

Kurzreferenz zu COSMAS II_{script}

Das Programm *c2script* wird mittels des Bash-Skripts *c2.sh* und einer "COSMAS II"-Skriptdatei als Parameter ausgeführt (z.B.: *./c2.sh test.scr*). Die Datei *test.scr* enthält den kompletten Befehlssatz und Beispielskripte in Kommentarform (durch Entfernen der Kommentarzeichen # können die Beispiele ausprobiert werden). Weitere Informationen befinden sich in der Datei README.

Die Konventionen für ein "COSMAS II"-Skript sind wie folgt:

- Jede Zeile enthält genau ein Kommando.
- Kommentare (eingeleitet durch das Zeichen # am Zeilenanfang) und Leerzeilen werden ignoriert. Bei allen set-Optionen (siehe Tab. 1) sind Kommentare in der Form 'set Option=Wert # Kommentar' erlaubt (z.B.: set SortHits=alphabetical R1A+L1A # Kommentar)
- Es ist genau ein beliebiges Trennzeichen (z.B. ein Space) zwischen Befehl und Parameter erforderlich.
- Die Wertzuweisung für eine Optionen (siehe Tab. 1) erfolgt mittels *Option=Wert* (ohne Leerzeichen).
- Die Befehlsreihenfolge gemäß dem Ablaufdiagramm (siehe Abb. 1) ist zu beachten.

Der von der Applikation unterstützte Befehlssatz gliedert sich in 4 Gruppen (<...> kennzeichnet Parameter):

- Recherche bezogene Befehle
 - Öffnen eines Archivs: *openArchiv <Archiv-Name>*
 - Auflisten aller Archive/Korpora: *listArchives*, *listCorpora*, *listUserCorpora*
 - Laden eines virtuellen Korpus: *loadCorpus | loadUserCorpus <Names des virtuellen Korpus>*
 - Generieren einer Suchanfrage: *bodyQuery <Query>* (siehe CII-Dokumentation zur Suchanfragesyntax)
- Setzen von Optionen: *set <Option=Wert>* (siehe Tab. 1)
- Ergebnis bezogene Befehle
 - Trefferstatistik anzeigen/exportieren: *statistics [<Dateiname>]*
 - Kookkurenzanalyse: *collocation*
 - Anfrage-Ergebnisse in eine Datei exportieren: *export <Dateiname>*
- Sonderbefehle
 - Einbinden einer Datei mit weiteren Befehlen: *include <Dateiname>*
 - Abbruch der Befehlsverarbeitung: *exit*

Außer für die Befehle *include* und *exit* gilt die in Abbildung 1 dargestellte Befehlsreihenfolge. Die verfügbaren Optionen (siehe Tab. 1) stehen in Klammern neben dem zugehörigen Befehl.

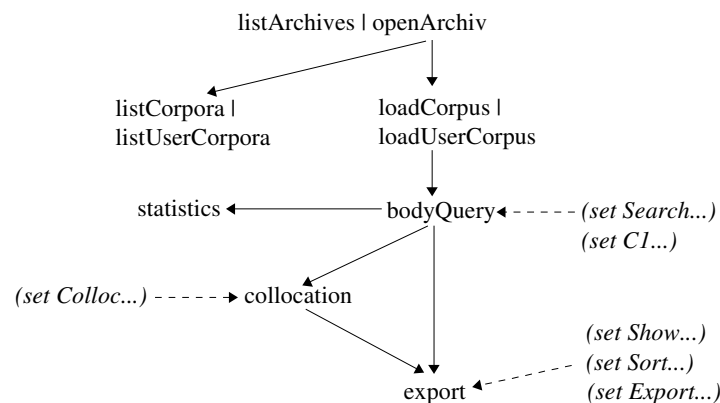


Abb. 1: Ablaufdiagramm einer skriptbasierten "COSMAS II"-Sitzung

Jede Option kann mit '*set Option=*' auf seinen voreingestellten ('default') Wert (innerhalb der Spalte 'Wert' fett gedruckt) zurückgesetzt werden. Für Einträge in der Spalte 'Wert' gilt: '[a,b]' bedeutet 'a oder b', '[0-9]+' bedeutet 'ein ganzzahliger Zahlenwert >= 0'. Falls die Werte einer Option kumulativ sind, dürfen sie mit '*set Option=Wert1+Wert2+...*' kombiniert werden (siehe 'SearchCase' und 'SearchLemma'). Die Werte 'aroundYear' (siehe SearchCorpusPresentation und SearchResultPresentation) und 'alphabetical' (siehe SortKwicAndFullText) benötigen einen weiteren Wert, welcher durch ein Leerzeichen getrennt folgt. Von runden Klammern umgebende Optionen (wie z.B. ExportFullText) sind bei 'ExportFormat=xml' nicht verfügbar.

Option	Wert	Bemerkungen	Abb.
C1Query	[off , on]	on => in der Suchanfrage wird die C1-Query-Syntax verwendet off => in der Suchanfrage wird die C2-Query-Syntax verwendet	10
C1MissingProxOp	[w1 , or]	w1 => weggelassener Verknüpfungsoperator bedeutet: Wortabstand /+w1 or => weggelassener Verknüpfungsoperator bedeutet: logisches „ODER“	10
CollocAutofocus	[off, on]		9
CollocClustering	[unique , multiple]		9
CollocConfidence	[normal , high, analytical]		9
CollocGranularity	[fine, medium , coarse, very_coarse]		9
CollocIgnoreFunctionWords	[off, on]		9
CollocLemmatization	[off , on]		9
(CollocLineNumbering)	[off , on]		9
CollocLLR	[off, on]		9
CollocSentenceBoundaries	[off, on]		9
CollocWordLeft	[0-9]+, default: 5		9
CollocWordRight	[0-9]+, default: 5		9
(ExportAnnotations)	[off , on]	funktioniert nur in Zusammenhang mit 'ExportFullText=on'	2
(ExportCorpusPresentation) (früher: ExportDocumentList)	[off , on]	Korpuspräsentation (siehe 'ShowCorpusPresentation')	2 u. 6
ExportExpansionLists	[off , on]	siehe <i>SearchExpansionLists</i>	2
ExportFontHeight	[0-9]+, default: 10		2
ExportFormat	[ascii , rtf, append, xml]	Der Modus <i>append</i> (Ausgabeformat ist ASCII) bedeutet, dass Dateien durch den export-Befehl nicht überschrieben werden.	2
(ExportFullText)	[off, on]	Belege (s. 'SortKwicAndFullText')	2 u. 5.1,5.3
(ExportFullTextDistanceType)	[characters, words, sentences, paragraphs]		2
(ExportFullTextLeftContext)	[0-9]+, default: 0		2
(ExportFullTextRightContext)	[0-9]+, default: 0		2
(ExportKwic)	[off , on]	KWIC (s. 'SortKwicAndFullText')	2 u. 5.1,5.2
(ExportKwicLeftContext)	[0-9]+, default: 0		2
(ExportKwicRightContext)	[0-9]+, default: 0		2
(ExportKwicDistanceType)	[characters, words, sentences , paragraphs]		2

Option	Wert	Bemerkungen	Abb.
(ExportMaxLines)	[0-9]+, default: 10000	'0' steht für eine unbegrenzte Anzahl.	2
(ExportMultiWordContext)	[off , on]	vgl. Bsp. 19 u. 20 in test.scr; zu 'test.scr' siehe Seite 1 in diesem Dokument	
ExportResultPresentation früher: ExportOverview	[off , on]	Ergebnispräsentation (siehe 'ShowResultPresentation')	2 u. 5.1
SearchCase	[everythingInsensitive , allCaseInsensitive, allCaseInsensitiveButFirst, allCaseSensitiveButFirst, allUpperCaseButFirst, allLowerCaseButFirst, firstCaseInsensitive, firstCaseSensitive, firstUpperCase, firstLowerCase, diacriticSensitivityOn, diacriticSensitivityOff]	Das Setzen aller Werte dieser Option ist kumulativ. Der Wert <i>everythingInsensitive</i> entspricht <i>allCaseInsensitive+diacriticSensitivityOff</i> . Der Wert <i>allCaseInsensitive</i> entspricht <i>allCaseInsensitiveButFirst+firstCaseInsensitive</i> . Die Werte ' <i>...UpperCase...</i> ' und ' <i>...LowerCase...</i> ' sind als Optionen in COSMAS II _{win} nicht verfügbar.	8
SearchCutValue	[0-9]+, default: 10000	funktioniert nur in Zusammenhang mit 'SearchLimit=...' (alle Werte außer 'off')	8
SearchExpansionLists	[freqDesc , freqAsc, alphaDesc, alphaAsc, retroDesc, retroAsc, off]		8
SearchLemma	[inflexionOnly , compounds, otherWordBuildingForms, specialWordForms]	Das Setzen der Werte für diese Option ist kumulativ. Der Wert <i>inflexionOnly</i> ist immer gesetzt bzw. nicht abwählbar.	4
SearchLimit	[off , linear, random_variable, random_fixed]	Der Wert <i>linear</i> bezieht sich auf 'die ersten n' und der Wert <i>random_...</i> auf 'n zufällig ausgewählte' Treffer. Beide Werte benötigen die Option <i>SearchCutValue</i> .	8
ShowCorpusPresentation früher: SortDocumentList	[byCorpus , byDoc, aroundYear ... , byDecade, byYear, byMonth, byDay, bySource, byCountry, byThema[...], byTextsorte, distribTexts, distribWords, distribGray]	'...' in 'aroundYear ...' steht für ein 4-stelliges Jahr (angezeigt werden alle Ergebnisse vor/nach dem 1. Januar) '...' in 'byThema[...]' steht für einen optionalen Wert zwischen 0 und 100 (vgl. Bsp. 9 in test.scr; zu 'test.scr' siehe Seite 1 in diesem Dokument); zu 'distribTexts', 'distribWords' u. 'distribGray' siehe Bsp. 11 in test.scr	6
ShowResultPresentation früher: SortOverview	[byDoc , byCorpus, aroundYear ... , byDecade, byYear, byMonth, byDay, bySource, byCountry, byThema[...], byTextsorte, rel_Freq_proz, rel_Freq_pMW, diffKoeff, haufKlass_Einzelwort, haufKlass_Einzelwort_CS, haufKlass_Lemma, haufKlass_automatisch]	'...' in 'aroundYear ...' steht für ein 4-stelliges Jahr (angezeigt werden alle Ergebnisse vor/nach dem 1. Januar) '...' in 'byThema[...]' steht für einen optionalen Wert zwischen 0 und 100 (vgl. Bsp. 9 in test.scr; zu 'test.scr' siehe Seite 1 in diesem Dokument); zu 'rel_Freq_*', 'diffKoeff' u. 'haufKlass_*' siehe Bsp. 12 bis 18 in test.scr	5.1
Sort	[off , on]	Sortierung aus- und einschalten	---
SortKwicAndFullText früher: SortHits	[unsorted , alphabetical ..., chronological, random_variable, random_fixed]	<i>chronological</i> bedeutet 'chronologisch aufsteigend'. '...' in 'alphabetical ...' steht für 'a+b+c{+iSC}' mit a,b,c ∈ {[L,R] [1,2,3],0}[A,D]}, z.B.: 'alphabetical 0A+L1D+R2A'. Anm.: 'iSC' ist optional und steht für 'Sonderzeichen ignorieren', '0A' für 'Suchwort aufsteigend', 'L1D' für 'erster Vorgänger absteigend' und 'R2A' für 'zweiter Nachfolger aufsteigend'.	3 u. 5.1-5.3 u. 7

Tab. 1: Übersicht aller Optionen des set-Befehls

Jede Anfrage beginnt mit einem **openArchiv**-Befehl und endet mit dem Eingabeende (Dateiende bzw. STRG+D bei direkter Eingabe), oder einem neuen openArchiv-Befehl.

Die Befehle **listArchives**, **listCorpora** und **listUserCorpora** liefern Listen aller zur Verfügung stehenden Archive bzw. aller zugehörigen (benutzerdefinierten) Korpora.

Mit dem **load(User)Corpus**-Befehl kann man einen virtuellen Korpus aus einem geöffneten Archiv laden.

Mit dem Befehl **bodyQuery** generiert man eine korpusbezogene Suchanfrage über die geladenen Korpora.

Jede mit dem **set**-Befehl gesetzte Option bleibt bis zum Ende der Anfrage gültig (außer der Wert wird durch erneutes Setzen verändert) und muss vor Ausführung seines zugehörigen Befehls gesetzt sein.

Der **collocation**-Befehl führt auf das Ergebnis der bodyQuery eine Kookkurenzanalyse durch. Falls eine Kookkurenzanalyse mit der Option „ExportFormat=xml“ ausgeführt wird, sollte beachtet werden, dass immer der gleiche Typ Ergebnisübersicht ausgegeben wird, unabhängig davon, wie die Optionen „ShowCorpusPresentation“ und „ShowResultPresentation“ gesetzt sind.

Der **export**-Befehl schreibt das Ergebnis der Recherche in eine Datei. Man kann nach jedem Export zu einem beliebigen Befehl zurückkehren und von dort die Ausführung des Skriptes fortführen.

Anmerkung zum XML-Format:

Die Option „ExportFormat=xml“ ist z.Z. nur für Kookkurenzanalysen implementiert.

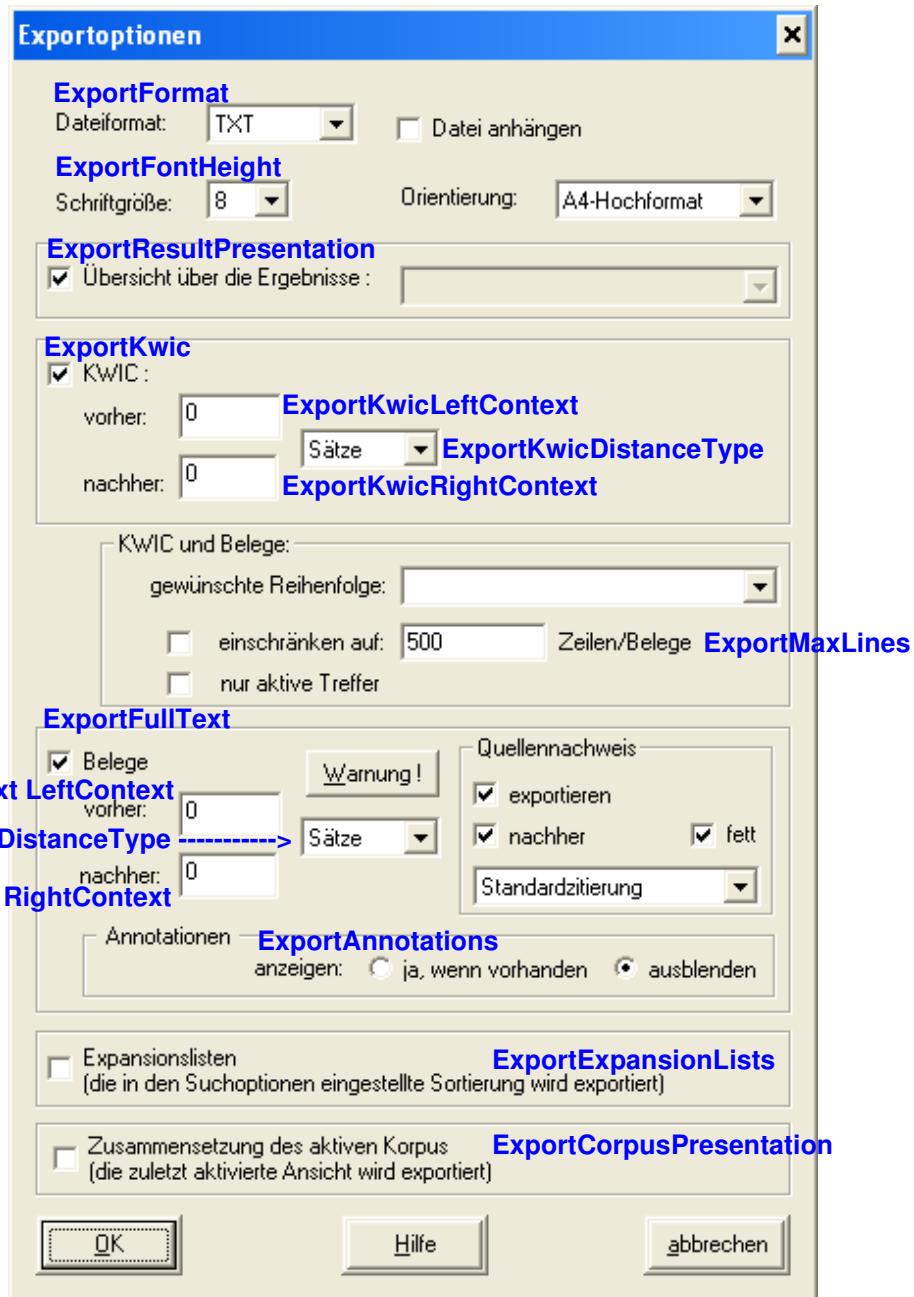


Abb. 2: Exportoptionen (Optionen -> Export)

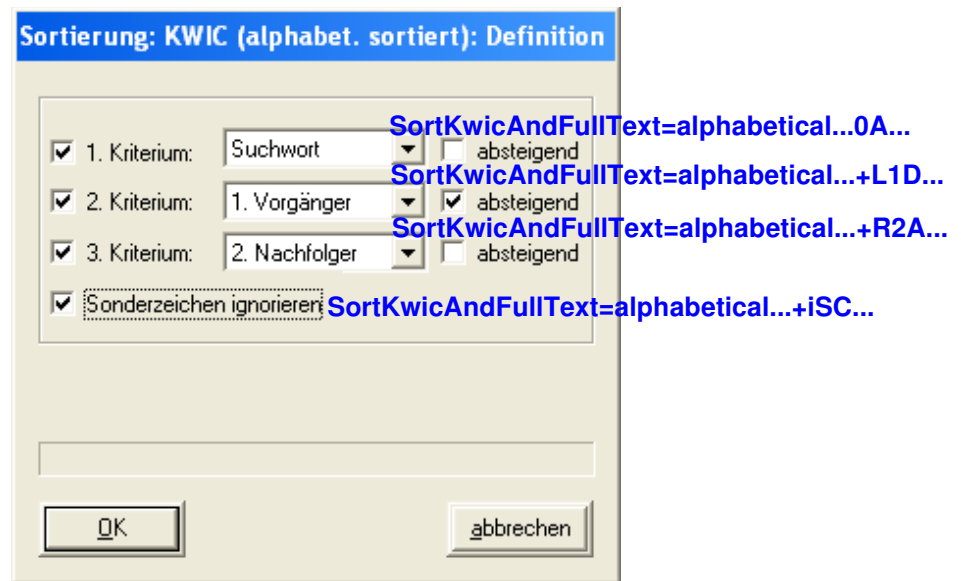


Abb. 3: Alphabetische Sortierung des KWIC (Ergebnispräsentation: Ansicht -> KWIC (alphabet. sortiert))

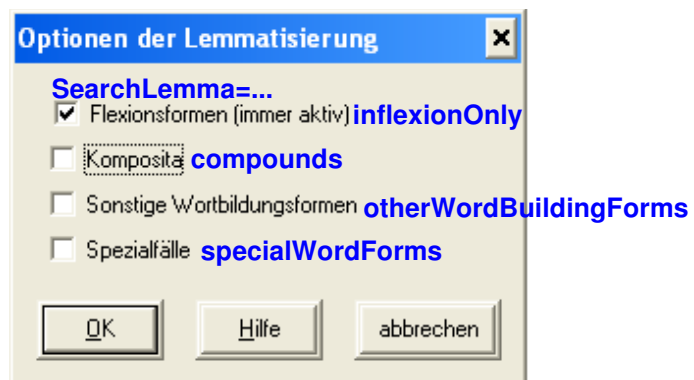


Abb. 4: Optionen der Lemmatisierung (Optionen -> Lemmatisierung)

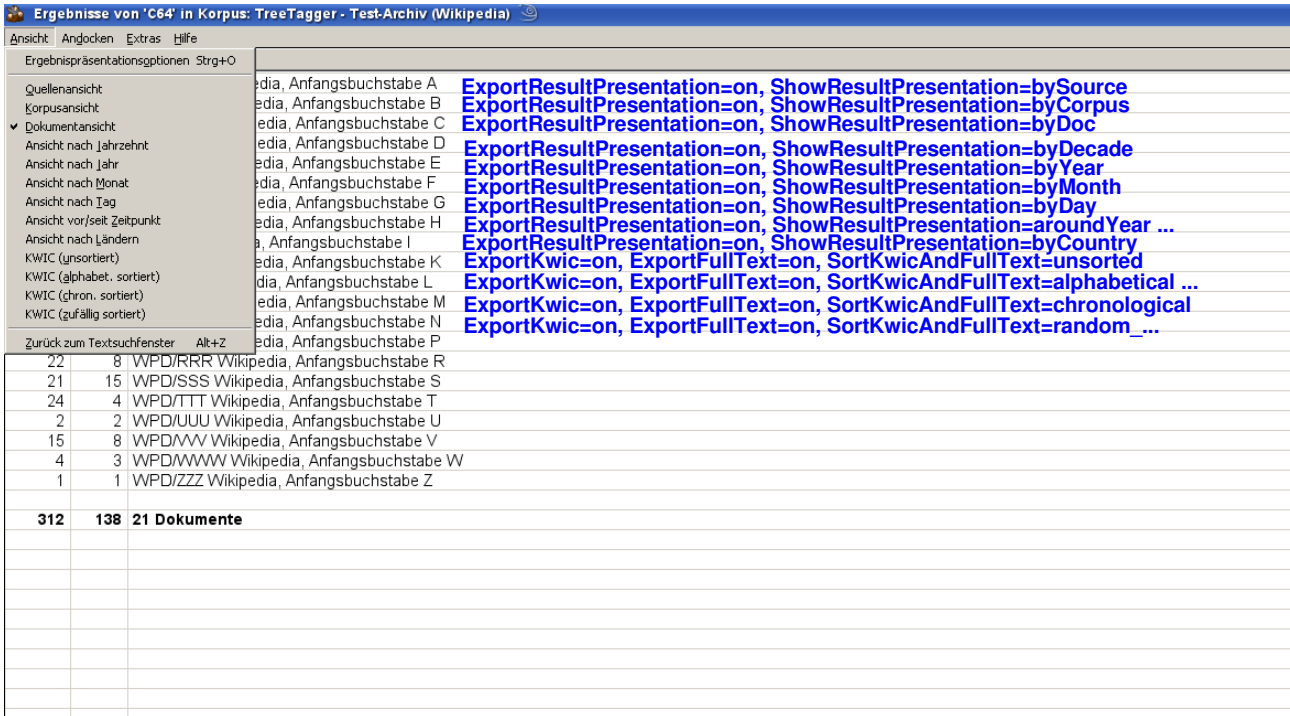


Abb. 5.1: Ergebnispräsentation (Doppelklick auf Suchergebnis)

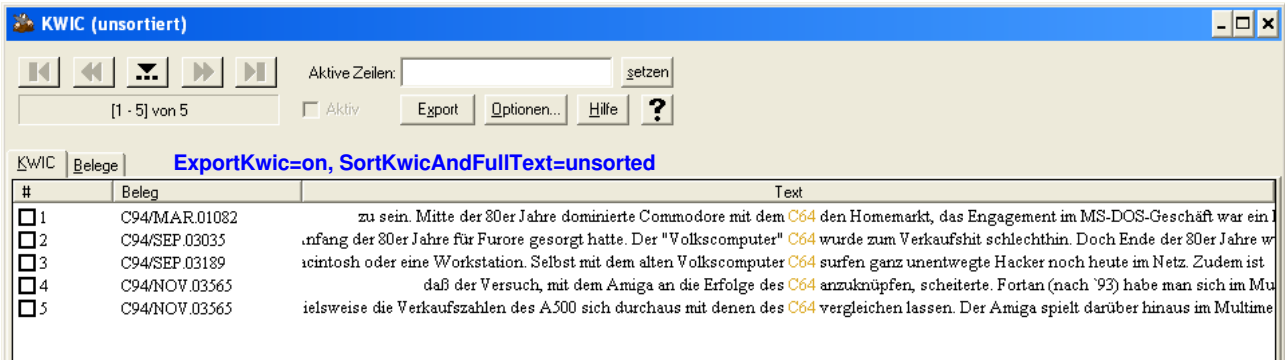


Abb. 5.2: KWIC (Ergebnispräsentation -> Ansicht -> KWIC (unsortiert))



Abb. 5.3: Belege (Ergebnispräsentation -> Ansicht -> KWIC (unsortiert) -> Klick auf Belege)

ExportCorpusPresentation=on, ShowResultPresentation=...

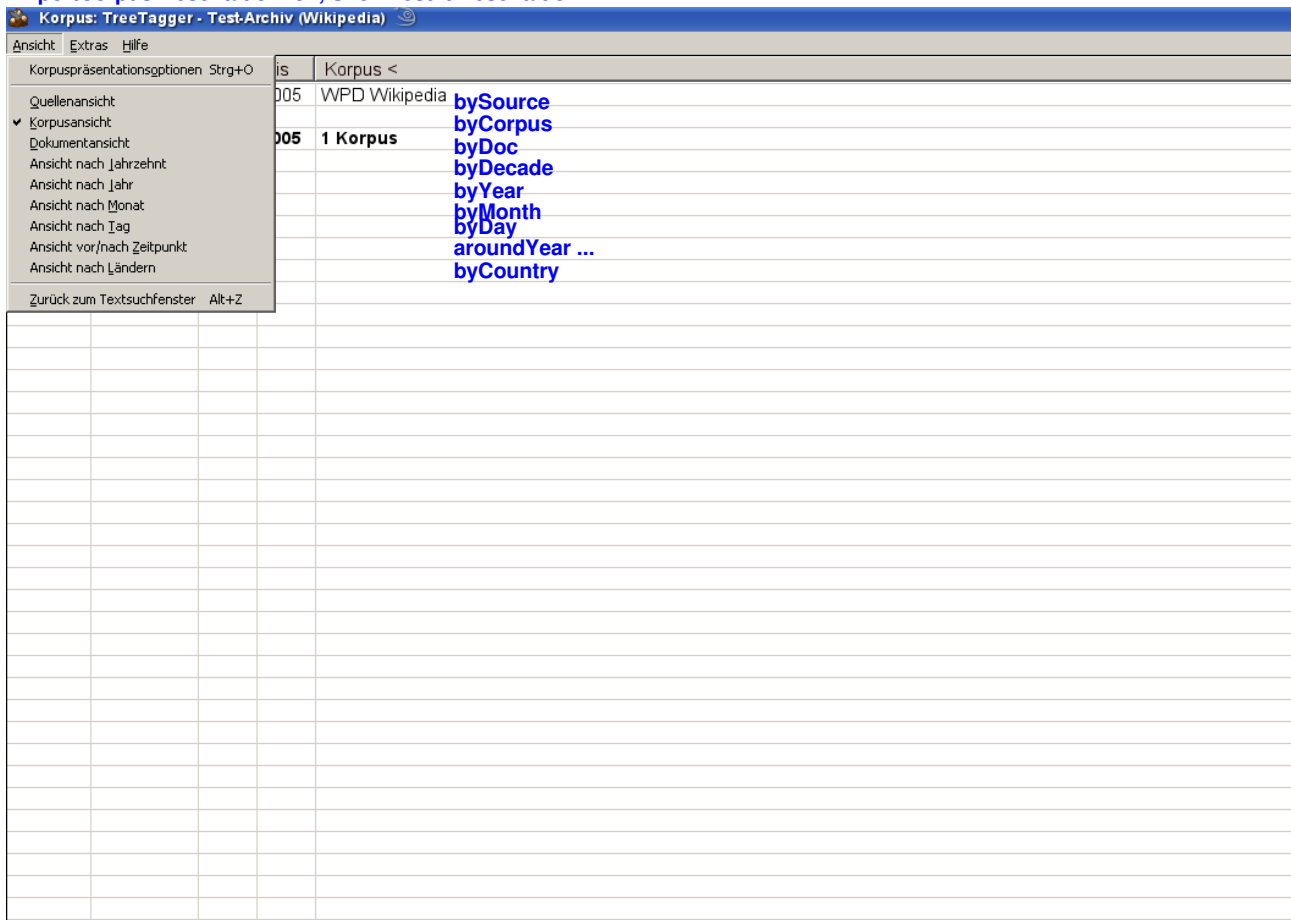


Abb. 6: Korpuspräsentation (Korpus -> Korpus anzeigen)

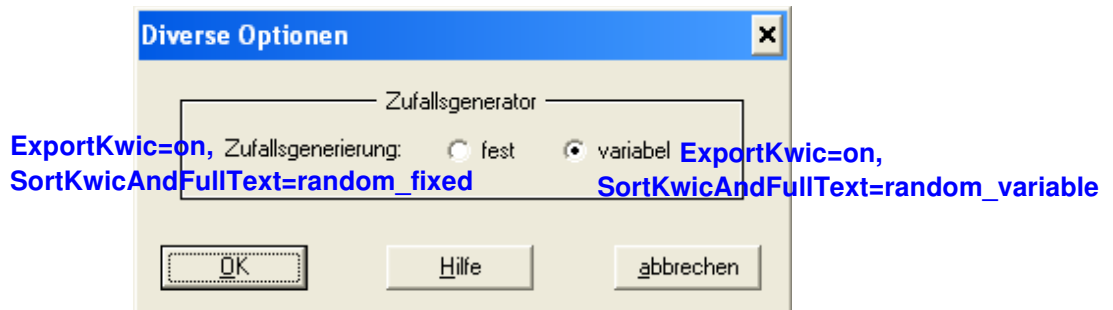


Abb. 7: Diverse Optionen (Optionen -> Divers)

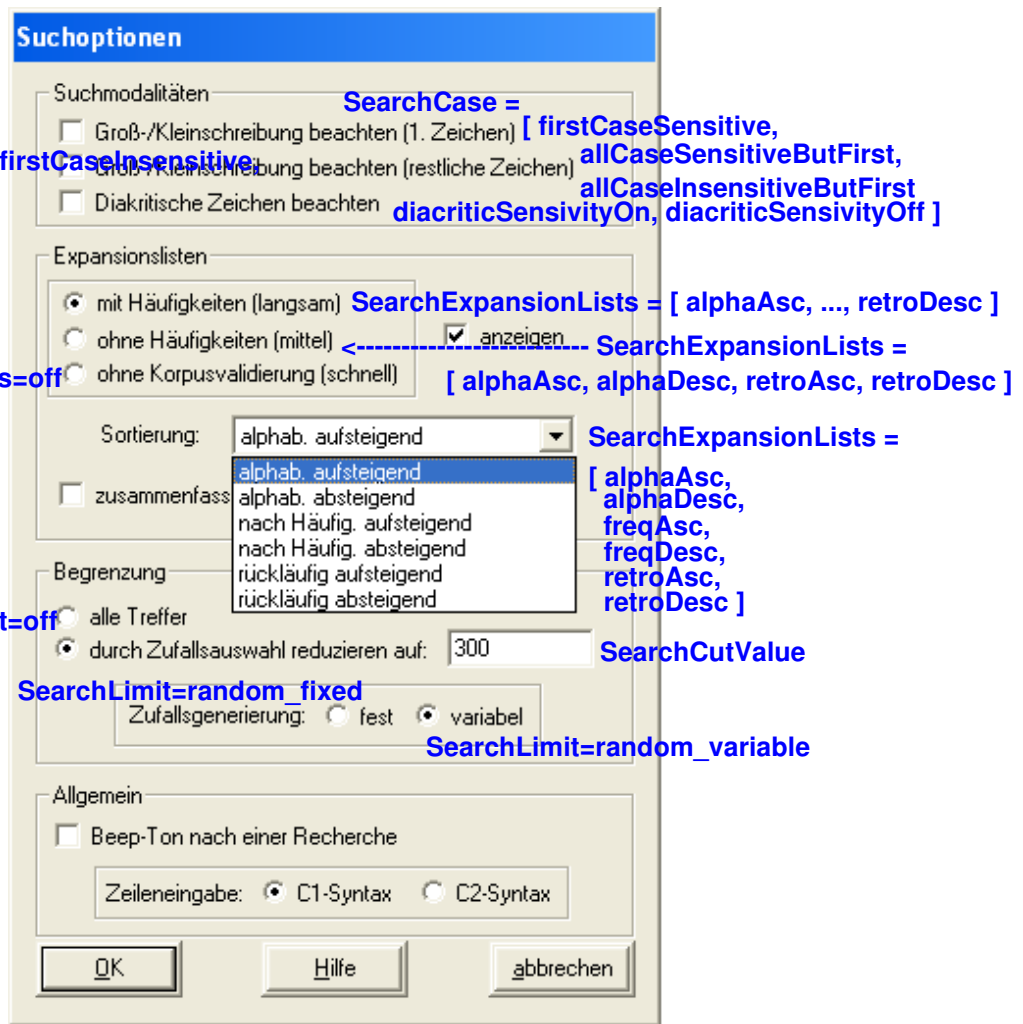


Abb. 8: Suchoptionen (Optionen -> Suche)

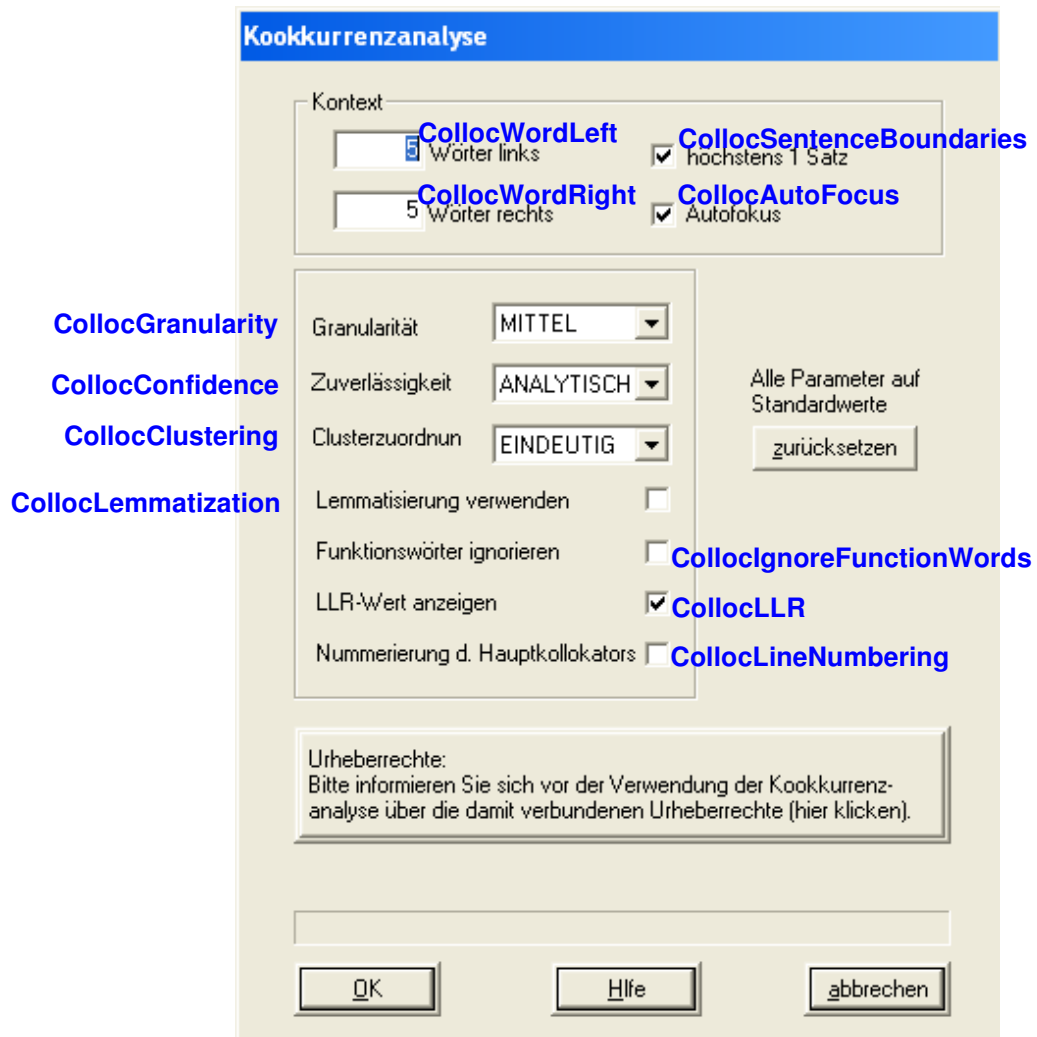


Abb. 9: Kookkurrenzanalyse (Ergebnispräsentation: Extras -> Kookkurrenzanalyse)

Über Cosmas II	Abmeldung	Recherche	Optionen
Archiv	Aktuelles Archiv	W - Archiv der geschriebenen Sprache	
Korpus	Suchanfrage	altes Buch	
Suchanfrage	Suche formulieren		
Wortformliste			
Ergebnisse	Eingabe:	Buch Seite	
Kookkurrenzanalyse	Weggelassener Verknüpfungsoperator bedeutet:	<input type="radio"/> Wortabstand /+w1 <input checked="" type="radio"/> logisches "ODER"	
Gesamt-KWIC	Beispiele:	C1MissingProxOp=[w1, or]	
Gesamt-Volltext		direkt nacheinander	
Export			

Abb. 10: Weggelassener Verknüpfungsoperator