von Studenten und Kollegen der [Englischen Abteilung](#) an der University of Liverpool und dem MA-Programm für Angewandte Linguistik an der Katholischen Universität von São Paulo für ihr Feedback bezüglich WordSmith (einschließlich der Bugs!), und Vorschläge für Funktionen, die dieses Programm beinhalten sollte. Forscher in vielen anderen Ländern haben mir als Alpha- und Beta-Tester sehr geholfen, und ich danke ihnen für ihre Geduld und ihr Feedback. Außerdem danke ich Nell Scott und allen anderen Familienmitgliedern, die mir mit ihrer wertvollen Unterstützung, sowie mit Feedback und Vorschlägen zur Seite gestanden sind.

Mike Scott

Zögern Sie nicht, mich unter meiner [Kontaktadresse](#) zu kontaktieren, wenn Sie weitere Verbesserungsvorschläge für **WordSmith Tools** haben.

12.6 Datumsformat

Date Format

Japanese date format year_month_day_hour_minute. At least it is logical, going from larger to smaller. Why aren't URLs organised in a logical order too?

12.7 Definitionen

12.7.1 Definitionen

Wörter

Ein Wort ist definiert als *Abfolge gültiger Zeichen zwischen zwei Worttrennzeichen*. Gültige Zeichen sind alle Buchstaben zwischen A und Z, alle Akzentzeichen, die in dem aktuellen [Zeichensatz](#) enthalten sind, sowie alle benutzerdefinierten Zeichen, die in einem Wort vorkommen können (z. B. ein Apostroph oder [Bindestrich](#)).

Es gibt für Wörter keine Höchstlänge, wenn ein Wort aber in einer Wortliste gespeichert werden soll, können Sie eine bevorzugte Höchstlänge (maximal 50 Zeichen) festlegen. Alle Wörter, die diese Länge überschreiten, werden an dieser Stelle mit einem + gekennzeichnet. Sie können in den [Texteigenschaften](#) festlegen, ob Sie Wörter mit Ziffern einbeziehen oder ausschließen möchten.

Cluster

Ein Cluster ist eine Gruppe aufeinander folgender Wörter in einem Text. Der Term *Phrase* wird hier nicht verwendet, da er in der Linguistik eine grammatische Beziehung zwischen den Wörtern impliziert. Bei der [WordList-Cluster-Verarbeitung](#) oder der [Concord-Cluster-Verarbeitung](#) muss diese grammatische Beziehung nicht vorhanden sein, obwohl Phrasen oder Idiome sehr häufig Cluster sind. Siehe auch: [allgemeine Informationen zu Clustern](#).

Sätze

Ein Satz ist definiert als *Punkt, Frage- oder Ausrufezeichen (.?!) direkt gefolgt von einem oder mehreren Worttrennzeichen und einem Großbuchstaben in der aktuellen Sprache bzw. einer Ziffer oder einem Währungszeichen*. (Für mehr Informationen siehe [Anfang und Ende von Textsegmenten](#) oder [Technische Informationen zu Viewer](#).)

Absätze

Absätze sind benutzerdefiniert. Für weitere Informationen siehe [Anfang und Ende von Textsegmenten](#).

Überschriften

Überschriften sind auch benutzerdefiniert, siehe [Anfang und Ende von Textsegmenten](#).

Siehe auch: [Texteigenschaften einstellen](#), [Gewichtung](#), [Gewichtetes Schlüsselwort](#), [Assoziierte Wörter](#)

12.7.2 Worttrennzeichen

Normalerweise wird angenommen, dass ein Wort vom nächsten Wort durch Leerstellen an Anfang und Ende getrennt wird. Aber **Oxford WordSmith Tools** inkludiert bei Worttrennzeichen auch bestimmte Standardcodes, die von den meisten Textverarbeitungsprogrammen verwendet werden: Seitenvorschubcode (12), Tabstopps (9), Wagenrücklauf (13) und Zeilenvorschub (10), Textende (26). Außerdem können [Bindestriche](#) als Worttrennzeichen erkannt werden, und damit Wörter wie *self-access* zwei Wörter gezählt werden.

Beachten Sie, dass in Chinesisch und Japanisch Wörter nicht auf diese Weise getrennt werden, also werden alle WordSmith-Funktionen, die Worttrennung erfordern, nicht funktionieren, es sei

denn der Text wurde vorher mit Worttrennzeichen markiert.

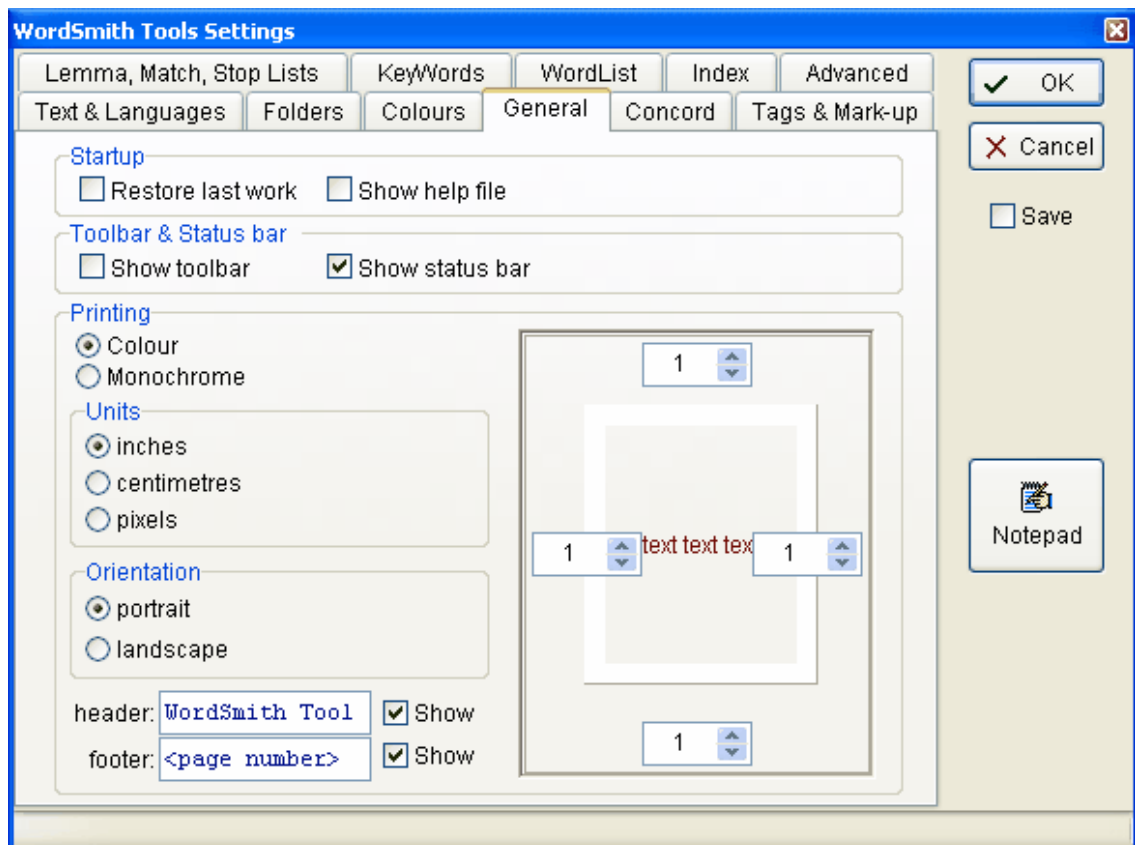
12.8 Demoversion

The demonstration version of **Oxford WordSmith Tools** offers *all* the facilities of the complete suite, except that any screen which shows a list (of words in a wordlist, or concordance lines, etc.) is limited to a small number of lines which can be shown or printed. (If you save data, all of it will be saved; it's just that you can't see it all in the demo version.)

See also: [Installing](#), [Version Information](#), [Contact Addresses](#).

12.9 Druckereinstellungen

Die Druckereinstellungen finden Sie unter *Einstellungen anpassen* | *Allgemein* im Oxford WordSmith Tools [Controller](#).



Programmstart

"Letzte Arbeit wiederherstellen" wird beim Start von WordSmith Ihre letzte Wortliste, Konkordanz oder Schlüsselwortliste wiederherstellen.

Symbolleiste & Statusleiste

In jedem Tool gibt es am unteren Rand des Fensters eine Statusleiste und oben eine Symbolleiste mit Schaltflächen. Standardmäßig wird die Symbolleiste ausgeblendet um zu

vermeiden, dass der Bildschirm unübersichtlich wird.

Drucken

Wenn Sie die Druckeinstellungen auf "schwarz-weiß" festlegen, werden alle Daten, die [farbig](#) markiert sind, kursiv oder fett gedruckt. Bei einem Farbdrucker werden die Ergebnisse farbig ausgedruckt, oder in Graustufen, wenn Ihr Drucker Graustufen drucken kann. Ändern können Sie auch die Einstellungen für Einheiten und Orientierung (Hochformat  oder Querformat ) , sowie Seitenränder, Standardkopfzeile und Standardfußzeile. Druckereinstellungen können Sie in der [Druckvorschau](#) verändern.

Bestätigung

Sie können Oxford WordSmith Tools so einstellen, dass ein Druckauftrag in der [Standarddatei \(wshell.ini\)](#) bestätigt wird. Wenn diese Datei die Zeile confirm printing=YES enthält, wird Ihnen bei jedem Druckauftrag mitgeteilt, welche Zeilen der aktuellen Konkordanz oder List gedruckt wurden.

Siehe auch: [Drucken](#)

12.10 Einschränkungen

Mit den Programmen in **Oxford WordSmith Tools** können Sie praktisch so viele Texte bearbeiten wie Sie wollen. Die Programme können Texte von einer CD-ROM lesen; das ermöglicht den Zugriff auf Korpora mit mehreren Millionen Wörtern. In der Praxis erreicht man die Grenzen aufgrund von a) [Speicherplatz](#) und b) Geduld.

Sie können gleichzeitig so viele Kopien von jedem Tool verwenden, wie Sie wollen. Jedes Tool ermöglicht es Ihnen, mit einer Datengruppe zu arbeiten.

[Zu ignorierende Tags](#) oder Tags mit einem Sternchen können bis zu 1.000 Zeichen lang sein. Wenn Sie nach Tags suchen, um festzulegen, ob Ihre [Textdateien gewissen Anforderungen entsprechen](#), werden nur die 2 ersten Megabyte des Texts geprüft. In [ASCII](#) entspricht das zwei Millionen Zeichen, in [Unicode](#) nur einer Million Zeichen.

Tipp

Drücken Sie F9, um den Dialog "Über" anzuzeigen -- darin wird das Versionsdatum und die Auslastung des [Arbeitsspeichers](#) angezeigt. Wenn der Arbeitsspeicher knapp wird, versuchen Sie a) einige Programme zu schließen, b) WordSmith Tools zu beenden und neu zu starten.

Siehe auch: [Spezifische Einschränkungen jedes Tools](#)

12.11 einzelne Wörter vs. Cluster

Wozu dient das?

Cluster sind Wörter, die wiederholt gemeinsam aufeinander folgend auftreten. Ihre Beziehung ist enger als bei Kollokationen, eher wie bei Komposita, Wortgruppen oder Phrasen. (Hier werden sie *Cluster* genannt, weil Gruppen und Phrasen bereits grammatikalisch definiert sind. Wenn solche Wörter mit einem Softwareprogramm gefunden werden, sind sie nicht automatisch fixierte, aus mehreren Wörtern bestehende *Einheiten*.) [Biber](#) nennt sie "lexical bundles" (dt. 'lexikalische Bündel').

Sprache besteht aus Phrasen und Text. Dabei es nicht hilfreich, Sprache als eine Frage der Auswahl eines Worts für eine grammatikalische Lücke zu betrachten, wie strukturalistische Theorien behaupten. Wörter stehen miteinander in Kontakt: Im Extremfall der Idiome sind sie eng aneinander gebunden, aber alle Wörter tendieren dazu, sich mit anderen Wörtern zu gruppieren. Diese Beziehungen sind möglicherweise grammatikalische Kollokationen (z. B. die Beziehung zwischen **abhängig** und **von**), [Kollokationen](#) oder semantische Prosodie (die

Tendenz, dass das Wort **verursachen** meist mit negativen Wörtern wie **Problem**, **Unfall**, etc. auftritt).

Oxford WordSmith Tools bietet zwei Möglichkeiten Wortcluster zu identifizieren, und zwar in [WordList](#) und in [Concord](#). Dabei werden verschiedene Methoden angewandt. In Concord werden nur Konkordanzzeilen verarbeitet, in WordList ganze Texte.

Wie funktioniert das in Concord?

Angenommen, Ihr Text beginnt etwa folgendermaßen:

Once upon a time, there was a beautiful princess. She snored. But the prince didn't.

Wenn Sie 2-Wort-Cluster gewählt haben, wird der Text wie folgt aufgeteilt:

Once upon

upon a

a time

(nicht "time there" aufgrund des Kommas!)

there was (etc.)

Wenn Sie 3-Wort-Cluster gewählt haben:

Once upon a

upon a time

there was a

was a beautiful

a beautiful princess

But the prince

the prince didn't

(etc.)

Also wird jeder N-Wort-Cluster gespeichert, wenn er eine Länge von N Wörtern erreicht. Cluster können nicht *über Interpunktionsgrenzen* (;,!,?) *hinausgehen*. (Interpunktionszeichen markieren Satzgrenzen, und es erscheint sinnvoll, dass Cluster nicht über Satzgrenzen hinausgehen dürfen.)

12.12 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.

12.13 Geschwindigkeit

Eine Wortliste mit 4,2 Millionen Wörtern zu erstellen, dauerte mit einem PC Baujahr 1993, 486/33 MHz mit 8MB [RAM](#) etwa 20 Minuten. Der Sortiervorgang am Ende der Verarbeitung dauerte etwa 30 Sekunden. Ein Pentium 200MHz mit 64MB RAM verarbeitete über 1,7 Millionen Wörter

pro Minute. Auf einem Pentium 100MHz mit 32MB RAM dauerte dieser Vorgang etwa dreieinhalb Minuten, mehr als eine Million Wörter pro Minute.

Die Erstellung einer Konkordanz auf demselben Pentium 100, unter Verwendung von 55MB Textdateien mit 9,3 Millionen Wörtern und einem Quad-Speed CD-ROM-Laufwerk, ergab folgende Ergebnisse:

Suchwort	Quelle	Geschwindigkeit
quickly	CD-ROM	6 Millionen Wörter pro Minute
quickly	Festplatte	12 Millionen WpM
the	CD-ROM	900.000 WpM
the	Festplatte	1 Million WpM
thez	CD-ROM	6 Millionen WpM
thez	Festplatte	16 Millionen WpM

Tests für eine Sammlung von Textdateien zwischen 4KB und 20KB und dem Suchwort *quickly* ergaben eine Geschwindigkeit von 2 Millionen Wörtern pro Minute, die bis auf 4 Millionen stieg, je größer die Dateien waren. Die Erstellung einer Wortliste basierend auf derselben Sammlung von Textdateien ergab eine Durchschnittsgeschwindigkeit von 800.000 WpM. Bei der 55MB Textdatei lag die Geschwindigkeit etwa bei 1,35 Millionen WpM.

Diese Daten lassen vermuten, dass folgende Faktoren die Konkordanzerstellung verlangsamen (in dieser Reihenfolge): Seltenheit des Suchworts (*the* war sehr viel langsamer als *quickly* oder das nicht existierende Wort *thez*), Textdateigröße (sehr kleine Dateien mit nur etwa 500 Wörtern (3KB) werden ca. dreimal langsamer verarbeitet als große Dateien) und Geschwindigkeit des Datenträgers (die Festplatte mit 12ms ist doppelt so schnell wie das veraltete Quad-Speed CD-ROM-Laufwerk). Wenn Concord ein Wort findet, muss es die Konkordanzlinie und die Kollokatoren speichern und anzeigen (damit Sie entscheiden können, ob der Vorgang [unterbrochen](#) werden soll, wenn die Ergebnisse nicht zufriedenstellend sind oder Sie bereits genügend Ergebnisse haben). Dadurch wird die Verarbeitung beträchtlich verlangsamt. Zweitens, beim Einlesen einer Datei wird das Dateiverwaltungssystem des Computers aktiviert, das für das Laden der Dateien einige Zeit benötigt verglichen mit der Zeit, in der Concord die Datei durchsucht. Drittens, die Geschwindigkeiten der Datenträger sind sehr unterschiedlich, wobei Disketten von allen Datenträgern am langsamsten sind.

Wenn Ihnen die Verarbeitung besonders langsam erscheint, schließen Sie so viele Programme wie möglich und starten Sie **Oxford WordSmith Tools** erneut. Oder installieren Sie mehr RAM. Informieren Sie sich darüber, wie Sie die Effizienz von Windows erhöhen können (virtueller Speicher, Datenträgerscaches, etc.) und arbeiten Sie mit einer großen, schnellen Festplatte. Sie können während der Berechnung in WordSmith andere Programme ausführen, diese benötigen jedoch viel Zeit des Prozessors. Ego-Shooter-Spiele werden möglicherweise nicht mehr einwandfrei ausgeführt, das [Drucken](#) eines Dokuments sollte jedoch kein Problem darstellen.

12.14 Handbuch für WordSmith Tools

Diese Hilfedatei bekommen Sie in der Form eines Handbuchs, wenn Sie die Software [installieren](#). Die Datei (**wshell.pdf**) ist im Adobe Acrobat(tm)-Format. Diese Hilfedatei enthält ein Inhaltsverzeichnis und einen sehr detaillierten Index (der anhand von **WordList** und **KeyWords** erstellt worden ist). Die meisten Menschen finden es einfacher, Text auf Papier zu lesen als in Hilfedateien.

Um Screenshots von **WordSmith** in Verwendung zu sehen: siehe [Kontaktadressen](#).

12.15 HTML, SGML und XML

Diese Formate dienen dem Informationsaustausch. HTML, Hypertext Markup Language, ist das bekannteste dieser Formate, und wird zur Verbreitung von Texten über das Internet verwendet. SGML, Standard Generalized Markup Language, wird von Verlegern und dem British National Corpus (BNC) verwendet. XML, Extensible Markup Language, ist eine Kombination aus HTML und SGML.

All diese Standardformate verwenden [Standardtext](#) mit zusätzlichen Markierungen (= Tags) meist in spitzen Klammern, wie z. B. <h1> und </h1>. Durch das Einfügen dieser Tags fügt man dem Text zusätzliche Informationen hinzu:

- 1 einen Header (<head>), der Informationen zu Autor und Ausgabe enthält
 - 2 Formatierung des Texts (z. B. fett, <i>kursiv</i>)
 - 3 Hervorhebung der wichtigen Textabschnitte (<h1> markiert eine Überschrift, <body> markiert den Hauptteil des Texts)
 - 4 wie Symbole und Sonderzeichen angezeigt werden sollen (ä entspricht ä)
- Siehe auch: [Tag-Übersicht](#)

12.16 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](#)

12.17 Internet

Verfügbar unter Einstellungen anpassen | Allgemein im WordSmith Controller. Geben Sie den vollständigen Dateinamen Ihres Internetbrowsers an. WordSmith kann so die in der Software inkludierte Hilfe in Form von .htm-Dateien und Screenshots aufrufen.

12.18 Internet

Verfügbar unter Einstellungen anpassen | Allgemein im WordSmith Controller. Geben Sie den vollständigen Dateinamen Ihres Internetbrowsers an. WordSmith kann so die in der Software inkludierte Hilfe in Form von .htm-Dateien und Screenshots aufrufen.

12.19 Internet

Verfügbar unter Einstellungen anpassen | Allgemein im WordSmith Controller. Geben Sie den vollständigen Dateinamen Ihres Internetbrowsers an. WordSmith kann so die in der Software inkludierte Hilfe in Form von .htm-Dateien und Screenshots aufrufen.

12.20 Kontaktadressen

Downloads

Sie erhalten die aktuelle Version auf [meiner Website](#) (nur auf Englisch verfügbar). Dort finden Sie auch einige kostenlose Extra-Downloads (Programme, Wortlisten, etc.) und Links zu anderen Websites, auf denen kostenlose Textkorpora zu Verfügung gestellt werden. Die offizielle Version der [Oxford University Press](#) wird seltener aktualisiert.

Screenshots

Besuchen Sie die [WordSmith 3 Screenshots](#)-Seite oder <http://www.lexically.net/wordsmith/version4/screenshots/index.html> (jeweils nur auf Englisch verfügbar) um herauszufinden, was man mit Oxford WordSmith Tools alles machen kann. Möglicherweise finden Sie dort wertvolle Anregungen für Ihre Arbeit und bekommen vielleicht auch einen Eindruck der Einschränkungen von WordSmith!

Kaufen

Für Details über Lieferanten besuchen Sie bitte <http://www.lexically.net/wordsmith/purchasing>.

[htm.](#)

Anregungen und Beschwerden

Wenn Sie nicht die offizielle OUP-Version besitzen, sondern eine Version von meiner Website haben, schreiben Sie bitte nicht an OUP sondern an mich persönlich (Mike.Scott@liv.ac.uk). Bitte schicken Sie mir eine möglichst genaue Beschreibung des aufgetretenen Problems sowie Ihrer PC-Ausstattung. Bitte fügen Sie keine Dateianlagen an, die größer als 200KB sind. Ich werde mich bemühen, Ihr Problem zu lösen.

12.21 lange Dateinamen

Diese **WordSmith**-Version verarbeitet lange Dateinamen einwandfrei.

12.22 letzte Datei wiederherstellen

By default, the last word list, concordance or key words listing that you saved or retrieved will be automatically restored on entry to **Oxford WordSmith Tools**. If the last Tool used is **Concord**, a list of your 10 most recent search-words will be saved too.

This feature can be turned off temporarily via a menu option or permanently in **wordsmith.ini** (in your lwsmith4 folder).

12.23 Menü- und Schaltflächenoptionen

Diese Funktionen können in jedem Tool angezeigt oder ausgeblendet sein, je nach Funktion des Tools oder des aktuellen Datenfensters -- die Titelleiste des aktuellen Fensters ist hervorgehoben.



Tipps

öffnet ein Fenster, in dem eine Karte von Oxford WordSmith Tools angezeigt wird. Sie sehen, wo Sie sich derzeit befinden, und was Sie als nächstes tun können. Weiters werden Tipps angezeigt, je nachdem mit welchem Tool Sie gerade arbeiten.



assoziierte Wörter

öffnet ein neues Fenster, in dem [assoziierte Wörter](#) angezeigt werden.



automatisch verbinden

verbindet ([lemmatisiert](#)) automatisch.



automatische Größenanpassung

jede angezeigte Zeile wird automatisch angepasst, sodass alle Daten angezeigt werden, die sichtbar sein sollen. In den meisten Fenstern ist die Größe der Zeilen fixiert, aber in manchen, z. B. in Viewer, kann die Höhe der Zeilen angepasst werden. So wird die Zeilenhöhe entsprechend der aktuell markierten Datenspalte angepasst.



assoziierte Gruppen

berechnet [assoziierte Gruppen](#) in einer Schlüsselwort-Datenbank



assoziierte Gruppen neu gruppieren

[gruppiert](#) die assoziierten Gruppen neu



Cluster

berechnet Konkordanz-[Cluster](#).



Kollokatoren

zeigt die [Kollokatoren](#) basierend auf Konkordanzdaten an.



berechnen

berechnet eine [neue Datenspalte](#) basierend auf Rechnerfunktionen und/oder vorhandenen

Daten.

Kollokator neu berechnen

berechnet Kollokatoren neu, z. B. nachdem Sie Konkordanzzeilen entfernt haben.

Spaltensummen

berechnet Summen, Minimum, Maximum, Durchschnitt und Standardabweichung für jede [Spalte](#) mit numerischen Daten.

Konkordanz

startet Concord in KeyWords und WordList, und erstellt eine Konkordanz der markierten Wörter basierend auf den ursprünglichen Originaltexten.

kopieren

[kopiert](#) Dateien an verschiedene Orte (Drucker, [Textdatei](#), [Zwischenablage](#), etc.).

2 Spalten verdoppeln

verdoppelt die Anzahl der Spalten, um beim Drucken Papier zu sparen.

bearbeiten

Liste oder Wortsuchen werden [bearbeitet](#) ([Suchen durch Eingabe](#)).

1 Modus "Bearbeiten" oder Modus "Eingabe"

wechselt zwischen dem Modus "Bearbeiten" und dem Modus "Eingabe"

Dateinamen

öffnet ein neues Fenster, in dem die [Namen der Dateien](#) angezeigt werden, aus denen die aktuellen Daten stammen. Wenn notwendig, können Sie die [Dateinamen bearbeiten](#).

F Dateien finden

findet die Textdateien, die alle markierten Wörter enthalten.

vergrößern

vergrößert alle Zeilen auf eine fixierte Höhe. Siehe verkleinern () unten.

? Hilfe (auch F1)

öffnet die WordSmith-Hilfe (diese Datei) mit kontextbezogener Hilfe.

zusammenführen

ein Eintrag wird mit einem anderen Eintrag zusammengeführt, z. B. Sätze in Viewer, Wörter in WordList ([Lemmatisierung](#)).

Layout

[Layouteinstellungen](#) können verändert werden: die Farbe der Spalten, das Ausblenden von Datenspalten, Schriftbilder und Spaltenbreite.

Links

berechnet [Links](#) zwischen Wörtern in einem Schlüsselwortplot.

markieren

markiert einen Eintrag zum [Zusammenführen](#) oder [Dateien finden](#).

Lemmata übereinstimmen

vergleicht jedes Element der Liste mit den Elementen einer [Lemmaliste](#) und verbindet alle damit übereinstimmenden Einträge.

= Matchliste

vergleicht Einträge in der aktuellen Liste mit den Einträgen in einer ["Matchlistendatei" oder -vorlage](#), und markiert alle gefundenen Elemente mit (-).

Transinformation

berechnet [Transinformation](#)swerte in einer [WordList-Indexliste](#).

Neu...

[zum Einstieg](#) in die verschiedenen Tools, z. B. Erstellen einer Konkordanz, Wortliste oder Schlüsselwortliste.

Öffnen...

bietet die Möglichkeit, gespeicherte Ergebnisse auszuwählen.

Muster

berechnet [Kollokationsmuster](#).

**Medien abspielen**

gibt eine [Mediendatei](#) wieder

**Plot**

öffnet ein neues Fenster, in dem ein [Concord-Verteilungsplot](#) oder [KeyWords-Plot](#) angezeigt wird.

**Drucken (auch F3)**

zeigt eine Druckvorschau für Ihre zu druckenden Daten an; kann in Datei drucken, was "[als Text speichern](#)" entspricht.

**aktualisieren**

liest die Textdatei wird neu ein (in Viewer) oder aktualisiert die Ansicht (in Druckvorschau).

**Ersetzen**

Suchen und Ersetzen, z. B. ersetzen von Laufwerk- oder Verzeichnisdaten beim [Bearbeiten von Dateinamen](#), wenn die Quelltexte verschoben wurden.

**neu sortieren**

sortiert Listen werden sortiert (z. B. nach Frequenz statt alphabetisch) in [Concord](#), [KeyWords](#) oder [WordList](#).

**Lineal**

Einblenden/Ausblenden von vertikalen Abschnitten in einer beliebigen Liste oder von Textabschnitten in einem [KeyWords-Plot](#). Klicken Sie auf Lineal im Menü, um die Anzahl der Linealabschnitte für einen [Plot](#) zu verändern.

**Speichern (auch F2)**

[speichert Ihre Daten](#) unter dem bestehenden Dateinamen; bei einer neuen Datei werden Sie zuerst nach einem Dateinamen gefragt.

**speichern als**

speichert, nachdem Sie einen Dateinamen angegeben haben.

**als Text speichern**

speichert .txt-Datei: Standardtext. Mit dieser Funktion gespeicherte Daten können mit Oxford WordSmith Tools nicht weiter untersucht werden.

**Suchen (auch F12)**

[sucht](#) innerhalb einer Liste.

**verkleinern**

verkleinert alle Zeilen auf eine kleinere fixierte Höhe. Siehe vergrößern () oben.

**skimmen**

in Viewer können Sie hiermit einen Text skimmen (= schnell durchlesen).

**Statistik**

öffnet ein neues Fenster, in dem eine [detaillierte Statistik](#) angezeigt wird..

Statusleiste

wechselt zwischen Ein- und Ausblenden der "Statusleiste" (am unteren Rand eines Fensters; enthält Kommentare und Angaben über den aktuellen Status)

**Statistikzusammenfassung**

öffnet ein neues Fenster, in dem die [Statistikzusammenfassung](#) angezeigt wird, z. B. die Proportion von Lemmata und Wort-Types.

**Spalten gegen Zeilen tauschen**

vertauscht Spalten und Zeilen. WordList-[Statistik](#) wird standardmäßig mit den Dateidaten in jeder Spalte angezeigt. Klicken Sie auf diese Schaltfläche, um die Zeilendaten mit den Spaltendaten zu vertauschen.

Symbolleiste

wechselt zwischen Ein- und Ausschalten der Symbolleiste, die dieselben dieselben Schaltflächen enthält, die Sie für [benutzerdefinierte Popupmenüs](#) festgelegt haben.

**Trennen**

alle Einträge, die zusammengeführt wurden, werden wieder getrennt, z. B. [lemmatisierte](#)

Einträge.



Quelltextansicht

[zeigt den Quelltext an](#) und markiert alle Wörter, die aktuell in der Liste markiert sind.



Microsoft Word(tm)

fügt formatierte Daten in Word ein.



WordList

innerhalb von KeyWords wird basierend auf den aktuellen Daten eine [Wortliste erstellt](#).



entfernen

[entfernt](#) gelöschte Einträge.

Siehe auch: [Tastenkombinationen](#), [Popupmenüs anpassen](#).

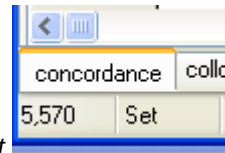
12.24 Modus "Bearbeiten" vs. Modus "Eingabe"

Most windows allow you to press keys either

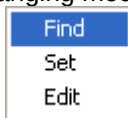
- to edit your data (edit mode), or
- to get quickly to a place in a list (type-in mode).

Concordance windows use key presses also for setting [categories](#) for the data, or for [blanking](#) out the search word.

In type-in mode, your key-presses are supposed to help you [get quickly](#) to the list item you're interested in, e.g by typing theocr to get to (or near to) theocracy in a word list. If you've typed in 5 letters and a match is found, the search stops.



Changing mode is done by right-clicking on the word Set and choosing from



the menu which opens up.

See also: [user-defined categories](#).

12.25 Plot-Verteilungswert

Wozu dient das?

Ein Verteilungswert beschreibt, wie gleichmäßig eine Gruppe von Werten verteilt ist. Denken Sie an Regenfälle in Großbritannien, die im Allgemeinen eher gleichmäßig über das ganze Jahr verteilt sind. Vergleichen Sie dies mit Ländern, in denen es eine Regenzeit gibt.

In der Linguistik ist es oft interessant zu wissen, wie ein Wort in einem Text verteilt ist, z. B. das englische Wort *skull* in Hamlet, und WordSmith konnte diese Information anhand eines Plots schon in der ersten Version anzeigen. Die Verteilungswertstatistik bietet mathematische Informationen und erleichtert den Vergleich von Texten.

Wie wird dieser Wert berechnet?

Die in KeyWords und Concord berechnete Plotverteilung wird mit der ersten der drei Formeln von [Oakes](#) (1998: 190-191) ermittelt, die er selbst als verlässlichste der drei Formeln auswies. Wie beim [Lineal](#) wird der Plot dazu in acht Segmente unterteilt.

Der P-Wert kann von 0 bis 1 reichen, wobei Werte zwischen 0,9 und 1 eine sehr gleichmäßige Verteilung bedeuten, und Werte von 0 bis 0,1 eine sehr ungleichmäßige Verteilung ([Katz](#), 1996).

Siehe auch: [Schlüsselwort-Plot](#), [Concord-Verteilungsplot](#).

12.26 Quelltexte finden

Für einige Berechnungen müssen die originalen Quelltexte zur Verfügung stehen. Wenn zum Beispiel in Concord [mehr Kontext](#) angezeigt werden soll, als für jede Zeile gespeichert wurde, muss Concord den Quelltext neu einlesen. Wenn Sie mit KeyWords einen [Verteilungsplot](#) erstellen wollen, muss KeyWords jeden Quelltext neu einlesen um festzustellen, welche Schlüsselwörter nahe beieinander stehen, und um die Position der Schlüsselwörter im Text und die Position der [Links](#) zu berechnen.

Das ist nicht möglich, wenn Sie die Quelldatei(en) in der Zwischenzeit verschoben oder gelöscht haben.

Siehe auch: [Dateinamen bearbeiten](#), [Quelldateien auswählen](#).

12.27 RAM-Auslastung

Je mehr Arbeitsspeicher (RAM) Sie auf Ihrem Computer haben, desto schneller ist das Programm und desto mehr kann es speichern. Während eines Vorgangs speichert jedes Programm die Ergebnisse in den Arbeitsspeicher. Eine Wortliste mit mehr als 80.000 Einträgen aus einem Text mit über vier Millionen Wörtern benötigt etwa 3 Megabyte Arbeitsspeicher. (Finnisch z. B. benötigt sehr viel mehr Platz.) Wenn der Arbeitsspeicher zu voll wird, versucht Windows Platz zu schaffen, indem es einige Ergebnisse in temporäre Verzeichnisse auf die Festplatte legt. Wenn das passiert, hören Sie wahrscheinlich viele Knackgeräusche, da der Rechner die Daten zuerst auf die Festplatte speichern und danach wieder einlesen muss. Normalerweise hören Sie in jedem Fall *manchmal* ein Knacken, da **Oxford WordSmith Tools** auf die Originaltext zugreifen muss, die sich auf der Festplatte befinden. Zu lautes anhaltendes Knacken deutet jedoch darauf hin, dass der Rechner seine natürlichen Grenzen erreicht hat.

Sie können herausfinden, wie viel Arbeitsspeicher Ihnen während eines Vorgangs noch zur Verfügung steht, indem Sie F9 drücken (in jedem Programm mit der Option "Über" im *Hilfe*-Hauptmenü). In der ersten Zeile wird die RAM-Auslastung angezeigt. Die anderen Zahlen betreffen Windows-Systemressourcen: diese Werte sollten keine Probleme bereiten, sinken sie jedoch unter 20%, dann sollten Sie die [Ergebnisse speichern](#), Windows beenden und neu starten.

Theoretisch können Wortlisten und Schlüsselwortlisten bis zu 2.147.483.647 einzelne Einträge beinhalten. Jedes dieser Wörter kann in Ihren Texten bis zu 2.147.483.647 Mal vorkommen. (Diese ungewöhnliche Zahl, die Hälfte von 2 hoch 32, ist die größte Ganzzahl mit Vorzeichen, die in 32 Bits gespeichert werden kann, und entspricht 2 Gigabyte.) Dieses theoretische Limit werden Sie höchst wahrscheinlich nicht erreichen: damit das englische Wort *the* 2.147.483.647 in Ihren Text vorkommt, müssten Sie etwa 30 Milliarden Wörter verarbeiten. (Das würde den Speicherplatz von 300 CD-ROMs bedeuten, da auf einer CD-ROM, die nur Text enthält, Platz für etwa 100 Millionen Wörter ist.) Dafür würde der Arbeitsspeicher schon längst nicht mehr ausreichen.

Wenn Sie über 64MB RAM oder mehr verfügen, sollten Sie im Arbeitsspeicher Platz für eine Kopie einer Wortliste basierend auf Millionen von Wörtern sowie gleichzeitig ein leistungsfähiges Textverarbeitungsprogramm sowie eine Textdatei haben.

Siehe auch: [Geschwindigkeit](#)

12.28 Referenzkorpus

Referenzkorpus

Textkorpus, der zu Vergleichszwecken dient. Vielleicht möchten Sie einen Textausschnitt mit dem [British National Corpus](#), einer Sammlung von Texten mit 100 Millionen Wörtern, vergleichen. Das ist sehr nützlich, wenn Sie [Schlüsselwörter](#) berechnen.

Im [Controller](#) können Sie die [Wortliste des Referenzkorpus festlegen](#), damit diese Wortliste von KeyWords und Concord verwendet wird.

12.29 Schnittstellen zwischen Tools

Schnittstellen mit Textverarbeitung, Tabellenkalkulation, Browser, etc.

Aus allen Fenstern, in denen Listen oder Texte angezeigt werden, kann ausgewählte Information sehr einfach in die [Zwischenablage](#) kopiert werden. (Verwenden Sie zum Einfügen Strg+Einfüg.)

 Wenn Sie dieses Symbol sehen, können Sie damit alle markierten Daten direkt in ein Microsoft Word(tm) Dokument einfügen.

Wenn Sie eine URL sehen (wie z. B. <http://www.lexically.net>) können Sie Ihren Browser öffnen, indem Sie auf die URL klicken.

Schnittstellen zwischen den verschiedenen Tools

Die Programme in **Oxford WordSmith Tools** sind miteinander über [wshell.exe](#) verknüpft (das Programm mit der Beschriftung "Oxford WordSmith Tools [Controller](#)", das in der linken oberen Ecke Ihres Bildschirms erscheint). Darin werden alle [Standardeinstellungen](#) verwaltet, z. B. Farben, Verzeichnisse, Schriftarten, Stopplisten, etc.

Generell, wenn Sie in **WordList** oder **KeyWords** Strg+C drücken, wird anhand des aktuellen Worts und der aktuellen Dateien direkt eine Konkordanz berechnet und angezeigt.

Drücken Sie Strg+W in **Concord** oder **KeyWords**, wird anhand der aktuellen Dateien eine Wortliste erstellt.

Jedes Tool wird so viele relevante Informationen wie möglich zu dem aufgerufenen Tool senden. Diese umfassen: das aktuelle Wort (das markierte Wort im Bildlauffenster) und die Textdateien, aus denen die aktuellen Informationen stammen.

Beispiel: nachdem eine Wortliste basierend auf drei Wirtschaftstexten berechnet wurde, stellen Sie fest, dass das Wort *hoffnungsvoll* häufiger vorkommt, als Sie erwartet hatten. Sie möchten anhand derselben Texte eine Konkordanz erstellen. Markieren Sie das Wort *hoffnungsvoll*, halten Sie Strg gedrückt und drücken Sie dazu C. Jetzt können Sie sehen, ob *hoffnungsvoll* Teil eines 3-Wort-[Clusters](#) ist, oder Sie untersuchen den Verteilungsplot.

Beispiel: Nachdem Sie eine Schlüsselwort-[Datenbank](#) basierend auf 300 Wirtschaftstexten erstellt haben, stellen Sie fest, dass das Wort *bid* ein gewichtetes Schlüsselwort zu sein scheint, und dass es mit den Wörtern *company*, *shares*, etc. assoziiert ist. Markieren Sie das Wort *bid* und drücken Sie Strg+C, dann wird eine Konkordanz derselben 300 Texte berechnet. Jetzt können Sie sich die Kontexte ansehen: ist *bid* ein Streben nach Macht oder Teil eines Angebotsvorgangs?

Beispiel: Sie haben eine Konkordanz des Worts *grün*. Drücken Sie jetzt Strg+W um eine Wortliste derselben Textdateien zu erstellen. Drücken Sie Strg+K um diese Wortliste mit einer Referenzkorpusliste zu vergleichen und die Schlüsselwörter in diesen Textdateien festzustellen.

12.30 Statusleiste

Die Statusleiste befindet sich am unteren Rand des Fensters. Dort kann ganze Fenster vergrößert oder verkleinert werden. Außerdem enthält die Statusleiste eine Reihe von Informationen über die aktuellen Daten. Die Statusleiste kann normalerweise über eine Option im Hauptmenü eingeblendet oder ausgeblendet werden. Wenn Sie mit der rechten Maustaste auf die Leiste klicken wird ein Popupmenü angezeigt, in dem Sie zwischen [Finden, Festlegen und Bearbeiten](#) wählen können.

12.31 Systemanforderungen

Diese **Oxford WordSmith Tools**-Version ist vorgesehen für Rechner mit:

- mindestens 256 MB RAM (128 MB sind möglicherweise ausreichend, jedoch nicht unter Windows XP)
- mindestens 40MB Speicherplatz auf der Festplatte
- Windows™ 98, NT, 2000, XP oder höher, oder einem Emulator eines dieser Systeme, wenn Sie unter einem Apple Mac oder Unix-System arbeiten. Für das Funktionieren unter Windows 95, zweite Fassung, kann nicht garantiert werden.

Das Programm läuft besser auf [schnellen](#) Rechnern, speziell Rechnern mit großem Arbeitsspeicher ([RAM](#)).

12.32 Tastenkombinationen

Fenster scrollen:

Strg+Pos1 Listenanfang

Strg+Ende Listenende

ist die Liste alphabetisch geordnet, [geben Sie das Suchwort ein](#)

ist die Liste horizontal verschiebbar:

Pos1 zum linken Rand

Ende zum rechten Rand

Strg+"Pfeil-nach-rechts" ein Wort nach rechts

Strg+"Pfeil-nach-links" ein Wort nach links

Hotkeys:

Strg+C **Concord** aus einem anderen Tool [aufrufen](#)

Strg+W **WordList** aus einem anderen Tool aufrufen

Strg+Einf gesperrten Abschnitt in die [Zwischenablage](#) kopieren

Umschalt-Cursor-Tasten einen Abschnitt sperren

F1 Hilfe 

F2 [Ergebnisse speichern](#) 

F3 Ergebnisse drucken 

F4 Einträge zusammenführen 

F5 Einträge für Zusammenführung markieren



F6 neu [sortieren](#) 

Strg/F6 [Sortierung](#) umkehren 

F7 Quelltextansicht 

F8 nach kurzen Sätzen suchen (**Viewer**)

F9	Dialog "Info" (zeigt Versionsdatum und Auslastung des Arbeitsspeichers an)
F12	in einer Liste suchen 
Strg/M	zwei Wortlisten oder KeyWords-Datenbanken zusammenführen
Alt/H	aufrufen der <i>Hilfe</i> -Untermenüs
Alt/S	aufrufen der Einstellungen-Untermenüs
Alt/W	aufrufen der Fenster-Untermenüs
Alt/X	Tool Beenden
Strg/Z	gelöschte Zeilen Entfernen

Siehe auch: [Menüelemente und Schaltflächen](#)

12.33 Tools zum Erkennen von Mustern

Bei fast jeder menschlichen Tätigkeit werden Werkzeuge benötigt, anfangen bei der Erzeugung von Geschirr bis hin zur Wettervorhersage. Computertools sind nützlich, weil sie bestimmte Aufgaben sehr schnell und einfach erledigen können, wodurch Werkzeuge das Durchführen komplexer Arbeiten ermöglichen. Es wird möglich Erkenntnisse zu gewinnen, da man eine Idee schnell und einfach ausprobieren und damit experimentieren kann, und aus Experimenten entstehen Erkenntnisse. Außerdem, durch die Umwandlung von Datensätzen in eine neue Form kann der Mensch Muster erkennen.

Das ist ironisch. Der Computer ist sehr schlecht dazu geeignet, Muster zu erkennen. Ein Computer eignet sich zum Addieren, Sortieren, etc. Ein Computer verfügt über einen Speicher, er kann aber nichts wissen oder verstehen, und es ist für einen Computer sehr aufwändig, gedruckte Zeichen zu erkennen (geschweige denn handgeschriebene Zeichen).

Nichtsdestotrotz kann der Computer den *Menschen* sehr gut dabei unterstützen, Muster und Tendenzen zu erkennen. Deshalb ist es wichtig Computertools wie **Oxford WordSmith Tools** richtig einzuschätzen. Ein Tool kann bei der Erfüllung einer Aufgabe helfen, es kann jedoch die Arbeit nicht selbstständig erledigen.

Tool versus Produkt

Es gibt Software, die als Produkt gestaltet ist. Ein Computerspiel ist ein in sich geschlossenes Produkt, ein elektronisches Wörterbuch ebenso. Andererseits Programme für Textverarbeitung, Tabellenkalkulation oder Datenbanken sind Tools (engl. für "Werkzeuge"), weil sie über ihre eigenen Grenzen hinaus funktionieren: Sie verwenden diese Programme für Aufgaben, die für den Hersteller nicht vorhersehbar sind. **Oxford WordSmith Tools** ist eine Gruppe von Tools, nicht von Produkten, wie bereits der Name besagt. Die verschiedenen Tools dienen dazu, viele Arten von Mustern zu erkennen, und das in praktisch allen Texten in einer Vielzahl von [Sprachen](#).

Erkenntnis durch Umwandlung

Die hier gestellten Ansprüche sind keineswegs religiöser, sondern psychologischer Natur. Das Verändern der Datenform, das Reduzieren der Daten und deren Umwandlung in ein neues Format fördert die menschliche Fähigkeit, Muster zu erkennen. Der Computer kann nichts "erkennen". (Wenn Sie 2 in einen Taschenrechner eingeben, und die Zahl immer wieder verdoppeln möchten, erkennt der Taschenrechner nicht, worauf Sie abzielen, und wird den Vorgang nicht automatisieren!) Menschen haben die Fähigkeit zu erkennen, insbesondere visuelle Muster.

Wird der Text zu einer Liste umgewandelt oder die Aufteilung der Schlüsselwörter in Quelltexten grafisch dargestellt, wird ein Mensch sehr wahrscheinlich ein Muster erkennen. Das passiert ganz automatisch. Manchmal sehen Mensch Muster, die nicht beabsichtigt sind (z. B. in einer Wolke). Es gibt keine Garantie dafür, dass das Muster "wirklich vorhanden" ist, es liegt im Auge des Betrachters.

Oxford WordSmith Tools soll dem Menschen dabei helfen, diese Muster zu erkennen, und

daraus Erkenntnisse zu gewinnen. Mit den WordSmith Tools gewinnen Sie Ihre eigenen Erkenntnisse über Ihre eigenen Daten aus Ihren eigenen Texten.

Toolarten

Alle Werkzeuge sind auf zwei Achsen positioniert: Spezialisierung und Dauer.

allgemein ----- spezialisiert

allgemein

Ein Spaten ist ein Gerät, das zum Umgraben und Ausheben von Erde verwendet wird und diese Arbeit erleichtert. Es kann aber auch zum Sandschaufeln oder Schneeräumen verwendet werden. Eine Nähmaschine kann zur Herstellung von Vorhängen oder Taschentüchern verwendet werden. Ein Textverarbeitungsprogramm ist allgemein verwendbar.

spezialisiert

Ein Fingerhut dient zum Schutz der Finger beim Nähen, und wird selten anderweitig verwendet. Ein Overlock-Nähmaschine dient zum Einfassen von Knopflöchern und zum Säumen: für diese Aufgaben ist sie besser geeignet als eine normale Nähmaschine, aber die Anwendungsbereiche sind speziell. Die Rechtschreibprüfung in einem Textverarbeitungsprogramm ist ziemlich spezialisiert.

temporär ----- permanent

temporär

Der Ast, mit dem ein Gorilla Früchte von einem Baum holt, ist ein temporäres Werkzeug. Nach der Verwendung wird das Werkzeug wieder zu einem einfachen Ast. Mit einem Brett zum Glattstreichen von Beton verhält es sich ähnlich. Das Brett wird nicht als Werkzeug bezeichnet, obwohl es als solches verwendet wird. Diese Art von Behelfswerkzeugen wird in Brasilien "quebra-galho" genannt (dt. 'Astbrecher').

permanent

Ein Stemmeisen wird als permanentes Werkzeug produziert, katalogisiert und verkauft. Es hat seinen eigenen Platz in unserem Vokabular. Einmal gekauft, muss es an einem bestimmten Platz sorgfältig aufbewahrt werden.

Die Tools (Werkzeuge) in Oxford WordSmith Tools wurden aus ursprünglich temporären Werkzeugen zu permanenten Werkzeugen entwickelt. Sie sollen außerdem als allgemeine

Werkzeuge dienen: sie sind das Schweizermesser  für Lexik. Es besteht keine Verletzungsgefahr, aber man muss wissen, wie man die Tools nützen kann.

Siehe auch: [Danksagung](#)

12.34 Tool-spezifische Einschränkungen

Concord: Einschränkungen

Sie können mit **Concord** eine praktisch unbegrenzte Anzahl von Konkordanzzeilen berechnen. Concord lässt [Suchwörter oder -phrasen](#) bis zu 80 Zeichen zu, obwohl Sie eine unbegrenzte Anzahl von Suchwörtern in Ihrer [Suchwortdatei](#) angeben können.

Jede Konkordanz kann eine unbegrenzte Anzahl von Kollokatoren speichern innerhalb eines maximalen [Kontextbereichs](#) von maximal 25 Wörtern links bzw. rechts des Suchworts.

WordList: Einschränkungen

Ein Haupteintrag kann tausende [Lemmata](#) enthalten, aber Sie können mit F4 nur 20 Elemente auf einmal zusammenführen. Wiederholen Sie diesen Vorgang bei Bedarf.

[Detaillierte Konsistenz](#)listen können bis zu 50 Dateien verarbeiten.

KeyWords: Einschränkungen

Ein Schlüsselwortplot pro Schlüsselwortanzeige. (Wenn Sie mehr möchten, rufen Sie dieselbe Datei in einem neuen Anzeigefenster auf.)

Anzahl der [Link](#)-Fenster pro Schlüsselwort-[Plot](#)anzeige: 20.

Anzahl der Fenster von [assoziierten Wörtern](#) pro Anzeige gewichteter Schlüsselwörter: 20.

Splitter: Einschränkungen

Jede Zeile einer großen Textdatei kann bis zu 10.000 Zeichen lang sein. Es sollte also spätestens nach 10.000 Zeichen durch eine Zeilenschaltung ein neuer Absatz begonnen werden.

Text Converter: Einschränkungen

Es können gleichzeitig bis zu 500 Strings gesucht und ersetzt werden.

Jeder zu suchende String und jeder zu ersetzende String kann bis zu 80 Zeichen lang sein.

Ein Sternchen darf nicht das erste oder letzte Zeichen innerhalb eines zu suchenden Strings sein.

Wenn das Sternchen als Platzhalter verwendet wird, liegt die Grenze bei 1.000 Zeichen.

Text Viewer: Einschränkungen

Wenn Sie beim Texte auswählen die Option "Anzeigen"  auswählen, werden von **Viewer** die ersten zehn ausgewählten Quelltextdateien aufgerufen.

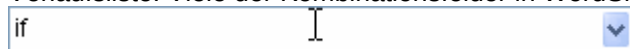
Wenn Sie Texte auswählen oder in die Mitte eines Texts springen (z. B. nachdem Sie in Concord  ausgewählt haben), wird **Viewer** nur 10.000 Zeichen jeder Datei verarbeiten, um die Geschwindigkeit der Verarbeitung sehr langer Dateien zu erhöhen. Sie können mit Viewer die Datei jedoch "neu einlesen", indem Sie  anklicken, um die Anzeige zu aktualisieren. Dabei wird von **Viewer** der gesamte Text verarbeitet.

Sie können nicht gleichzeitig an zwei oder mehr durch Alignment synchronisierten Texten arbeiten.

Siehe auch: [Allgemeine Einschränkungen](#)

12.35 Verlauf

Verlaufsliste: Viele der Kombinationsfelder in WordSmith, wie dieses Suchwortfeld



, speichern früher eingegebene Wörter, die Sie aufrufen können, indem Sie auf den kleinen abwärts gerichteten Pfeil rechts neben der Schaltfläche klicken.

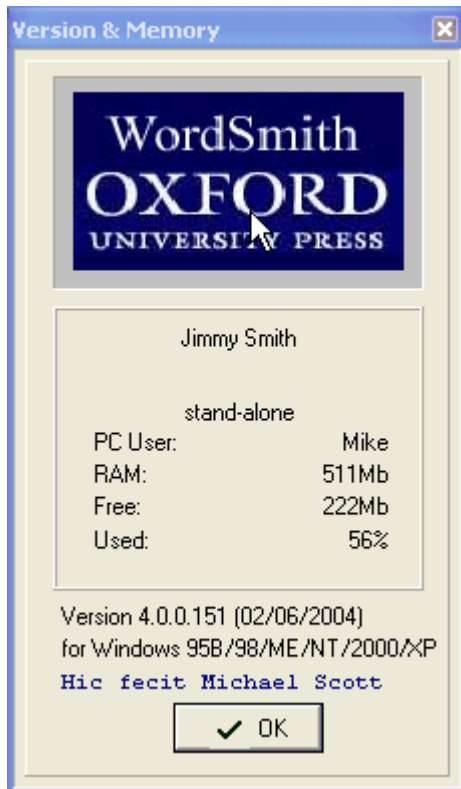
12.36 Versionsinformation

Diese Hilfedatei gilt für die aktuelle Version von **Oxford WordSmith Tools**.

Die Version von **Oxford WordSmith Tools** wird unter der Option *Über* (F9) angezeigt. Dort finden Sie auch Ihren registrierten Namen und die Menge des verfügbaren [Arbeitsspeichers](#). Wenn Sie eine Demoversion haben, wird dies direkt unter Ihrem Namen angezeigt.

Überprüfen Sie das Datum in diesem Feld, das anzeigt, wie aktuell Ihre Version ist. Werden Vorschläge in das Programm implementiert, dann werden die verbesserten Versionen zum Download zur Verfügung gestellt. Bewahren Sie Ihre Lizenznummer für aktualisierte Versionen auf.

Sie können auf die Grafik WordSmith Oxford Tools in dem Feld "Über" klicken, um Ihre aktuelle Lizenznummer anzuzeigen.



Siehe auch: [32-bit-Version: Unterschiede](#), [Demoversion](#), [Kontaktadressen](#).

12.37 Verzeichnisse\Laufwerke

Found in main Settings menu in all Tools. Default folders can be altered in Oxford WordSmith Tools or set as [defaults](#) in `wordsmith.ini`.

- Concordance Folder: for your concordance files.
- KeyWords Folder: for your key-word list files.
- WordList Folder: where you will usually [save](#) your word-list files.
- Texts Folder: where your text files are to be found.
- Downloaded Media: where your [sound & video files](#) will be stored after downloading the first time from the Internet.
- Settings: where your settings files (.ini files and some others) are kept.

If you write the name of a folder which doesn't exist, Oxford WordSmith Tools will create it for you if possible. (On a network, this will depend on whether you have rights to create folders and [save](#) files.)

If you change your Settings folder, you should let WordSmith copy any `.ini` and other settings files which have been created so that it can keep track of your language preferences, etc.

Note: in a network, drive names such as **G:**, **H:**, **K:** change according to which machine you're running from, so that what is **G:\texts\my text.txt** on one terminal may be **H:\texts\my text.txt** on another. Fortunately network drives also have names structured like this: [\\computer_name\drive_name\](#). You will find that these names can be used by **WordSmith**, with the advantage that the same text files can be accessed again later.

Tip

Use different folders for the different functions in Oxford WordSmith Tools. In particular, you may end up making a lot of word lists and key word lists if you're interested in making [databases](#) of key words. It is theoretically possible to put any number of files into a folder, but accessing them seems to slow down after there are more than about 500 in a folder. Use the batch facility to produce very large numbers of word list or key words files. I would recommend using a `\keywords` folder to store `.kdb` files, and `\keywords\genre1`, `\keywords\genre2`, etc. for the `.kws` files for each genre.

See also: [finding source texts](#).

12.38 Zahlen

Abhängig von Ihren Texteeinstellungen möchten Sie möglicherweise Zahlen in Ihre Wortlisten aufnehmen bzw. daraus ausschließen.

12.39 Zeichensätze

12.39.1 Übersicht

You need "[plain text](#)" in WordSmith. Not Microsoft Word `.doc` files -- which contain text and a whole lot of other things too that you cannot normally see.

To handle a text in a computer, programs need to know how the text is encoded. In its processing, the software sees only a long string of numbers, and these have to match up with what you and I can recognise as "characters". For many languages like English with a restricted alphabet, encoding can be managed with only 1 "byte" per character. On the other hand a language like Chinese, which draws upon a very large array of characters, cannot easily be fitted to a 1-byte system. Hence the creation of other "multi-byte" systems. Obviously if a text in English is encoded in a multi-byte way, it will make a bigger file than one encoded with 1 byte per character, and this is wasteful of disk and memory space. So, at the time of writing, 1-byte character sets are still in very widespread use.

In practice, your texts are likely to be encoded in a Windows 1-byte system, older texts in a DOS 1-byte system, and newer ones, especially in Chinese, Japanese, Greek, in Unicode. What matters most to you is what each character looks like, but WordSmith cannot possibly sort words correctly, or even recognise where a word begins and ends, if the encoding is not correct. WordSmith has to know (or try to find out) which system your texts are encoded in. It can perform certain tests in the background. But as it doesn't actually understand the words it sees, it is much safer for you to [define the character set in advance](#), especially if you process texts in German, Spanish, Russian, Greek, Polish, Japanese, Farsi, Arabic etc.

Three main kinds of character set, each with its own flavours, are [Windows](#), [DOS](#), and [Unicode](#).

Tip

To check results after changing the code-page, select [Choose Texts](#) and View the file in question. While viewing you can change Text Characteristics until it looks right. If you can't get it to look right, you've probably not got a cleaned-up [plain text](#) file but one straight from a word-processor. In that case, take it back into the word-processor and [save it as text](#) again as a plain text file in Windows format, which is more up-to-date than DOS formats.

See also: [Choosing Accents & Symbols](#), [Accented characters](#); [Choosing Language](#)

12.39.2 Akzentzeichen und Symbole

Diese Funktion in den [Konkordanzeinstellungen](#) und anderen Dialogfeldern ermöglicht das Einfügen von Symbolen und Akzentzeichen in Ihr Suchwort, Stoppwort oder Kontextwort, etc. Wählen Sie einfach das Symbol aus und ziehen Sie es mit der Maus an die Stelle, wo Sie es einfügen möchten, oder Sie fügen das Symbol mit einem Doppelklick ein. Angezeigt werden alle verfügbaren Akzentzeichen, basierend auf aktuellen Einstellungen wie [Zeichensatz](#) und [Texteigenschaften](#), sowie eine Reihe zusätzlicher Symbole, wie z. B. typografische Apostrophe. Einige dieser Symbole ((**r**),(**c**),(**tm**) und typografische Apostrophe) können nur angezeigt werden, wenn Sie die Texte im Windows [ANSI](#)-Format gespeichert haben. Siehe auch: [Sprachen auswählen](#)

12.39.3 ANSI und ASCII

ASCII text, ANSI text, Text Only and DOS text are all names for plain text.

Most word-processors insert special hidden codes into text files to help them keep track of page numbers, bold type and so on. Oxford WordSmith Tools can handle them anyway but you'll get cleaner results if you use plain text without the hidden codes.

If your source texts were saved as "Text Only with line breaks" there will probably be one `<Enter>` every 70 or 80 characters at the end of each text line. If they were saved as "Text Only", the `<Enters>` will be equivalent to paragraph breaks. I recommend saving as "Text Only".

The Windows program **Notepad** (*Start | Program Files | Accessories*) makes plain text files or **.txt** files. It uses basic character sets e.g. **A** to **Z**, numbers and common punctuation symbols. The main difference is in the accented characters. For more on this, see [character sets](#).

See also : [HTML,SGML & XML](#).

12.39.4 DOS

DOS (text format before Windows) offered a range of character sets called "codepages". They all shared the same codes for the standard English alphabet (**a**, for example is always code 97) and common punctuation symbols, but included varying symbols for box-drawing, foreign language accents, etc.

If you process texts in German, Spanish, Russian, Greek, Polish, etc. you may need to find out which codepage was used when the texts were originally typed.

For example, the character **ã** is coded one way in codepage 850 (Multilingual) but differently in codepage 860 (Portuguese). It is simply not available at all in codepage 437 (the default codepage in the UK and USA). To alter or examine codepages, see a DOS manual or check the topic out on the web.

When it loads up, **Oxford WordSmith Tools** detects the current DOS code-page, so the codepage is only likely to need altering if you are using texts produced when another codepage was in use.

12.39.5 Windows

Windows character set codes are different from those in DOS or Unicode. (The **£** symbol is code 156 in DOS but 163 in Windows.) In Windows 95 or later you can get non-Western fonts enabled via Microsoft Plus. If your texts were written using a Windows word-processor and [saved as text](#) in Windows, the accented characters will obey the Windows codes. You will have access to a few more symbols than in DOS (e.g. ®,©,™ and curly apostrophes).

Windows **Western** (1252) format includes:

Anglo-Saxon, Basque, Catalan, Danish, Dutch, English, Middle English, Finnish, French, German, Icelandic, Italian, Norwegian, Old Norse, Portuguese, Spanish, Swedish

Windows **Baltic** (1257) format includes:

Estonian, Latvian, Lithuanian

Windows **Central European** (1250) format includes:

Albanian, Bosnian, Croatian, Czech, Hungarian, Polish, Romanian, Serbian, Slovak, Slovene, Upper Sorbian, Lower Sorbian

Windows **Cyrillic** (1251) format includes:

ByeloRussian, Bulgarian, Macedonian, Russian, Serbian (1251), Ukrainian

Windows **Greek** (1253) handles Greek

and Windows **Turkish** (1254) handles Turkish (what else?)

12.39.6 Unicode

Unicode ist ein Textformatstandard, der ein Zeichen in zwei "Byte" kodiert. Dadurch können mehr als 65.000 verschiedene Zeichen und Symbole dargestellt werden, was wesentlich ist für Sprachen wie Chinesisch, Japanisch, Cherokee und viele andere.

Beim Auswählen von Texten können Sie eine Schaltfläche anklicken, um zu überprüfen, ob die Textdateien in Unicode kodiert sind.

12.40 Zip-Dateien

Zip-Dateien sind Dateien, die standardmäßig komprimiert wurden. **WordSmith** kann *.zip*-Dateien in dieser neuen Version jetzt bearbeiten und speichern.

Wozu dient das?

Abgesehen von dem Vorteil, dass Zip-Dateien wesentlich kleiner als Originaldateien sind, geht so weniger Speicherplatz verloren, und zwar aus dem folgenden Grund: Jede Textdatei, auch wenn sie nur ein Wort enthält, benötigt auf Ihrer Festplatte etwa 4.000 Byte, abhängig von Ihrem System bis zu 32.000 Byte Speicherplatz. Wenn Sie 100 kurze Dateien haben, würden Sie so viele tausend Byte Speicherplatz verlieren. Wenn Sie 100 kurze Dateien "zippen", ist dieser Speicherplatz möglicherweise für diese eine Datei ausreichend. Zip-Dateien werden aufgrund dieser Vorteile häufig bei Datentransfers über das Internet verwendet. Wenn Sie über viele Wortlisten verfügen, so ist es sehr viel effizienter, diese in einer Zip-Datei zu speichern. Die Nachteile des Zippens sind a) die Zeit, die es in Anspruch nimmt und b) die Tatsache, dass Zip-Dateien nur von Softwareprogrammen gelesen werden können, die dieses Standardformat "verstehen". Es gibt zahlreiche Zip-Programme auf dem Markt, u. a. *PKZip*(tm) und *Winzip*(tm). Eine gezippte Wortliste kann von diesen Programmen zwar entzippt, aber danach nicht bearbeitet werden. **WordSmith** kann die Liste zuerst entzippen und dann öffnen.

Wie mache ich das?

Wenn Sie die Option "Zip-Datei erstellen" sehen, können Sie diese auswählen und die Ergebnisse werden als gezippte Datei mit der Dateierweiterung *.zip* in das von Ihnen gewählte Verzeichnis gespeichert.

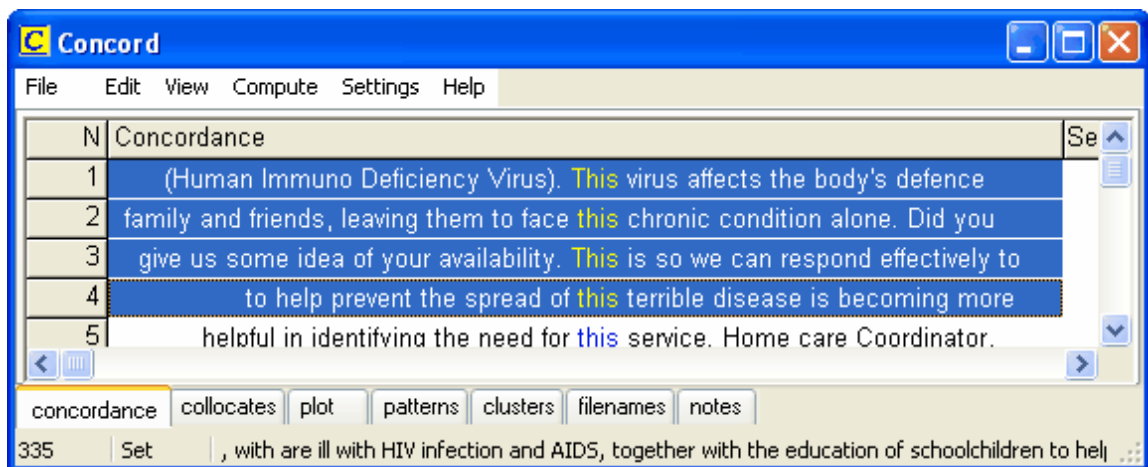
Wenn Sie eine gezippte Wortliste, Konkordanz, Textdatei, etc. öffnen möchten, und diese Zip-Datei enthält mehr als eine Datei, können Sie entscheiden, welche der Dateien Sie öffnen wollen. Anderenfalls läuft dieser Prozess im Hintergrund und beeinflusst andere **WordSmith**-Prozesse nicht.

12.41 Zwischenablage

Sie können einen Datenbereich entweder mit der Maus markieren, oder indem Sie bei gedrückter Umschalttaste den Bereich mit den Cursortasten markieren. Danach kopieren Sie die Daten mit Strg+Einfg oder Strg+C in die Zwischenablage. In einem Textverarbeitungsprogramm können Sie die kopierten Daten in Ihren Text einfügen (oder "speziell einfügen"). Das ist normalerweise einfacher als die Daten [als Textdatei zu speichern](#) (oder in eine Datei zu [drucken](#)) und grafische Darstellungen können auf diese Weise auch kopiert und gespeichert werden.

Beispiel

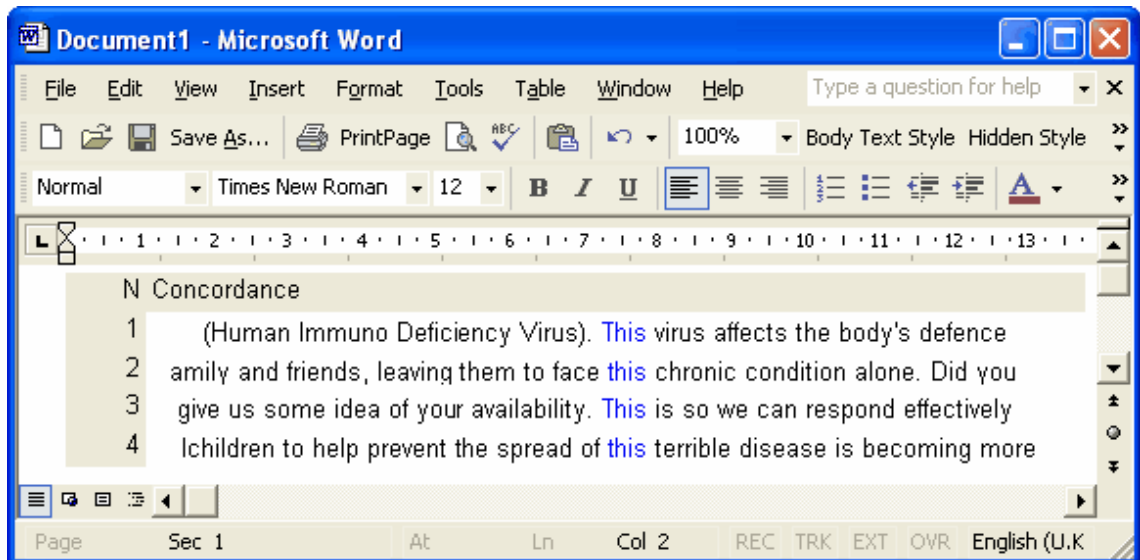
1. Wählen Sie einige Daten aus. In diesem Beispiel wurden die ersten vier Zeilen (von 335) einer Konkordanz ausgewählt, nur sichtbarer Text und keine Klassen- oder Dateinameninformation.



2. Halten Sie Strg gedrückt und drücken Sie dazu Einfg oder C. Die Daten befinden sich nun in der Windows-Zwischenablage und können in einem anderen Programm wie Excel, Word, Editor, etc. in eine Datei eingefügt werden.

Die Daten werden automatisch in zwei Formaten in die Zwischenablage kopiert:

als Bild



Dieses Format verwendet man meistens in wissenschaftlichen Aufsätzen, Dissertationen, etc., und im Fall eines Plots ist kein anderes Format möglich. In dieser Konkordanz werden Sie nur die Wörter erhalten, die Sie in Ihrer Konkordanzzeile sehen (nicht die gesamte Zeile).

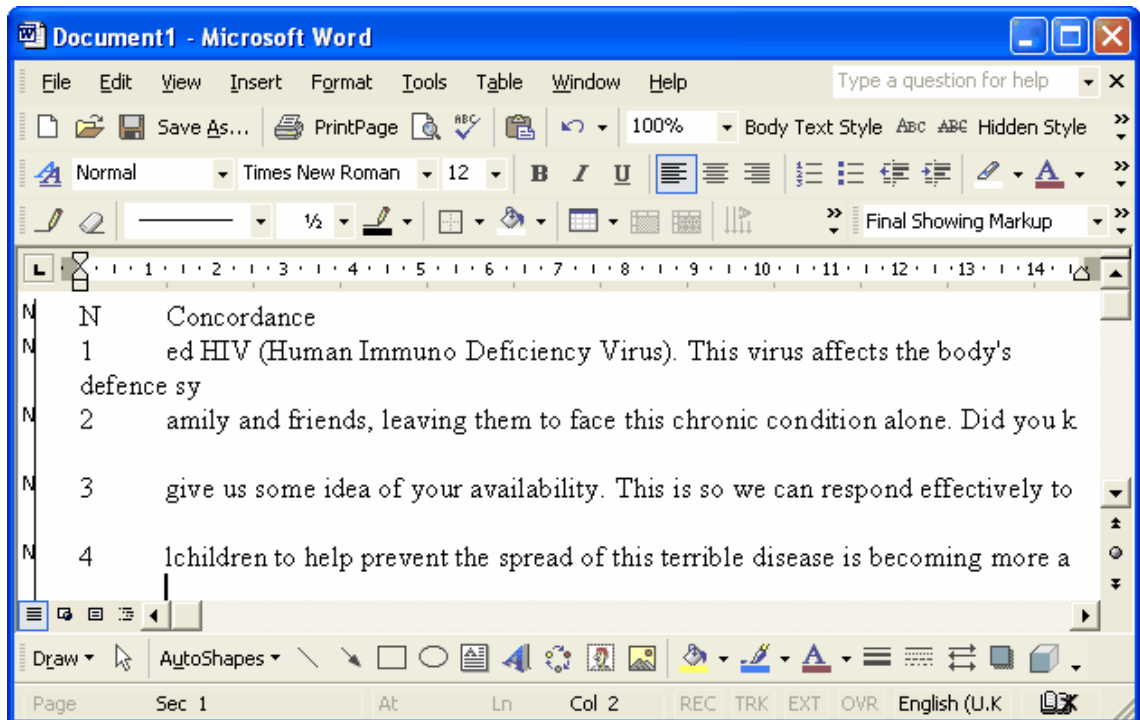
In dieser Grafik sind Farben und graphische Daten enthalten. Wenn Sie danach auf die Grafik klicken, können Sie die Gesamtgröße der Grafik verändern und jedes Wort oder jede Linie bearbeiten (was jedoch einigermaßen kompliziert ist).

In Word ist es empfehlenswert, das Bild mit der Option *Bild (Erweiterte Metadatei)* unter dem Menü *Bearbeiten | Inhalte einfügen* einzufügen. Was Sie in Word sehen ist der Anzeige in Concord sehr ähnlich.

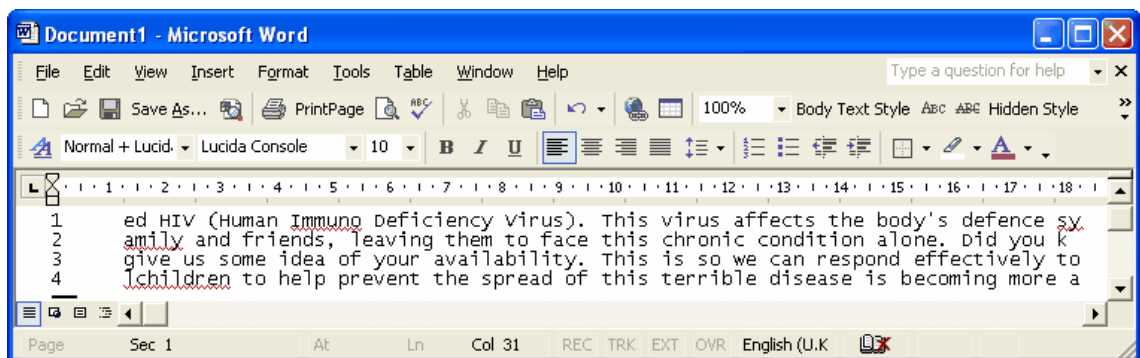
als Standardtext

Die zweite Möglichkeit ist, die Information als Standardtext einzufügen, wenn Sie die Konkordanzlinien für Unterrichtszwecke bearbeiten wollen, oder die Zeilen in eine Tabelle z. B. in MS [Excel](#)(tm) einfügen wollen (was günstiger ist, wenn Sie grafische Daten haben, wie z. B. [Concord-Verteilungsplots](#) oder [KeyWords-Plots](#)).

In diesem Fall werden die Konkordanz oder andere Daten als Standardtext kopiert und die Spalten durch Tabstopps getrennt. Mit dem Windows Texteditor kann man nur dieses Dateiformat lesen und bearbeiten. Microsoft Word fügt die Daten als Text ein (mit Umschalt +Einfg oder Strg+V). Es werden maximal so viele Zeichen eingefügt, wie Sie in den [Einstellungen](#) für "Als Standardtext speichern" festgelegt haben, standardmäßig sind 80 Zeichen festgelegt.



Hier wurden die Konkordanzzeilen kopiert, aber sie sind nicht sehr übersichtlich angeordnet und es ist schwierig, das Suchwort zu erkennen (*this*). Wenn Sie das Suchwort übersichtlich anordnen möchten, sollten Sie eine dicktengleiche Schrift wie z. B. Courier oder Lucida Console verwenden. Das sieht dann folgendermaßen aus:



Beachten Sie, dass bei Lucida Console und Schriftgröße 10 ein Text mit 80 Zeichen und Zahlen am linken Rand über 18 cm breit ist. Um Wortumbrüche zu vermeiden, kann das Papierformat in Word auf Querformat umgestellt werden. Sie können auch die Anzahl der Zeichen pro Zeile auf 50 oder 60 verringern.

Fehlerbehebung

Section



XIII

13 Fehlerbehebung

13.1 abgestürzt

Das Programm ist abgestürzt!

Lösung: **Oxford WordSmith Tools** schließen und erneut ausführen. Sollte das nicht funktionieren, beenden Sie Windows und versuchen Sie es erneut.

13.2 Apostrophe nicht gefunden

Apostrophe nicht verarbeitet

Wenn die Originaldatei mit Microsoft Word(tm) gespeichert wurde, findet **Concord** möglicherweise keine Apostrophe oder Anführungszeichen darin. Grund dafür ist, dass mit Word bestimmte Zeichen automatisch formatiert werden können (AutoFormat). Ein normaler Apostroph oder ein normales Anführungszeichen wird in diesem Fall durch das entsprechende typografische Zeichen ersetzt, das abhängig von der Position am Anfang oder am Ende eines Worts entweder nach links oder rechts geschwungen ist. Diese typografischen Zeichen nicht gleich wie gerade Apostrophe oder Anführungszeichen.

Lösung: Ziehen Sie das Symbol aus dem unteren Zeichensatz, wenn Sie Ihr [Suchwort](#) eingeben, oder ersetzen Sie mit Text Converter alle typografischen Apostrophe und Anführungszeichen durch gerade Apostrophe und Anführungszeichen.

Siehe auch: [Einstellungen](#)

13.3 Demolimit

Demolimit erreicht

Sie haben das Programm möglicherweise gerade heruntergeladen, aber Ihre Registrierungsdetails noch nicht angegeben. Um dies zu tun, gehen Sie in das Oxford WordSmith Tools Hauptfenster, und wählen Sie im Menü *Einstellungen | Registrieren*.

Wenn Sie keine Lizenznummer mit 20 Zeichen besitzen, kontaktieren Sie bitte die [Oxford University Press](#). Der **einzige** Unterschied zwischen einer [Demoversion](#) und einer Vollversion ist: Mit einer Vollversion werden alle Daten angezeigt bzw. gedruckt, mit einer Demoversion jedoch können Sie nur etwa 25 Zeilen der Ergebnisse sehen.

13.4 Drucker druckt nicht

Druckproblem

Wenn auf Ihrem Ausdruck einige Spalten leer sind, andere jedoch korrekt gedruckt wurden, haben Sie möglicherweise einen Drucker, der nur schwarz-weiß drucken kann, nicht jedoch Graustufen. Ändern Sie im [Controller](#) die Einstellung (*Einstellungen anpassen | Allgemein*) auf schwarz-weiß.

13.5 eigenartige Symbole

eigenartige Symbole

eigenartige Symbole beim Arbeiten mit Oxford WordSmith Tools

1. Überprüfen Sie Ihre Textdateien. Lesen Sie die Dateien im **Editor**. Enthalten die Texte viele eigenartige Symbole? Diese Symbole können versteckte Kodierungen sein, die Ihr Textverarbeitungsprogramm verwendet. Lösung: Öffnen Sie die Datei in Ihrem

Textverarbeitungsprogramm und *speichern* Sie die Datei unter einem neuen Namen im [Standardtext](#)format, auch als "Nur-Text" oder **.txt** bezeichnet.

2. *Wählen Sie Texte aus*, markieren Sie die Textdatei, und bevor Sie mit *OK* bestätigen, klicken Sie auf *anzeigen*. Enthält die Datei eigenartige Symbole? Lösung: Ändern Sie die *Texteinstellungen*; ändern Sie einen DOS-Zeichensatz zu einem Windows-Zeichensatz oder umgekehrt. Der Text war in [ASCII](#) kodiert, aber **Oxford WordSmith Tools** hat fälschlicherweise ANSI erkannt.

3. Eigenartige Symbole in einer Wortliste können auch Ursache von Rechtschreibfehlern in den originalen Textdateien sein.

Griechisch, Russisch, etc.

4. Wenn der Text auf Russisch, Griechisch, etc. ist, benötigen Sie eine geeignete Schriftart, die Sie z. B. für Windows 95 über Microsoft Plus erhalten.

5. Wenn Sie mehrere Listen geöffnet haben, in denen *verschiedene* Zeichensätze verwendet werden, und Sie verändern die [Schriftart](#) oder [Texteigenschaften](#), werden die Listen aktualisiert, um die aktuelle Schriftart und den aktuellen Zeichensatz anzuzeigen, es sei denn Sie minimieren zuvor die Fenster, die von den Änderungen betroffen sind.

eigenartige Symbole, wenn WordSmith-Daten in einer anderen Anwendung geöffnet werden

Oxford WordSmith Tools kann "[Speichern](#)" oder "Speichern als" und "[Als Text speichern](#)", indem in eine Datei [gedruckt](#) wird. Mit "Speichern" und "Speichern als" wird die Datei in einem speziellen Format gespeichert, um die Daten mit **Oxford WordSmith Tools** wieder verwenden zu können. Dieses Format ist nicht für ein Textverarbeitungsprogramm geeignet. Die Idee dahinter ist, dass Sie Ihre Arbeit speichern und an einem anderen Tag fortsetzen können. "Als Text speichern" andererseits bedeutet, die Ergebnisse als Standardtext zu speichern, indem Sie in eine Datei "gedruckt" werden. Diese Funktion ist nützlich, wenn Sie aus **WordSmith** nicht auf Papier drucken möchten, sondern die Daten in eine Tabelle oder ein Textverarbeitungsprogramm (wie z. B. **Microsoft Word**) einfügen möchten. Es ist normalerweise schneller, den ausgewählten Text in die [Zwischenablage](#) zu kopieren.

13.6 Ergebnisse für Concord und WordList unterschiedlich

Ergebnisse für Concord und WordList unterschiedlich

Es ist möglich, dass in **WordList** eine bestimmte Anzahl eines [\(Wortlisten\)-Clusters](#) gefunden wird, aber in **Concord** eine andere Anzahl. Dies geschieht aufgrund der unterschiedlichen Berechnungsverfahren. WordList verarbeitet Texte Wort für Wort und ignoriert Interpunktion (außer Bindestriche und Apostrophe). Wenn **Concord** nach einem [\(Konkordanz\)-Cluster](#) sucht, wird die Interpunktion berücksichtigt.

13.7 Liste häufig gestellter Fragen

Siehe auch: [Protokollierung](#).

Dies sind häufig gestellte Fragen (FAQs).

Es gibt eine sehr viel längere Liste von Erklärungen unter [Fehlermeldungen](#).

[Apostrophe werden nicht verarbeitet](#)

[Ist das Russisch, Griechisch oder Englisch? Es werden eigenartige Symbole angezeigt](#)

[Das Programm ist abgestürzt](#)

[Das Programm startet nicht!](#)

[Es dauert viel zu lange!](#)

[Tasten reagieren nicht](#)

[Zeile überschreitet das Demolimit](#)

[Falsche Übereinstimmung zwischen Ergebnissen aus Concord und WordList](#)

[Keine Tags sichtbar in Konkordanz](#)

[Druckproblem](#)
[Text ist nicht lesbar wegen der Farben](#)
[Zu viel oder zu wenig Platz zwischen den Spalten](#)
[Wortliste nicht sortiert](#)
[Won't slice pineapples](#)

13.8 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.)

13.9 Spaltenabstand

Der Spaltenabstand ist falsch.

Sie können den Abstand verändern, indem Sie auf die Schaltfläche [Layout](#) klicken.

13.10 startet nicht

Das Programm startet nicht!

Unerwarteter Fehler.

13.11 Tagproblem in Concord

keine Tags sichtbar in Konkordanz

Wenn Sie in **Concord** *Nächstes Tag* ausgewählt haben, und in Ihrer Konkordanz dennoch keine Tags angezeigt werden, haben die [zu ignorierenden Tags](#) wahrscheinlich dasselbe Format. Zum Beispiel, wenn unter *Tags ignorieren* <*> festgelegt ist, werden alle Tags wie <Titel>, <Zitat>, etc. bei der Konkordanzsuche ignoriert, es sei denn, Sie legen diese Tags in einer [Tagdatei](#) fest. Lösung: Geben Sie eine Tagdatei an und berechnen Sie die Konkordanz neu.

13.12 Tasten reagieren nicht

Tasten reagieren nicht

Wenn beim Drücken einer Taste nichts passiert, ist wahrscheinlich das falsche Fenster aktiv. Windows ist so gestaltet, dass Benutzer verschiedene Programme gleichzeitig auf demselben Bildschirm ausführen können, sodass jedes Fenster anders auf verschiedene Tastenkombinationen reagiert. Sie können sehen, welches Fenster gerade aktiv ist, da die Titelleiste sich farblich von den anderen geöffneten Fenstern unterscheidet. Sie können das Problem lösen, indem Sie das Fenster anklicken und dann die gewünschte Taste drücken.

13.13 unleserliche Farben

Text ist aufgrund der Farben nicht lesbar

Lösung: Wählen Sie *Einstellungen | Farben*. Sie können hier die Farben Ihrem Computerbildschirm entsprechend festlegen. Es gibt auch Einstellungen für Schwarz-Weiß-Bildschirme.

13.14 Wortliste nicht sortiert

Wortliste nicht sortiert

Wörter werden entsprechend der Microsoft-Routinen sortiert, die sprachabhängig sind. Wenn Sie Texte auf Spanisch verarbeiten, aber die Spracheinstellungen auf "Englisch" festgelegt sind, werden Sie Ergebnisse erhalten, die nicht Spanisch sortiert sind (z. B. **LL** kommt direkt vor **LM**).
Lösung: Wählen Sie die richtige [Sprache](#) aus und berechnen Sie die Wortliste neu.

13.15 zu langsam

Es dauert viel zu lange

Wenn Sie viel Text verarbeiten und einen alten PC mit geringem Arbeitsspeicher besitzen, kann die Verarbeitung sehr lange dauern. Sie werden ein Rattern von der Festplatte hören, wenn der [Arbeitsspeicher](#) knapp ist. Lösung: Sie können die Geschwindigkeit Ihres Computers erhöhen, indem Sie zusätzlichen Arbeitsspeicher installieren (*großer* Unterschied!), die Festplatte defragmentieren, einen Datenträgercache verwenden oder die Einstellungen Ihres virtuellen Datenträgers anpassen. Wenn Sie **Oxford WordSmith Tools** auf einem Netzwerk ausführen, fragen Sie Ihren Netzwerkadministrator, ob die Geschwindigkeit aufgrund des Netzwerkzugangs so niedrig ist.

Zweite Lösung: Beenden Sie alle Programme, die Sie gerade nicht benötigen. Dadurch kann unter Umständen die Kapazität des Systemspeichers verbessert werden.

Dritte Lösung: Beenden Sie Windows und starten Sie es erneut. Dadurch kann unter Umständen die Kapazität des Systemspeichers verbessert werden.

Vierte Lösung: Speichern und von der lokalen Festplatte neu laden, nicht vom Netzwerk.

Fehlermeldungen

Section



XIV

14 Fehlermeldungen

14.1 .ini-Datei nicht gefunden

.ini-Datei nicht gefunden

Beim Starten sucht **WordSmith** nach der Datei **wshell.ini**, in der die aktuellen [Standardeinstellungen](#) gespeichert sind. Sollten Sie diese Datei verschoben oder umbenannt haben, muss sie wiederhergestellt werden. Die Datei musste in demselben Verzeichnis wie die Tools gespeichert werden.

14.2 anderes Tool kann nicht aufgerufen werden

Can't call other Tool

Inter-Tool communication has got disrupted. [Save](#) your work, first. Then, if necessary, close down **Oxford WordSmith Tools** altogether, then start the main **wordsmith.exe** program again.

14.3 Arbeitsspeicher ist knapp

Arbeitsspeicher ist knapp!

Der Vorgang konnte nicht ausgeführt werden, weil der Arbeitsspeicher ([RAM](#)) knapp ist.

14.4 auf die Datei konnte nicht zugegriffen werden

Auf die Datei konnte nicht zugegriffen werden

Auf die Datei wird möglicherweise schon von einem anderen Programm zugegriffen. Die meisten Textverarbeitungsprogramme markieren geöffnete Textdateien mit "in Verwendung" und diese Dateien lassen sich von anderen Programmen nicht öffnen, auch nicht schreibgeschützt. Schließen Sie die Textdatei in Ihrem Textverarbeitungsprogramm.

14.5 auf Textdatei kann nicht zugegriffen werden – wird sie anderweitig verwendet?

Kein Zugriff auf Textdatei möglich: wird sie anderweitig verwendet?

Auf die Datei konnte nicht zugegriffen werden. Sie wird vielleicht von einer anderen Anwendung verwendet. Wenn das der Fall ist, schließen Sie die Datei in dieser Anwendung und versuchen Sie es erneut.

14.6 Basislistenfehler

base list error

WordSmith is trying to access a word or concordance line above or below the top or bottom of the data computed. This is a bug.

14.7 Bestätigungsmeldungen: Datei neu einlesen?

Datei neu einlesen?

Dies ist eine Bestätigungsmeldung. Um fortzufahren liest **Viewer** die Datei von der Festplatte neu ein. Davon werden alle Änderungen beeinflusst, die Sie bereits an der angezeigten Datei vorgenommen haben. Sie sollten diese Datei zuerst speichern und es danach erneut versuchen. Außerdem wird Viewer versuchen, die gesamte Textdatei zu lesen (bis zu 16.368 Sätzen).

Befindet sich eine große Datei auf einem langsamen CD-ROM-Laufwerk, so kann das eine Weile dauern.

14.8 bricht bei xx Wörtern ab -- Tagdatei hat mehr!

Die Anzahl der Einträge in der Tagdatei liegt über dem aktuellen Limit. Oder es ist gar keine Tagdatei!

14.9 can't merge list with itself!

Can't merge list with itself

You can only merge 1 word list or key word database with 1 other at a time. Select (by clicking while holding down the Control key) 2 file-names in the list of files.

14.10 Datei enthält keines der festgelegten Tags

File contains none of the tags specified

You specified tags, but none of them were found.

14.11 Datei hat "Löcher"

File has "holes"

Your text file is defective. It may well contain useful text, but it also contains at least one unrecognised character such as character(0). The problem could have arisen because it was transferred from one system to another, part of the disk is corrupted, or else maybe the file contains unrecognised graphics, or else it is not a plain text file but e.g. a [Word document](#). You will see the context where the problem occurred and will be told roughly how far into the text it was detected.

WordSmith can proceed if you wish but you get a chance to skip the text.

You can solve this problem -- which will come each time you choose that text file -- by reading the text file into a word processor and re-saving it as a plain .txt file. Also, in [File Utilities](#) there is a tool for finding such files.

14.12 Datei konnte nicht gelesen werden

Can't read file

If this happens when starting up **Oxford WordSmith Tools**, there is probably a component file missing. One example is **sayings.txt**, which holds sayings that appear in the main [Controller](#) window. If you've deleted it, I suggest you use **notepad** to start a new **sayings.txt** and put one blank line in it.

If you get this message at another time, something has gone wrong with a disk reading operation. The file you're trying to read in may be corrupted. This happens easily if you often handle very large files, especially if it's a long time since you last ran *Scandisk* to check whether any clusters in your files have got lost. See your DOS or Windows manual for help on fragmentation.

14.13 Datei konnte nicht gelesen werden

Nicht lesbar

Das passiert möglicherweise, weil das Ablagesystem auf Ihrem Speicherplatz defekt ist. Besonders häufig tritt dieser Fehler aus, wenn Sie große Dateien in Ihrem Textverarbeitungsprogramm bearbeiten. In diesem Fall ist es empfehlenswert, zuerst die

Festplatte unter *Eigenschaften* auf *Fehler* zu prüfen.

14.14 Datei konnte nicht gespeichert werden

Nicht gespeichert

Das passiert möglicherweise, wenn Sie entweder versuchen, das Verzeichnis in ein schreibgeschütztes Verzeichnis im Netzwerk zu speichern, oder wenn Ihr Speicherplatz voll oder das Ablagesystem auf Ihrer Festplatte defekt ist. Letzteres kommt sehr häufig vor, insbesondere wenn Sie große Dateien in Ihrem Textverarbeitungsprogramm bearbeiten. In diesem Fall ist es empfehlenswert, zuerst die Festplatte unter *Eigenschaften* auf *Fehler* zu prüfen. Wenn Sie in einem Netzwerk arbeiten, können Sie in manchen Laufwerken und Verzeichnissen [speichern](#) und in anderen nicht. Sie können das Problem lösen, indem Sie entweder auf dem Laufwerk A: speichern oder auf einem Festplattenlaufwerk, auf dem Sie Schreibrechte besitzen.

14.15 Datei nicht gefunden

Datei wurde nicht gefunden

Diese Meldung, sowie die Meldung [Originaltext nicht gefunden](#), erscheint, wenn **WordSmith** auf den Originalquelltext zugreifen muss, mit dem eine Liste erstellt wurde, ihn aber nicht finden kann. Wurde der Originaltext gelöscht oder verschoben? Ist die Datei noch vorhanden, kann man im Fenster "Dateiname" () in dieser Liste den [Dateinamen bearbeiten](#). Diese Meldung kann auch erscheinen, wenn Sie den Dateinamen selbst eingegeben haben. Möglicherweise haben Sie den Dateinamen falsch eingegeben. Ist es ein [langer Dateiname](#) unter Windows 95 or NT? Bei der Eingabe eines [Dateinamens](#) müssen sowohl der vollständige Pfadname als auch der Dateiname angegeben werden.

14.16 Dateinamen dürfen nicht identisch sein!

Dateinamen dürfen nicht identisch sein

Eine Datei kann nicht mit sich selbst verglichen werden.

14.17 Datenträgerproblem -- Datei nicht gespeichert

Disk problem: File not saved

Something has gone wrong with a disk writing operation. Perhaps there's not enough room on the drive. If so, delete some files on that drive.

14.18 Details für Matchliste nicht angegeben.

Details für Matchliste nicht angegeben.

Sie haben die Schaltfläche [Matchliste](#) gedrückt, aber dann keine gültige Matchliste ausgewählt oder keine Filtervorlage eingegeben. Versuchen Sie es erneut.

14.19 dies ist eine Demoversion

Dies ist eine Demoversion

Wahrscheinlich möchten Sie eine Vollversion. Dazu müssen Sie Ihre Demoversion [registrieren](#).

14.20 diese Funktion ist noch nicht bereit

Diese Funktion ist noch nicht bereit!

Temporäre Meldung für Funktionen, die sich noch in der Testphase befinden.

14.21 diese Meldung nicht mehr anzeigen...

Registrieren Sie Ihre Demoversion. Dies ist eine "Registrierungsaufforderung" für die [Demoversion](#).

14.22 Drucker notwendig

WordSmith benötigt für die Installation einen Druckertreiber, auch wenn Sie nie etwas ausdrucken möchten. Dazu müssen Sie keinen Drucker kaufen oder den Drucker anschalten. Die Funktion [Druckvorschau](#) in Concord, WordList, KeyWords, etc. benötigt lediglich Informationen bezüglich des Papierformats. Wenn Sie die Meldung erhalten, dass kein Drucker installiert wurde, wählen Sie in Windows Start | Einstellungen | Drucker und Faxgeräte und installieren Sie einen Standarddrucker (egal welchen).

14.23 Format der Lizenznummer nicht korrekt

Die Lizenznummer setzt sich wie folgt zusammen:

2 Buchstaben, ein Punkt, dann 4 Ziffern, ein Punkt, 4 Ziffern, usw.
Beispiel: XX.1234.5678.9012.3456 (Punkt nach jedem 4. Buchstaben)

14.24 Formular unvollständig

Formular unvollständig

Sie haben versucht ein Formular zu schließen, in dem eines oder mehrere leere Felder ausgefüllt werden müssen, bevor **WordSmith** fortfahren kann.

14.25 Funktion kann noch nicht vollständig ausgeführt werden

Funktion kann noch nicht vollständig ausgeführt werden

Diese Funktion befindet sich noch in der Entwicklung und kann noch nicht vollständig ausgeführt werden.

14.26 Grenze bei Zusammenführung erreicht

Grenze bei Zusammenführung erreicht: Versuchen Sie erneut zusammenzuführen.

Es kann nur eine bestimmte Anzahl von Wörtern in einem Arbeitsschritt [lemmatisiert](#) werden.

Wenn Sie die Grenze erreicht haben und diese Meldung erhalten,

1. drücken Sie F4 um zu lemmatisieren,
2. heben Sie den Haupteintrag erneut hervor,
3. drücken Sie F5 und setzen Sie die Lemmatisierung so mit jedem Eintrag fort, den Sie zum Haupteintrag hinzufügen möchten,
4. drücken Sie danach F4 um die Wörter zusammenzuführen.

14.27 Internetverbindung konnte nicht hergestellt werden

Internetverbindung konnte nicht hergestellt werden.

Diese Funktion setzt voraus, dass a) Sie einen Internetbrowser auf Ihrem Computer installiert haben und b) dass Ihr System die Internet-URL-Endung **.htm** mit diesem Browser "assoziiert".

14.28 kein Platz für berechnete Variable

Kein Platz für berechnete Variable

Es ist nicht genügend Platz für die Variable, die Sie berechnen möchten.

14.29 keine aktuelle WordSmith-Datei

Keine aktuelle WordSmith-Datei

Die Datei, die Sie öffnen möchten, wurde zwar mit WordSmith erstellt, aber entweder

- die Datei wurde mit einer früheren Version (1-3) erstellt oder
- die Datei wurde mit einer WordSmith 4 Beta-Version erstellt und das Format musste leider verändert werden.

Wurde die Datei mit einer der Versionen 1-3 erstellt, können Sie die Datei möglicherweise mit dem [Converter](#) konvertieren.

14.30 Keine assoziierten Gruppen identifiziert

Keine assoziierten Gruppen identifiziert.

Verändern Sie die Einstellungen und versuchen Sie es erneut.

14.31 keine assoziierten Wörter gefunden

Keine assoziierte Dateien gefunden.

Verändern Sie in den *Einstellungen* Mindest- und Höchstfrequenzen, und versuchen Sie es danach erneut.

14.32 keine Cluster gefunden

Keine Cluster gefunden

Verändern Sie in den *Einstellungen* die Clustereinstellungen und versuchen Sie es erneut.

Entweder die Anzahl der Konkordanzlinien ist zu gering, um die Mindestanzahl zu finden, oder die Cluster-Länge ist zu hoch.

14.33 keine Einträge in der KeyWords-Datenbank

Keine Einträge in der KeyWords-Datenbank

Verändern Sie die Einstellungen und versuchen Sie es erneut.

14.34 keine gelöschten Zeilen zu entfernen

Keine gelöschten Zeilen zu entfernen

Sie haben Alt+Z gedrückt, aber keine Zeilen zu [entfernen](#). Es ist alles in Ordnung.

14.35 keine Kollokatoren gefunden

Keine Kollokatoren gefunden

Verändern Sie im [Controller](#) die Einstellungen (*Einstellungen anpassen* | *Concord* | *Mindestfrequenz*) und versuchen Sie es erneut. Die Anzahl der Konkordanzlinien war zu gering, um die Mindestanzahl zu finden.

14.36 keine Konkordanzeinträge

No concordance entries found

If you got no concordance entries, either a) there really aren't any in your text(s), b) there's a problem with the specification of what you're seeking, or c) there's a problem with the text selection. Check how you've spelt the search-word and context word. If you're using [accented text](#), check the format of your texts. If you're using a [search-word file](#), ensure this was prepared using a plain Windows word-processor such as Notepad.

Have you specified any [wildcards](#) (* and ?) accurately? If you are looking for a question-mark, you may have put "?" correctly but remember that question-marks usually come at the ends of words, so you will need "*?".

Tip

Bung in an [asterisk](#) or two. You're more likely to find book* than book.

14.37 keine Schlüsselwörter für Plot

Keine Schlüsselwörter für Plot

Haben Sie alle Schlüsselwörter gelöscht?

14.38 keine Schlüsselwörter gefunden

Keine Schlüsselwörter gefunden

Verändern Sie die Einstellungen und versuchen Sie es erneut. Entweder die Mindestfrequenz ist zu hoch und/oder der [P-Wert](#) ist zu niedrig festgelegt, um Schlüsselwörter zu finden. Für sehr kurze Texte ist möglicherweise eine Mindestfrequenz von 2 notwendig.

14.39 keine solchen Dateien gefunden

Keine entsprechende(n) Datei(en) gefunden

Sie haben den Namen einer nicht vorhandenen Datei eingegeben. Bei der Eingabe eines [Dateinamens](#) müssen sowohl der vollständige Pfadname als auch der Dateiname angegeben werden.

14.40 keine Statistik verfügbar

Keine Statistik verfügbar

Einige Arten von Wortlisten, die von **Oxford WordSmith Tools** erstellt wurden, z. B. die Wortliste einer Schlüsselwort-Datenbank, beinhalten Wörter in alphabetischer Reihenfolge oder nach Frequenz geordnet, jedoch keine Statistikinformationen der originalen Textdateien. Aus diesem Grund können Sie die Statistik in **WordList** nicht aufrufen. Möglicherweise erhalten Sie diese Meldung auch, wenn die Statistikdatei, die Sie aufrufen möchten, beschädigt ist.

14.41 keine WordSmith-Datei

Die Datei, die Sie öffnen möchten, wurde nicht mit WordSmith erstellt. WordSmith erstellt Dateien, die Ergebnisse enthalten, mit den Dateierweiterungen **.LST**, **.CNC**, **.KWS**, etc. Dies sind spezielle WordSmith-Dateiformate, die nicht von Microsoft Word geöffnet werden können. Umgekehrt kann WordSmith eine [Standardtext-Datei](#) oder eine Word **.doc**-Datei normalerweise nicht als Datendatei sondern nur als Textdatei zur Verarbeitung erkennen.

Siehe auch: [Daten aus früheren Versionen konvertieren](#)

14.42 keine Wörter aus KeyWords-Stoppliste

Keine Wörter aus KeyWords-Stoppliste

WordSmith konnte Ihre Stopplistendatei nicht lesen, oder die Datei war leer.

14.43 keine Wörter aus Konkordanz-Stoppliste

Keine Wörter aus Konkordanz-Stoppliste

14.44 keine Wörter aus Lemmaliste

Keine Wörter aus Lemmaliste

WordSmith konnte Ihre Lemmalistendatei nicht lesen, oder die Datei war leer.

14.45 keine Wörter aus Matchliste

Keine Wörter aus Matchliste

WordSmith konnte Ihre [Matchlisten](#)datei nicht lesen, oder die Datei war leer, oder Sie haben vergessen, die gewünschte Aktion zu aktivieren (eine der Optionen lautet *Keine*). Oder aber es sind nur Zahlen in der aktuellen Spalte, die Sie mit einer Wortliste oder einer Vorlage übereinstimmen möchten. Möglicherweise gibt es auch keines der von Ihnen festgelegten Wörter!

14.46 keine Wörter aus Stoppliste

Keine Wörter aus Stoppliste

WordSmith konnte Ihre Stopplistendatei nicht lesen, oder die Datei war leer.

14.47 keine Wörter aus Tagliste

Keine Wörter aus Tagliste

WordSmith konnte Ihre Tagdatei nicht lesen, oder die Datei war leer.

14.48 keine Wortlisten ausgewählt

Keine Wortlisten ausgewählt

Damit **WordSmith** weiß, welche Wortlisten verglichen werden sollen, müssen Sie diese Wortlisten auswählen, indem in jedem Verzeichnis auf eine Wortliste klicken. Wenn den Vorgang abbrechen wollen, klicken Sie auf "Abbrechen".

14.49 KeyWords-Datei ist fehlerhaft

Schlüsselwortdatei ist fehlerhaft.

Jede Datei, die von **Oxford WordSmith Tools** erstellt wird, hat je nach Typ ihre eigene vorgegebene Dateierweiterung (z. B. **.KWS**, **.KDB**) und eine eigene interne Struktur. Dateien mit derselben Erweiterung, die von einem anderen Programm erstellt wurden, sind nicht mit **WordSmith** kompatibel. Es ist nicht zielführend, eine **.KWS**-Datei zu **.TXT** umzubenennen oder umgekehrt. **WordSmith** erkennt, dass die Datei, die Sie öffnen wollen, nicht von der aktuellen **KeyWords**-Version erstellt wurde.

14.50 KeyWords-Datenbank ist fehlerhaft

Schlüsselwort-Datenbank ist fehlerhaft.

Jede Datei, die von **Oxford WordSmith Tools** erstellt wird, hat je nach Typ ihre eigene vorgegebene Dateierweiterung (z. B. **.KDB**, **.KWS**) und eine eigene interne Struktur. Dateien mit derselben Erweiterung, die von einem anderen Programm erstellt wurden, sind nicht mit **WordSmith** kompatibel. Es ist nicht zielführend, eine **.KDB**-Datei zu **.TXT** umzubenennen oder umgekehrt. **WordSmith** erkennt, dass die Datei, die Sie öffnen wollen, nicht von der aktuellen **KeyWords**-Version für eine Schlüsselwort-Datenbank erstellt wurde.

14.51 Konkordanzdatei ist fehlerhaft

Concordance file is faulty

Each type of file created by **Oxford WordSmith Tools** has its own default filename extension (e. g. **.CNC**, **.LST**) and its own internal structure. If you have another file with the same extension produced by another program, this will not be compatible. It would not be sensible to rename a **.CNC** file to **.TXT**, or vice-versa! **WordSmith** has detected that the file you're calling up wasn't produced by the current version of **Concord**.

14.52 Konkordanz-Stopplistendatei nicht gefunden

Concordance stop list file not found

You typed in the name of a non-existent file. If typing in a [filename](#), remember to include the full drive and folder as well as the filename itself.

14.53 Konvertierungsdatei nicht gefunden

Konvertierungsdatei nicht gefunden

Sie haben den Namen einer nicht vorhandenen Datei eingegeben. Bei der Eingabe eines [Dateinamens](#) müssen sowohl der vollständige Pfadname als auch der Dateiname angegeben werden.

14.54 Laufwerk ungültig

Drive not valid

WordSmith is unable to access this drive. This could happen if you attempt to access a disk drive which doesn't exist, e.g. drive P: where your drives include A:, C:, D: and E:.

14.55 Limit für dateibasierte Suchwörter erreicht

Limit für dateibasierte Suchwörter erreicht

Es können nicht mehr als 15 Suchwörter auf einmal verarbeitet werden, es sei denn Sie verwenden eine [Datei mit Suchwörtern](#), die Sie in **Concord** als Batch gruppiert mit einem Limit von 500 verarbeiten können.

14.56 Liste der Fehlermeldungen

Liste der Fehlermeldungen

Siehe auch: [Fehlerbehebung](#).

[Wörter können nur als ASCII gespeichert werden.](#)

[Anderes Tool kann nicht aufgerufen werden.](#)

[Verzeichnis kann nicht erstellt werden, da bereits ein Verzeichnis mit diesem Namen vorhanden ist.](#)

[Listen können nicht zusammengeführt werden.](#)

[Die Datei kann nicht gelesen werden.](#)

[Zeichensatz an Sprache \(<language>\) anpassen und auf <x> zurücksetzen.](#)

[Konkordanzdatei ist fehlerhaft](#)

[Konkordanz-Stopplisten-datei nicht gefunden](#)

[Konvertierungsdatei nicht gefunden](#)

[Zielverzeichnis nicht gefunden](#)

[Laufwerkproblem: Datei wurde nicht gespeichert.](#)

[Verteilung muss mit Konkordanz gespeichert werden.](#)

[Laufwerk ungültig](#)

[Internetverbindung konnte nicht hergestellt werden.](#)

[Neues Verzeichnis mit diesem Namen konnte nicht erstellt werden.](#)

[Auf die Datei konnte nicht zugegriffen werden](#)

[Datei enthält keines der festgelegten Tags.](#)

[Datei wurde nicht gefunden](#)

[Dateinamen dürfen nicht identisch sein.](#)

[Vollständiger Pfadname \(Laufwerk:\Verzeichnis\) wird benötigt.](#)

[Funktion kann noch nicht vollständig ausgeführt werden](#)

[INI-Datei wurde nicht gefunden.](#)

[Konkordanzdatei ist ungültig.](#)

[Dateiname ist ungültig](#)

[KeyWords-Datenbank ist ungültig.](#)

[KeyWords-Datei ist ungültig.](#)

[Wortlisten-Vergleichsdatei ist ungültig.](#)

[WordList-Datei ist ungültig.](#)

[Grenze bei Zusammenführung erreicht: zusammenzuführen und neu versuchen.](#)

[Schlüsselwortdatei ist fehlerhaft.](#)

[Schlüsselwort-Datenbank ist fehlerhaft.](#)
[Limit für dateibasierte Suchwörter \(500\) erreicht.](#)
[Schnittstelle zwischen den Tools unterbrochen.](#)
[Details für Matchliste nicht angegeben.](#)
[Muss eine Zahl sein](#)
[Lizenznummer für Netzwerk bereits in Verwendung oder umgekehrt](#)
[Kein Zugriff auf Textdatei möglich: wird sie anderweitig verwendet?](#)
[Keine assoziierte Dateien gefunden.](#)
[Keine assoziierten Gruppen identifiziert.](#)
[Keine Cluster gefunden](#)
[Keine Kollokatoren gefunden](#)
[Keine Konkordanzeinträge gefunden.](#)
[Keine Wörter aus Konkordanz-Stoppliste](#)
[Keine gelöschten Zeilen zu entfernen](#)
[Keine Einträge in der KeyWords-Datenbank](#)
[Keine Schlüsselwörter gefunden](#)
[Keine Schlüsselwörter für Plot](#)
[Keine Wörter aus KeyWords-Stoppliste](#)
[Keine Wörter aus Lemmaliste.](#)
[Keine Wörter aus Matchliste](#)
[Kein Platz für berechnete Variable](#)
[Keine Statistik verfügbar](#)
[Keine Wörter aus Stoppliste](#)
[Keine entsprechende\(n\) Datei\(en\) gefunden](#)
[Keine Wörter aus Tagliste](#)
[Nummer ist ungültig.](#)
[Keine Wortlisten ausgewählt.](#)
[Zugriff auf die originale Textdatei nicht möglich, da nicht vorhanden.](#)
[Lizenznummer ist nicht korrekt.](#)
[Lizenznummer muss aus 20 Zeichen bestehen.](#)
[Arbeitsspeicher ist knapp!](#)
[Quelldatei\(en\) nicht gefunden](#)
[Stopplistendatei nicht gefunden](#)
[Stopplistendatei nicht eingelesen](#)
[Tagdatei nicht gefunden.](#)
[Taglistendatei konnte nicht eingelesen werden](#)
[Diese Funktion ist noch nicht bereit!](#)
[Dies ist eine Demoversion](#)
[Dieses Programm funktioniert nur unter Windows 95 oder höher.](#)
[Störende Meldung "Demoversion registrieren" bei Programmstart nicht mehr erhalten.](#)
[Grenze für zu ignorierende Elemente \(50\) überschritten.](#)
[Satzgrenze \(8000\) überschritten.](#)
[Zwei Dateien sind notwendig.](#)
[Bricht bei xx Wörtern ab -- Tagdatei hat mehr!](#)
[Schlüsselwortdatenbanken konnten nicht zusammengeführt werden.](#)
[Wieso ist die Suche fehlgeschlagen?](#)
[Wortlistendatei nicht gefunden.](#)
[Wortlisten-Vergleichsdatei ist fehlerhaft.](#)
[Wortlistendatei ist fehlerhaft.](#)
[Oxford WordSmith Tools ist abgelaufen: neue Version laden](#)
[Oxford WordSmith Tools ist bereits geöffnet.](#)
[WordSmith-Versionen stimmen nicht überein](#)
[XX Tage verbleiben.](#)

14.57 Lizenznummer für Netzwerk wird bereits verwendet

Lizenznummer für Netzwerk bereits in Verwendung oder umgekehrt

Der Lizenznummer für die Verwendung in einem Netzwerk gilt nicht für die Verwendung auf einem einzelnen Rechner und umgekehrt. Wenn Sie diese Meldung erhalten, installieren Sie die Software bitte dementsprechend neu.

14.58 Lizenznummer ist nicht korrekt

Lizenznummer ist nicht korrekt

Die Lizenznummer stimmt nicht mit den Anforderungen der aktualisierten Vollversion überein! Die alte Lizenznummer von früheren Versionen wird nicht mehr verwendet. **WordSmith** funktioniert zwar noch, aber nur im [Demomodus](#).

14.59 muss eine Zahl sein

Muss eine Zahl sein

Das, was Sie eingetippt haben, ist keine Zahl. Vorsicht besonders bei kleingeschriebenem **L** und **1** sowie dem Buchstaben **O** anstelle der Zahl **0**.

14.60 neues Verzeichnis mit diesem Namen konnte nicht erstellt werden

Neues Verzeichnis konnte nicht erstellt werden

Ein Verzeichnis und eine Datei können nicht den gleichen Namen haben. Gibt es bereits eine *Datei* mit dem Namen C:\TEMP\FRED, ist es nicht möglich, ein *Unterverzeichnis* in C:\TEMP mit dem Namen FRED zu erstellen. Wählen Sie einen anderen Namen.

14.61 nichts aktiviert

Nichts aktiviert

In einigen Formularen gibt es Optionen "Aktiviert", die aktiviert bzw. deaktiviert werden können. Sind diese Optionen nicht aktiviert, werden sie zwar angezeigt, von **WordSmith** aber ignoriert.

14.62 nur für die Nutzung am PC mit dem Namen XXX

Nur für die Nutzung am PC mit dem Namen XXX

Die Software wurde für die Nutzung auf einem anderen PC registriert. Wenn Sie diese Meldung erhalten, installieren Sie die Software bitte dementsprechend neu.

14.63 Programm funktioniert nur unter Windows 98 oder höher

Dieses Programm funktioniert nur unter Windows 98 oder höher.

Seit der WordSmith-Version 4.0 ist dies ein 32-Bit-Programm (und eine 32-Bit-Hilfedatei).

14.64 Quelldatei(en) nicht gefunden

Quelldatei(en) nicht gefunden

Sie haben den Namen einer nicht vorhandenen Datei eingegeben. Bei der Eingabe eines [Dateinamens](#) müssen sowohl der vollständige Pfadname als auch der Dateiname angegeben werden.

14.65 Schlüsselwortdatenbanken konnten nicht zusammengeführt werden.

Wahrscheinlich war nicht genügend Arbeitsspeicher ([RAM](#)) vorhanden, um diesen Vorgang durchzuführen.

14.66 Schlüsselwörter können wegen unterschiedlicher Sprachen nicht berechnet werden

Can't compute key words as languages differ

Key words can only be computed if both the text file and the reference corpus are in the same primary language. You can compute KWs using 2 different varieties of English or 2 different varieties of Spanish, but not between English and French.

14.67 Schnittstelle zwischen den Tools unterbrochen

Schnittstelle zwischen den Tools unterbrochen.

Der Oxford WordSmith Tools [Controller](#) oder ein einzelnes Tool versucht erfolglos, ein anderes Tool aufzurufen. Das ist entweder auf einen Fehler in einem anderen geöffneten Programm oder zu wenig Arbeitsspeicher zurückzuführen. Da die [Schnittstellen](#) zwischen den Tools in dieser Programmgruppe von entscheidender Bedeutung sind, sollten Sie WordSmith beenden und erneut ausführen.

14.68 Stopplistendatei nicht eingelesen

Stopplistendatei nicht eingelesen

Beim Einlesen vom Datenträger ist ein Fehler aufgetreten. Die Datei ist möglicherweise beschädigt. Das passiert häufig, wenn Sie sehr große Dateien bearbeiten, speziell wenn Sie Ihren Datenträger schon längere Zeit nicht mehr auf *Fehler überprüft* haben, um zu kontrollieren, ob Cluster in Ihren Dateien verloren gegangen sind. Hilfe zu Fragmentierung finden Sie in ihrem DOS- oder Windows-Handbuch.

14.69 Stopplistendatei nicht gefunden

Stopplistendatei nicht gefunden

Sie haben den Namen einer nicht vorhandenen Datei eingegeben. Bei der Eingabe eines [Dateinamens](#) müssen sowohl der vollständige Pfadname als auch der Dateiname angegeben werden.

14.70 Tagdatei nicht eingelesen

Taglistendatei konnte nicht eingelesen werden

Beim Einlesen vom Datenträger ist ein Fehler aufgetreten. Die Datei ist möglicherweise beschädigt. Das passiert häufig, wenn Sie sehr große Dateien bearbeiten, speziell wenn Sie Ihren Datenträger schon längere Zeit nicht mehr auf *Fehler überprüft* haben, um zu kontrollieren, ob Cluster in Ihren Dateien verloren gegangen sind. Hilfe zu Fragmentierung finden Sie in ihrem DOS- oder Windows-Handbuch.

14.71 Tagdatei nicht gefunden

Tagdatei nicht gefunden

Sie haben den Namen einer nicht vorhandenen Datei eingegeben. Bei der Eingabe eines [Dateinamens](#) müssen sowohl der vollständige Pfadname als auch der Dateiname angegeben werden.

14.72 Transinformation inkompatibel

Transinformationsliste ist nicht kompatibel

Eine Transinformationsliste stammt aus einer Indexdatei, und beinhaltet die Information aus welcher Indexdatei sie berechnet wurde. Beim Öffnen einer Transinformationsliste wird auch die entsprechende Indexdatei geöffnet. Wenn die Indexdatei auf Ihrem PC nicht gefunden wird oder umbenannt wurde, erhalten Sie diese Meldung. Die Transinformation kann zwar aufgerufen werden, aber bezüglich Frequenz und alphabetischer Liste wird auf eine andere Indexdatei verwiesen und außerdem ist es nicht möglich, direkt aus dieser Liste eine Konkordanz zu erstellen..

14.73 ungültige KeyWords-Berechnung

Ungültige KeyWords-Berechnung

Um mit **KeyWords** die Schlüsselwörter in einer Textdatei durch Vergleich mit einem Referenzkorpus zu berechnen, müssen beide Dokumente in derselben Sprache verfasst und gleich sortiert sein (alphabetisch, aufsteigend). Wenn Sie diese Meldung erhalten, dann versuchen Sie Schlüsselwörter zu berechnen, die diese Kriterien nicht erfüllen. Lösung: Öffnen Sie jede Wortliste und prüfen Sie, ob sie funktioniert und aufsteigend alphabetisch sortiert ist. Überprüfen Sie, dass beide Dateien mit denselben Spracheinstellungen erstellt wurden, und berechnen Sie falls nötig eine der Dateien oder beide Dateien neu.

14.74 ungültige KeyWords-Datenbank

KeyWords-Datenbank ist ungültig.

Jede Datei, die von **Oxford WordSmith Tools** erstellt wird, hat je nach Typ ihre eigene vorgegebene Dateierweiterung (z. B. **.KWS**, **.KDB**) und eine eigene interne Struktur. Dateien mit derselben Erweiterung, die von einem anderen Programm erstellt wurden, sind nicht mit **WordSmith** kompatibel. Es ist nicht zielführend, eine **.KDB**-Datei zu **.TXT** umzubenennen oder umgekehrt. **WordSmith** erkennt, dass die Datei, die Sie öffnen wollen, nicht von der aktuellen **KeyWords**-Version für eine Datenbank erstellt wurde.

14.75 ungültige Konkordanzdatei

Konkordanzdatei ist ungültig.

Jede Datei, die von **Oxford WordSmith Tools** erstellt wird, hat je nach Typ ihre eigene vorgegebene Dateierweiterung (z. B. **.CNC**, **.LST**) und eine eigene interne Struktur. Dateien mit derselben Erweiterung, die von einem anderen Programm erstellt wurden, sind nicht mit **WordSmith** kompatibel. Es ist nicht zielführend, eine **.CNC**-Datei zu **.TXT** umzubenennen oder umgekehrt. **WordSmith** erkennt, dass die Datei, die Sie öffnen wollen, nicht von der aktuellen **Concord**-Version erstellt wurde.

14.76 ungültige Nummer

Nummer ist ungültig.

Entweder Sie haben gerade Daten eingegeben, oder **Oxford WordSmith Tools** wollte gerade Daten einlesen (z. B. aus **wshell.ini**, der Datei für [Standardeinstellungen](#)), die keine Buchstaben sondern nur Ziffern beinhalten dürfen. Der Großbuchstabe **O** wird von Computern nicht als Äquivalent zu der Ziffer **0** angesehen. Es ist auch möglich, dass die Zahl neben einem Buchstaben oder Symbol steht, z. B. **€ 30**. Erscheint diese Meldung gleich beim Start von **WordSmith**, überprüfen Sie bitte **wshell.ini** auf Fehler.

14.77 ungültige WordList-Datei

WordList-Datei ist ungültig.

Jede Datei, die von **Oxford WordSmith Tools** erstellt wird, hat je nach Typ ihre eigene vorgegebene Dateierweiterung (z. B. **.LST**, **.CNC**) und eine eigene interne Struktur. Dateien mit derselben Erweiterung, die von einem anderen Programm erstellt wurden, sind nicht mit **WordSmith** kompatibel. Es ist nicht zielführend, eine **.LST**-Datei zu **.TXT** umzubenennen oder umgekehrt. **WordSmith** erkennt, dass die Datei, die Sie öffnen wollen, nicht von der aktuellen **WordList**-Version erstellt wurde.

14.78 ungültige WordList-Vergleichsdatei

Wortlisten-Vergleichsdatei ist ungültig.

Jede Datei, die von **Oxford WordSmith Tools** erstellt wird, hat je nach Typ ihre eigene vorgegebene Dateierweiterung (z. B. **.LST**, **.CNC**) und eine eigene interne Struktur. Dateien mit derselben Erweiterung, die von einem anderen Programm erstellt wurden, sind nicht mit **WordSmith** kompatibel. Es ist nicht zielführend, eine **.CNC**-Datei zu **.TXT** umzubenennen oder umgekehrt. **WordSmith** erkennt, dass die Datei, die Sie öffnen wollen, nicht von **WordList** als Vergleichsdatei erstellt wurde.

14.79 ungültiger Dateiname

Dateiname ist ungültig

[Dateinamen](#) dürfen keine Leerzeichen oder bestimmte Zeichen, wie z. B. **?** und ***** enthalten.

Unter Windows bis Windows 95 waren Dateinamen sogar auf acht Buchstaben, einen Punkt und drei weitere Buchstaben begrenzt. Versuchen Sie es erneut.

14.80 Verteilung wird mit Konkordanz gespeichert

Dispersions go with concordances

They can't be [saved](#) separately.

14.81 Verzeichnis kann nicht erstellt werden, da bereits ein Verzeichnis mit diesem Namen vorhanden ist

Can't make folder as that's an existing filename

If you already have a *file* called C:\TEMP\FRED, you can't make a *sub-folder* of C:\TEMP called FRED. Choose a new name.

14.82 vollständiger Pfadname (Laufwerk:\Verzeichnis) wird benötigt

Vollständiger Pfadname (Laufwerk:\Verzeichnis) wird benötigt.

Bei der Eingabe eines [Dateinamens](#) müssen sowohl der vollständige Pfadname als auch der Dateiname angegeben werden.

14.83 Wieso ist die Suche fehlgeschlagen?

Mit der Standard-Suchfunktion (F12 oder ) für eine Datenliste wird die aktuell markierte Spalte durchsucht. Wenn Sie innerhalb von Daten in einer anderen Spalte suchen möchten, müssen Sie diese Spalte zuerst anklicken.

Standardmäßig werden bei einer Suche "ganze Wörter" gesucht. Verwenden Sie * am Ende oder am Anfang des Worts oder der Nummer, um eine erfolgreiche Suche durchzuführen, z. B. im Falle von Daten, die aus mehreren Wörtern bestehen. (Der Vorteil der Suche mit Sternchen besteht darin, dass Sie entweder einen Präfix, Suffix, oder beides festlegen können, im Gegensatz zu der Windows Standardsuchoption "nur ganzes Wort suchen".)

14.84 WordSmith Tools abgelaufen

Meldung nur für Benutzer mit zeitlicher begrenzter Lizenznummer. Ihre Version von **Oxford WordSmith Tools** ist nicht mehr gültig und läuft nun im [Demomodus](#). Laden Sie eine neue Version aus dem [Internet](#).

14.85 WordSmith wird bereits ausgeführt

Versuchen Sie nicht, **Oxford WordSmith Tools** zu starten, wenn das Programm bereits ausgeführt wird. Drücken Sie Alt+Tab, um zu der geöffneten WordSmith-Komponente zurückzukehren. (Sie können jedoch mehrere Kopien jedes Tools gleichzeitig ausführen.)

14.86 WordSmith-Versionen stimmen nicht überein

Da die verschiedenen Tools miteinander [verknüpft](#) sind, ist es wichtig sicherzustellen, dass alle Dateien der Komponenten miteinander kompatibel sind. Wenn Sie diese Meldung erhalten, ist mindestens eine Komponente anders datiert als die anderen.

Lösung: laden Sie die Komponenten, die Sie benötigen, von einer der Kontakt-[Websites](#) herunter.

14.87 Wörter können nur als ASCII gespeichert werden

Can only save WORDS as Plain Text

Oxford WordSmith Tools can't save graphics as a text file. If you get this error message, you can only save this type of data by copying to the [clipboard](#) and pasting it into your word-processor.

14.88 Wortlistendatei ist fehlerhaft

Jede Datei, die von **Oxford WordSmith Tools** erstellt wird, hat je nach Typ ihre eigene vorgegebene Dateierweiterung (z. B. **.LST**, **.KWS**) und eine eigene interne Struktur. Dateien mit derselben Erweiterung, die von einem anderen Programm erstellt wurden, sind nicht mit **WordSmith** kompatibel. Es ist nicht zielführend, eine **.LST**-Datei zu **.TXT** umzubenennen oder umgekehrt. **WordSmith** erkennt, dass die Datei, die Sie öffnen wollen, nicht von der aktuellen **WordList**-Version erstellt wurde.

14.89 Wortlistendatei nicht gefunden

Sie haben den Namen einer nicht vorhandenen Datei eingegeben. Bei der Eingabe eines [Dateinamens](#) müssen sowohl der vollständige Pfadname als auch der Dateiname angegeben werden.

14.90 Wortlisten-Vergleichsdatei ist fehlerhaft

Jede Datei, die von **Oxford WordSmith Tools** erstellt wird, hat je nach Typ ihre eigene vorgegebene Dateierweiterung (z. B. **.LST**, **.KWS**) und eine eigene interne Struktur. Dateien mit derselben Erweiterung, die von einem anderen Programm erstellt wurden, sind nicht mit **WordSmith** kompatibel. Es ist nicht zielführend, eine **.LST**-Datei zu **.TXT** umzubenennen oder umgekehrt. **WordSmith** erkennt, dass die Datei, die Sie öffnen wollen, nicht von **WordListals** Vergleichsdatei erstellt wurde.

14.91 XX Tage verbleiben

Meldung nur für Benutzer mit zeitlicher begrenzter Lizenznummer. Nach Ablauf des Zeitraums ist **WordSmith** nur noch im [Demo](#)modus verfügbar.

14.92 Zeichensatz an Sprache (<language>) anpassen und auf <x> zurücksetzen.

Character set reset to <x> to suit <language>

Prior to version 2.00.07, **Oxford WordSmith Tools** handled fewer [character sets](#) and [languages](#) than it does now. Accordingly, data saved in the format used before that version may not "know" what language it was based on. If you get this message when opening up an old **WordSmith** data file, it's because **WordSmith** doesn't know what language it derived from. Through gross linguistic imperialism, it will by default assume that the language is English! If the data are okay, just click the save button so that next time it will "know" which language it's based on. If not, reset the language to the one you want in the [Controller](#), *Adjust Settings | Text*, then re-save the list.

14.93 Zielverzeichnis nicht gefunden

Destination folder not found

WordSmith couldn't find that folder; perhaps it's mis-spelt.

14.94 zu viele Sätze

Die Grenze liegt bei 8.000. Erledigen Sie die Aufgabe stückweise.

14.95 zu viele übereinstimmende Gruppen zu ignorieren

Die Grenze liegt bei 50. Führen Sie alle übrigen Einträge manuell zusammen.

14.96 Zugriff auf die originale Textdatei nicht möglich, da nicht vorhanden

Zugriff auf die originale Textdatei nicht möglich, da nicht vorhanden

Um fortfahren zu können, muss **WordSmith** auf die originale Textdatei, auf der die Liste basiert, zugreifen. Diese wurde jedoch verschoben oder umbenannt.

Wenn Sie in einem Netzwerk arbeiten, ist möglicherweise die Netzwerkverbindung nicht zugeordnet oder ausgefallen. Unter Umständen ist auch nicht die richtige oder gar keine CD-ROM im Laufwerk.

14.97 zwei Dateien notwendig

Für dieses Verfahren benötigen Sie zwei Dateien. Wählen Sie zwei Dateinamen aus der Dateiliste aus (indem Sie Strg gedrückt halten).

Index

- " -

"range" 105

- . -

.ini-Dateien 48

- 2 -

2. Sortierung 78

25 Zeilen 198

- 3 -

32-bit-Version 170

- A -

Ablaufdatum 217

Absatz: Anfang und Ende 168

Absätze (festlegen) 124

Absätze: Definition 173

Abschnitt: Anfang und Ende 168

accents window 19

adjusting with mouse 131

advanced concordance settings 70

aktuelle Version überprüfen 17

Akzentzeichen 192

Akzentzeichen und Symbole 192

Alignment 130

Alignment zweier Texte mit Viewer 130

als HTML speichern 46

als Text speichern 46

als XML speichern 46

Alt+Tabulator 55

altering your data 33

alternative Suchwörter 80

andere Tools aufrufen 185

Anfang und Ende von Sätzen 168

Anführungszeichen 198

Anhalten 52

ansi 192

Apostrophe bei Sortierung 78

ascii 192

associate defined 85

associates 86

assoziierte Einträge 106

assoziierte Schlüsselwortgruppen neu gruppieren 94

assoziierte Schlüsselwortgruppen: neu gruppieren 94

Audiodatei-Tags 162

auf vorherige Ergebnisse zugreifen 45

Ausrichtung 41

aussteigen 45

Auswahl kopieren 33

automatische, dateibasierte Konkordanzerstellung 74

- B -

Babbage 200

Baltisch 24

Batch auswählen 86

batch processing: file-names 22

batch processing: folders 22

Batch-Verarbeitung 19

beenden 45

Behandlung von Tag-Arten 161

Beispiel 89

Beispiel Schlüsselwörter 89

benutzerdefinierte Einstellungen 48

benutzerdefinierte Kategorien 59

Berechnung in KeyWords 91

Bereich 104

Bestätigungsmeldungen: Datei neu einlesen? 203

bevorzugte Texte (Favouriten) 23

bibliography 170

Bindestriche 53

Bindestrich-Einstellungen 171

blanking out entries 58

BNC Sampler version 174

BNC: innerhalb von Texten auswählen 165

BNC: Tagdatei 166

BNC: Textformat 177

BNC: Zwischen Texten auswählen 164

BNC-Markierungen für Sätze und Überschriften 168
 boolean and/not 73
 Boolesche Funktion 'oder' 80
 bricht bei xx Wörtern ab 204
 browsing original 68
 Bugs 171

- C -

CD-ROM: Geschwindigkeit 176
 CD-ROM: Speicherplatz 184
 CD-ROM-Version: Standardeinstellungen 48
 changing from edit to type-in mode 183
 character sets 191
 characters in save as text 81
 Charles Babbage 200
 Cherokee 193
 Chinesisch 193
 Chi-Quadrat 91
 Choose Languages: overview 6
 choosing files from standard dialogue box 27
 class instructions 40
 clumps 86
 Cluster 175
 Cluster in KeyWords 87
 Cluster: Definition 173
 Cocoa-Tags 161
 codepages 191
 codes 191
 collocates 63
 collocates: display 61
 collocates: horizons 60
 collocates: minimum frequency 60
 collocates: sorting 79
 collocation associates 86
 collocation: settings 60
 collocation: specifications 60
 column totals 29
 comparison display 102
 Concord: clusters 64
 Concord: collocation 63
 Concord: creating exercises 58
 Concord: Einschränkungen 188
 Concord: Kategorien 59
 Concord: mehrere Suchwörter 74

Concord: Muster 78
 Concord: nächstes Tag 75
 Concord: overview 58
 Concord: sortieren 78
 Concord: sound and video 67
 Concord: source text file 68
 Concord: speichern und drucken 67
 Concord: stretching the display to see more 68
 Concord: Textsegmente 79
 Concord: Tipps zum Anfang 12
 Concord: Übersicht 5
 Concord: uniform plot 65
 Concord: viewing options 67
 Concord: what you see and can do 68
 Concord: Wildcards 80
 Concord: zapping unwanted lines 74
 concordance batch processing 70
 concordance display 68
 concordance settings 70
 consistency lists: sorting 119
 context word 73
 Controller (wshell.exe) 5
 convert data from old version 136
 converter 152
 count data frequencies 31
 custom .dll file 33
 custom processing 33
 cut spaces 81

- D -

Danksagung 172
 Das Programm tut nicht das, was ich möchte 199
 database statistics 90
 date format 173
 datebasierte Lemmatisierung 106
 Dateibasierte Suchwörter oder -phrasen 74
 Dateien auswählen 86
 Dateinamen 180
 Dateinamen unter Windows 95 180
 Dateinamenanzeige 51
 Datei-Utilities: Datei-Chunker 151
 Datei-Utilities: Index 147
 Datei-Utilities: Übersicht 7
 Datei-Utilities: umbenennen 152
 Daten als Textdatei 93

Daten auf N Einträge reduzieren 46
 Daten in Word kopieren 194
 Datenbank aufbauen 88
 Definition eines gewichteten Schlüsselworts 89
 definition of associate 85
 Definition von Gewichtung 89
 Definitionen 173
 demonstration version 174
 den Daten Notizen hinzufügen 19
 Desktop 179
 dezimale Ansicht 132
 Dezimalstellen 41
 Dickens-Text 24, 45
 directories 190
 disambiguation 86
 dispersion 65
 dispersion plot: sorting 79
 displaying comparisons 102
 DOS codes 192
 drucken 44
 Druckereinstellungen 174
 Druckvorschau 44

- E -

edit mode 183
 editing concordances 74
 ein Wort finden 50
 eine bevorzugte Textauswahl anlegen 23
 eine Gruppe bevorzugter Textdateien speichern 23
 Eine Konkordanz aus Schlüsselwörtern erstellen 88
 eine Tagdatei erstellen 166
 Einen Plot berechnen 92
 Einführung zu WordSmith Tools 3
 Einschränkungen 175
 Einstellungen anpassen 19
 Einträge finden 113
 Einträge löschen 56
 Einträge zusammenführen 106
 Eintragstypen finden 113
 einzelne Wörter 175
 English 200
 entfernen 56
 Entity-Referenzen 160
 Ergebnisse speichern 49

error messages: base list error 203
 error messages: can only save words as ASCII 217
 error messages: can't call other tool 203
 error messages: can't make folder as that's an existing filename 216
 error messages: can't merge list with itself! 204
 error messages: can't read file 204
 error messages: character set reset to <x> to suit <language> 218
 error messages: concordance file is faulty 210
 error messages: concordance stop list file not found 210
 error messages: destination folder not found 218
 error messages: disk problem -- file not saved 205
 error messages: dispersions go with concordances 216
 error messages: drive not valid 211
 error messages: file contains "holes" 204
 error messages: file contains none of the tags specified 204
 error messages: no concordance entries found 208
 Erste Schritte 3
 Erste Schritte mit Concord 12
 Erste Schritte mit KeyWords 13
 erste Schritte mit WordList 14
 Erste Zeile in Klammern 147
 Erstellung einer Datenbank 88
 erstmalige Verwendung von WordSmith 45
 erweiterte Einstellungen 37
 exakte Sortierung in WordList 124
 Excel 46
 exercises 58

- F -

Farben 27
 Farben in Tags 166
 Farben verändern 27
 Fehlerbehebung 199
 Fehlerbehebung: 200
 Fehlerbehebung: abgestürzt 198
 Fehlerbehebung: Akzentzeichen 198
 Fehlerbehebung: Demolimit 198
 Fehlerbehebung: Drucker druckt nicht 198
 Fehlerbehebung: eigenartige Symbole 198

- Fehlerbehebung: Ergebnisse für Concord und WordList unterschiedlich 199
- Fehlerbehebung: es dauert viel zu lange 201
- Fehlerbehebung: Farben nicht lesbar 201
- Fehlerbehebung: Spaltenabstand 200
- Fehlerbehebung: Tagproblem in Concord 200
- Fehlerbehebung: Tasten reagieren nicht 200
- Fehlerbehebung: Viewer 134
- Fehlerbehebung: Wortliste nicht sortiert 201
- Fehlermeldungen 211
- Fehlermeldungen: .ini-Datei nicht gefunden 203
- Fehlermeldungen: Arbeitsspeicher ist knapp 203
- Fehlermeldungen: auf die Datei konnte nicht zugegriffen werden 203
- Fehlermeldungen: auf Textdatei kann nicht zugegriffen werden - wird sie anderweitig verwendet? 203
- Fehlermeldungen: Datei konnte nicht gelesen werden 204
- Fehlermeldungen: Datei nicht gefunden 205
- Fehlermeldungen: Dateinamen dürfen nicht identisch sein 205
- Fehlermeldungen: dies ist eine Demoversion 205
- Fehlermeldungen: diese Funktion ist noch nicht bereit 206
- Fehlermeldungen: Dieses Programm funktioniert nur unter Windows 98 oder höher. 213
- Fehlermeldungen: Drucker notwendig aber nicht gefunden 206
- Fehlermeldungen: Formular unvollständig 206
- Fehlermeldungen: Funktion kann noch nicht vollständig ausgeführt werden 206
- Fehlermeldungen: Grenze bei Zusammenführung erreicht 206
- Fehlermeldungen: Internetverbindung konnte nicht hergestellt werden 207
- Fehlermeldungen: kein Platz für berechnete Variable 207
- Fehlermeldungen: Keine assoziierten Gruppen identifiziert 207
- Fehlermeldungen: keine assoziierten Wörter gefunden 207
- Fehlermeldungen: keine Cluster gefunden 207
- Fehlermeldungen: keine Einträge in der KeyWords-Datenbank 207
- Fehlermeldungen: keine gelöschten Zeilen zu entfernen 208
- Fehlermeldungen: keine Kollokatoren gefunden 208
- Fehlermeldungen: keine Schlüsselwörter für Plot 208
- Fehlermeldungen: keine Schlüsselwörter gefunden 208
- Fehlermeldungen: keine solchen Dateien gefunden 208
- Fehlermeldungen: keine Statistik verfügbar 209
- Fehlermeldungen: keine WordSmith-Datei 209
- Fehlermeldungen: keine Wörter aus KeyWords-Stoppliste 209
- Fehlermeldungen: keine Wörter aus Konkordanz-Stoppliste 209
- Fehlermeldungen: keine Wörter aus Lemmaliste 209
- Fehlermeldungen: keine Wörter aus Matchliste 209
- Fehlermeldungen: keine Wörter aus Stoppliste 209
- Fehlermeldungen: keine Wörter aus Tagliste 210
- Fehlermeldungen: keine Wortlisten ausgewählt 210
- Fehlermeldungen: KeyWords-Datei ist fehlerhaft 210
- Fehlermeldungen: KeyWords-Datenbank ist fehlerhaft 210
- Fehlermeldungen: Konvertierungsdatei nicht gefunden 211
- Fehlermeldungen: Limit für dateibasierte Suchwörter erreicht 211
- Fehlermeldungen: Lizenznummer für Netzwerk wird bereits verwendet 213
- Fehlermeldungen: Lizenznummer ist nicht korrekt 213
- Fehlermeldungen: Lizenznummer muss aus 20 Zeichen bestehen. 206
- Fehlermeldungen: Matchliste 205
- Fehlermeldungen: muss eine Zahl sein 213
- Fehlermeldungen: Neues Verzeichnis konnte nicht erstellt werden 213
- Fehlermeldungen: Nicht gespeichert 205
- Fehlermeldungen: nichts aktiviert 213
- Fehlermeldungen: Quelldatei(en) nicht gefunden 213
- Fehlermeldungen: Schnittstelle zwischen den Tools unterbrochen 214
- Fehlermeldungen: Stopplistendatei nicht eingelesen 214
- Fehlermeldungen: Stopplistendatei nicht gefunden 214
- Fehlermeldungen: Tagdatei nicht eingelesen 214
- Fehlermeldungen: Tagdatei nicht gefunden 215
- Fehlermeldungen: ungültige KeyWords-Datei 215

Fehlermeldungen: ungültige KeyWords-Datenbank 215
 Fehlermeldungen: ungültige Konkordanzdatei 215
 Fehlermeldungen: ungültige Nummer 216
 Fehlermeldungen: ungültige WordList-Datei 216
 Fehlermeldungen: ungültige WordList-Vergleichsdatei 216
 Fehlermeldungen: ungültiger Dateiname 216
 Fehlermeldungen: Vollständiger Pfadname wird benötigt 217
 Fehlermeldungen: Zugriff auf die originale Textdatei nicht möglich, da nicht vorhanden 219
 file associations 176
 File Utilities: compare 2 files 150
 File Utilities: find duplicates 151
 file-types 176
 filtern 42
 Finden durch Eingabe 50
 folders 190
 Format 41
 frequency of happi* 31
 Funktionstasten bei Suche 113
 Für die Nutzung am PC mit dem Namen XXX 213

- G -

Gänsefüßchen 198
 gesamte Wortsuche 80
 Geschwindigkeit 176
 gewichtetes Schlüsselwort definieren 89
 Gewichtung definieren 89
 Globalität des Plots 183
 Griechisch 24
 Griechische Schriftart 22
 Groß- und Kleinschreibung beachten 80
 grow and shrink 68

- H -

Handbuch für WordSmith Tools 177
 hexadezimale Ansicht 132
 hide tags 81
 hide words 81
 Hilfe beim Programmstart anzeigen 48
 holes in file 204
 horizons 60
 Hotkeys 186

HTML
 XML 177
 HTML Header: ausschneiden 163
 HTML und SGML Tags 161

- I -

Idiome 175
 in Tabelle, etc. exportieren 46
 Indexlisten: Verwendung 107
 Information über die WordSmith-Version 189
 innerhalb von Texten auswählen 165
 instructions folder (Verzeichnis mit Anweisungen) 16
 interface 178
 international versions 178
 Internet Explorer 180

- J -

Japanisch 193
 Joiner 149

- K -

keine aktuelle WordSmith-Datei 207
 key key-words 90
 KeyWords database 90
 KeyWords: Anzeige 95
 KeyWords: Berechnung 91
 KeyWords: Cluster 87
 KeyWords: Einschränkungen 188
 KeyWords: Links 91
 KeyWords: sortieren 95
 KeyWords: Tipps 90
 KeyWords: Tipps zum Anfang 13
 KeyWords: Übersicht 6
 KeyWords: Zweck 85
 Kollokationsmuster 78
 Konkordanz mit Tags 72
 Konkordanzen zusammenführen 102
 Konsistenzanalyse (detailliert) 104
 Konsistenzanalyse (einfach) 105
 Kontaktadressen 179
 Kopieren: alle 33
 Kopieren: festlegen 33

Kopieren: markierte 33
 Kyrillisch 24

- L -

lange Dateinamen 180
 langsam 201
 Languages Chooser: Andere Sprachen 143
 Languages Chooser: Einstellungen speichern 144
 Languages Chooser: Schriftart 142
 Languages Chooser: Sortierreihenfolge 142
 Languages Chooser: Sprache 141
 Languages Chooser: Übersicht 140
 Layout 41
 Layout speichern 41
 Lemmata 106
 Lemmata automatisch zusammenführen 99
 lemmatising with custom .dll 33
 Lemma-Übereinstimmungen: WordList 100
 Letzte Datei 179
 letzte Datei wiederherstellen 179
 Lineal 93
 Liste der Schaltflächen 180
 localisation 178
 Logdatei zur Fehlersuche 37
 Log-Likelihood-Test 91
 Log-Likelihood-Wert 117

- M -

Markierung des Endes einer Überschrift 53
 Markierung einer Überschrift 53
 Markierung eines Absatzendes 53
 Markierung eines Absatzes 53
 Markierung eines Satzendes 53
 Markierung eines Satzes 53
 Matchliste 42
 mehrere Listen 19
 Menüauswahl 180
 MI3-Wert 117
 Microsoft Word 194
 Minimal Pairs 144
 Minimal Pairs: aim 144
 Minimal Pairs: Anforderungen 144
 Minimal Pairs: choosing files 144
 Minimal Pairs: Output 145

Minimal Pairs: Programm verwenden 146
 Minimal Pairs: Regeln und Einstellungen 145
 Minimal Pairs: Übersicht 7
 mit mehreren geöffneten Fenstern arbeiten 55
 MI-Wert 117
 moving sentences 131
 multimedia concordancing 67
 Multimedia-Tags 162
 Multimedia-Tags definieren 162
 multiple file analysis 90

- N -

nach einem Wort bzw. Wortteil suchen 50
 nächstes Tag 75
 negative Gewichtung 89
 Netzwerke: Standardeinstellungen 48
 Netzwerkeinstellungen 16
 Netzwerk-Standardeinstellungen 48
 Netzwerkversion 16
 neu in Version 4 5
 neu sortieren 56
 neu sortieren: Concord 78
 neu sortieren: KeyWords 95
 neue Datenspalte berechnen 30
 neue Version herunterladen 17
 neuer Benutzer 45
 nicht lesbar 201
 Notizen 19
 number of concordance entries 81
 numbering: paragraphs 131
 numbering: sentences 131

- O -

Option "Über" 175
 Optionen für Standardeinstellungen 48
 ordering details 174
 over-writing 153
 Oxford University Press 174
 Oxford WordSmith Tools installieren 16

- P -

paragraph numbering 131
 percentages v. raw numbers 81

Phrasen 175
 Plotanzeige 93
 Plots und Links 91
 Plot-Verteilungswert 183
 Popupmenü 37
 Popupmenü anpassen 37
 Portugiesisch 24
 potato-peeling machine 200
 price 174
 Protokollierung 37
 purple marks 68
 purpose of Text Converter 152
 P-Wert 92

- Q -

Quelltexte 184
 Quelltexte finden 184

- R -

RAM-Auslastung 184
 raw numbers v. percentages 81
 Referenzkorpus 185
 Registrierungsaufforderung 206
 registry 176
 relevante Dateien finden 40
 replacing 153
 research uses 58
 re-sorting: collocates 79
 re-sorting: consistency lists 119
 re-sorting: dispersion plot 79
 restore last file 180
 restricted search 73
 Russisch 24
 Russische Schriftart 22

- S -

Satz: Anfang und Ende 168
 Sätze (festlegen) 124
 Sätze: Definition 173
 Schaltflächen 180
 Schlüsselwort-Datenbanken konnten nicht
 zusammengeführt werden 214
 Schlüsselwörter grafisch darstellen 92

Schnittstellen zwischen Tools 185
 Schrägstrich 80
 Schriftart verändern 22
 Schriftarten 22
 Schriftbild 41
 search by typing 183
 sentence numbering 131
 sentence only 81
 setting up a training session 40
 skimmen 132
 sortieren: Concord 78
 sortieren: KeyWords 95
 sortieren: WordList 124
 Sortierung von Zahlen 78
 sound & video tagged files 67
 Spalte "Klasse" 59
 Spalten drucken 174
 Spaltenbreite 41
 Speichernutzung 184
 Speicherplatz 184
 spezifische Einschränkungen 188
 Splitter 147
 Splitter: Dateinamen 148
 Splitter: Symbole 148
 Splitter: Übersicht 7
 Splitter: Wildcards 148
 splitting 132
 Sprache 24
 Standardeinstellungen 48
 Standardeinstellungen für BNC-Tags 163
 Standardeinstellungen speichern 48
 standardisierte oder durchschnittliche
 Type-Token-Relation 122
 startet nicht 200
 statistics of a database 90
 Statistik 119
 Statusleiste 180, 186
 Sternchen 80
 stoplist.cod 156
 Stopplisten 51
 student use 58
 Suchen durch Eingabe 50
 Suchen und Ersetzen 50
 Suchwort-Syntax 80
 summary statistics 31
 Symbole 192

Symbolleiste 180
Systemanforderungen 186

- T -

tag context 70
Tag-Arten 160
Tagdatei 166
Tagdatei automatisch laden 161
Tag-Konkordanz erstellen 72
Tags als Selektoren 163
Tags sortieren 75
Tags werden in Concord nicht angezeigt 200
Tags: Übersicht 160
Tagtypen 160
Tastatur 186
Tastenkombinationen 186
teacher instructions 40
teaching uses 58
Teile von Daten speichern 49
teilweise speichern 49
Text Converter folders 153
Text Converter settings 153
Text Converter: asterisk 154
Text Converter: Einschränkungen 188
Text Converter: Konvertierungsdatei 156
Text Converter: syntax 154
Text Converter: Übersicht 7
Text Converter: verschieben, wenn 156
Text Converter: Vorlage Konvertierungsdatei 157
Text Converter: wildcards 154
Text in eine Wortliste importieren 120
Textdatei: verwenden, um eine Wortliste zu erstellen 120
Textdateien speichern 24
Textdateien zusammenführen 149
Texte auswählen 24
Texte mit Tags 160
Texte: auswählen 24
Texte: mehr Texte 24
Texteigenschaften 53
Textendesymbole 148
Textformate 53
Textsegmente in Concord 79
the ~ operator 73
Tools zur Erkennung von Mustern 187

training students 40
Transinformation: Anzeige 117
Transinformation: Berechnen 116
Transinformationswerte 114
trennen 106
Trennzeichen des Textendes 147
troubleshooting: apostrophes not found 198
troubleshooting: pineapple-slicing 200
troubleshooting: quotation marks not found 198
troubleshooting: smart quotations 198
Two word-list analysis 85
type-in mode 183
Type-Token-Relationen 122

- U -

über Menü suchen 113
Überschriften (festlegen) 124
Überschriften: Anfang und Ende 168
Überschriften: Definition 173
Umgang mit Zahlen im Text 191
undefined tags 81
ungleichmäßige Verteilung 183
Unicode-Kodierung 193
university or school work 40
unleserlich 201
Unterschiede in Version 4 170
user-defined process 33

- V -

Verarbeitung unterbrechen 52
Verlaufsliste 45, 80
Version Checker: overview 8
version francaise 178
Version überprüfen 17
Versionen stimmen nicht überein 217
Versionsdatum 189
Viewer 130
Viewer: aligning the sentences 131
Viewer: Anzeigoptionen 132
Viewer: bearbeiten 131
Viewer: Einschränkungen 188
Viewer: Einstellungen 132
Viewer: markierte Wörter 132
Viewer: sentence joining 132

Viewer: skimmen 132
Viewer: Tags 132
Viewer: technische Aspekte 133
Viewer: translation mis-matches 133
Viewer: Übersicht 9
Viewer: ungewöhnliche Sätze 134
viewing original text file 68
vorherige Auswahl löschen 24
vorherige Listen 45

- W -

Was ist neu 5
WebGetter: Anzeige 138
WebGetter: Einschränkungen 139
WebGetter: Einstellungen 137
WebGetter: Übersicht 10, 136
what is a concordance 58
Wie baue ich eine Datenbank auf? 88
Wie viel Text? 175
Wie viele Wörter? 175
Wieso ist die Suche fehlgeschlagen? 217
Wieso nicht? 199
Window Management 55
Windows 2000 186
Windows 98 186
Windows character set codes 192
Windows NT 186
Windows XP 186
WordList Indexlisten: anzeigen 108
WordList: anhand einer Textdatei erstellen 120
WordList: Cluster 110
WordList: Einschränkungen 188
WordList: Einstellungen für Minimum und Maximum 124
WordList: Einträge verändern 38
WordList: Groß- und Kleinschreibung beachten 123
WordList: Sortierreihenfolge 124
WordList: Sortierung 201
WordList: the basic display 125
WordList: Tipps zum Anfang 14
WordList: Übersicht 6
WordList: Zweck 99
WordList-Einträge bearbeiten 38
WordList-Index erstellen 109

WordSmith controller: Concord: settings 81
WordSmith controller: KeyWords settings 96
WordSmith controller: WordList settings 127
WordSmith Tools: Handbuch 177
WordSmith Tools: Installation 16
WordSmith wird bereits ausgeführt. 217
WordSmith-Version 189
Wort: Definition 173
Wortliste aus KeyWords-Daten erstellen 92
Wortliste ist fehlerhaft 217
Wortlisten in verschiedenen Sprachen können nicht verglichen werden 215
Wortlisten neu sortieren 38, 124
Wortlisten vergleichen 101
Wortlisten zusammenführen 102
Wortlistendatei nicht gefunden 218
Wortlisten-Vergleichsdatei ist fehlerhaft 218
Wortmuster 78
Worttrennzeichen 173
wshell.exe (Controller) 5
wshell.ini und Netzwerke 16

- X -

XX Tage verbleiben 218

- Z -

Zahlen 53
zahlreiche Dateien umbenennen 152
Zeichen innerhalb eines Worts 53
Zeilenanfänge entfernen 165
Zentraleuropäisch 24
Zip-Dateien 193
zu viele Sätze 218
zu viele übereinstimmende Gruppen zu ignorieren 218
zufälliges Löschen von Einträgen 46
zum Text hinzufügen 86
Zweck von Splitter 147
Zweck von Viewer 130
zwei Dateien notwendig 219
Z-Wert 117
Zwischen Texten auswählen 164
Zwischenablage 194

Endnotes 2... (after index)

Back Cover