

A. Verbindung zum FU-Login-Server mit ZEDAT-Kennung: A.1

- ▶ **Windows:** PuTTY öffnen (http://www.zedat.fu-berlin.de/tip4u_03.pdf)
(PuTTY auf Desktop abspeichern, nicht aus Anleitung starten!)
 - a. Beim ersten Anmelden: Hostkey akzeptieren [OK]
 - b. Bei „login as“ eigenen ZEDAT-Usernamen eingeben [ENTER]
 - c. Passwort eingeben (PW wird nicht angezeigt!) [ENTER]
- ▶ **Mac/Linux:** Terminal öffnen (Spotlight/DashHome > Terminal)
 - a. ~\$ ssh ZEDATKENNUNG@login.fu-berlin.de [ENTER]
 - b. Beim ersten Anmelden: Hostkey akzeptieren **yes** [ENTER]
 - c. Passwort eingeben (PW wird nicht angezeigt!) [ENTER]
- ▶ Um Verbindung zu beenden: USER@login:~\$ **exit** [ENTER]

B. Einstellungen setzen (muss nur einmalig ausgeführt werden!):

- ▶ Verbindung zum Login-Server herstellen (siehe A.)
- ▶ USER@login:~\$ sh /home/s/structeng/cqp.sh [ENTER]
- ▶ Einstellungen aktivieren: vom Server ab- & neu anmelden (siehe A.)

C. CQP aufrufen, Korpus laden/wechseln, erste Eingaben:

- ▶ Verbindung zum Login-Server herstellen (siehe A.)
- ▶ USER@login:~\$ **cqp** startet CQP
- ▶ [no corpus]> **show corpora** listet Korpora auf (optional)
- ▶ [no corpus]> **BNC** lädt BNC-Korpus
- ▶ **BNC> BROWN** wechselt z. BROWN-Korpus
- ▶ **BROWN> "fabulous"** erstellt Konkordanz zu 'fabulous'

Navigationsbefehle in cqp A.2

- LEERTASTE** Konkordanz/Info *seitenweise* durchblättern
- PFEILTASTE** Konkordanz/Info *zeilenweise* durchblättern
- q** Konkordanz/Info verlassen (ohne ENTER)
- cat Last** Letzte Konkordanz erneut ausgeben
- A = Last** Speichert letzte Konkordanz als „A“
(neue Suche überschreibt vorige als **Last**)
- show** Zeigt Korpora/gespeicherte Konkordanzen
- show cd** Zeigt verfügbare Attribute/Annotation
- cat Last > "DATEINAME.txt"**
Speichert **Last** als Datei im Userspace
www.zedat.fu-berlin.de > Datenablage
- exit** CQP verlassen (Server verlassen: 2x **exit**)

Positionsattribute in cq@fu A.3

(Verfügbare Attribute: info BROWN oder show cd)

- word** ‚as is‘, d.h. wie im Korpus/Text
 - lemma** Grundform (im BNC: **hw**, *head word*)
 - pos** Part-of-Speech-Tag (Wortklasse)
 - class** Übergeordnete Wortklasse, z.B. **SUBST**, **VERB**, **ADV**
 - show +pos** Attribut **pos** in Konkordanz ausgeben
 - show -pos +class** Attribut aus- (-) oder einschalten (+)
- (Historische Korpora nicht lemmatisiert; i.d.R. **word/pos**)

Elementare CQP-Befehle B.1

Das CQP-Prinzip: **ein Token in eine Eckklammer**. Innerhalb der Klammer können für das Token beliebig viele & sehr komplexe Bedingungen formuliert & kombiniert werden:

[attribut₁="WERT" & attribut₂="WERT" & attribut₃="(WERT₁|WERT₂|WERT₃)" ...]

Beispielhaft am BROWN-Korpus (Penn-Tagset, s. Infodatei **info BROWN**):

- | | | |
|---|------------|--|
| BROWN> "jump" | (Kurzform) | } ‚as-is‘, findet <i>jump</i> , aber nicht <i>jumps</i> , <i>jumping</i> , <i>jumped</i> |
| BROWN> [word="jump"] | (Langform) | |
| BROWN> "jumps?(ed)?" | | <i>jump</i> , <i>jumps</i> , <i>jumped</i> (s. „Reguläre Ausdrücke“) |
| BROWN> [word="jump(ing)?s?(ed)?" | | wie oben, nur anders (und mit <i>-ing</i> -Form) |
| BROWN> [lemma="jump"] | | Lemma: <i>jump</i> , <i>jumps</i> , <i>jumping</i> , <i>jumped</i> |
| BROWN> [lemma="jump" & pos="V.+"] | | Verb formen von <i>jump</i> (aber nicht Nomen) |
| BROWN> [lemma="round" & pos="NN JJ"] | | <i>round</i> , die Nomen oder Adjektiv sind |
| BROWN> [lemma="round" & pos!="N.*"] | | <i>round</i> , die nicht (!=) Nomen sind |
| BROWN> [lemma="round" & class="VERB" & pos!="VBG"] | | <i>round</i> als Verb, außer <i>-ing</i> -Form |
| BROWN> [pos="V.*"] | | alle Verben (in Brown: keine Modalverben) |
| BROWN> [pos="V(V H).*"] | | alle <i>lexikalisch en</i> Verben & <i>have</i> -Formen |
| BROWN> [] | | irgendein Token (hier: ALLE Korpus-tokens) |

Beispiele für Mehrwortsuchen (z.B. Phrasen, Kollokationen, viele Komposita, etc.):

- | | |
|---|---|
| BROWN> [lemma="the"] [class="SUBST"] | <i>the</i> gefolgt von N (sg./pl./Eigennamen) |
| BROWN> [lemma="the"] [pos="NNS"] | <i>the</i> gefolgt von N im Plural (NNS) |
| BROWN> [class="ADV"] [pos="JJ"] | alle ADV-ADJ-Sequenzen („Bigramme“) |
| BROWN> [pos="JJ"] {2} | alle ADJ-ADJ-Sequenzen („Bigramme“) |
| BROWN> [word="the"] [] {0,2} [pos="N.*"] | <i>the</i> , gefolgt von 0–2 Token, gefolgt von N |
| BROWN> [] "and" [] | <i>and</i> ; jeweils 1 Treffer links & rechts |
| BROWN> [pos="N.*"] "and"%c [pos="N.*"] | <i>N-and-N</i> -Koordination (%c: <i>and</i> , <i>And</i> , <i>AND</i> ...) |

Namen von Attributen müssen in Kleinbuchstaben geschrieben werden (also **pos**, **lemma**, **class**, **word**). Die Werte bei **pos** müssen in Großbuchstaben eingegeben werden, z.B. **JJ** = Adjektiv, **NN** = Nomen. Werte von **word** sind ‚as is‘-Korpusformen und werden gefunden, wie eingegeben; um Groß- & Kleinschreibung zu finden, muss bei **word %c** angehängt werden: **[word="the"]** bzw. **"the"** findet *the*, aber **[word="the"%c]** oder **"the"%c** findet *the*, *The*, *THE*, *thE*, etc. **%d** ignoriert Diakritika, **%cd** ignoriert Groß-/Kleinschreibung & Diakritika gleichzeitig.

Werte von Attributen (alles zwischen den Anführungszeichen) können mit Platzhaltern und Operatoren erweitert werden, was Recall & Präzision der Abfragen verbessert. Siehe Infobox „Reguläre Ausdrücke“.

Know-your-corpus-Hinweis: Tagsets sind korpusabhängig! Siehe **info BROWN / show cd**.

Reguläre Ausdrücke & Operatoren (Auswahl) B.2

- irgendein Zeichen

Wiederholungsoperatoren
(für voriges Zeichen/Gruppe):

- * 0 oder mehr mal
- + 1 oder mehr mal
- ? 0 oder 1 mal
- {3} genau 3 mal
- {1,} min. 1 mal
- {2,3} min. 2, max. 3 mal
- & UND
- | ODER
- != NICHT
- () Sequenz gruppieren
- [] Klasse gruppieren
- \ ‚Escape‘-Zeichen, d.h.
- \? findet „echtes“ ? im Text
- \. findet „echten“ .
- %c *case-insensitive*

Beispiele:

- "tests?"** optionales s: *test*, *tests*
- "tests" "\?"** ... *tests*?
- "s.ng"** *sing*, *sang*, *sung*, *song*
- "s[iaun]g"** *sing*, *sang*, *sung*
- "improve.*"** *improve(d)*, *improves*, *improvement(s)*, ...
- "improve?.*"** wie oben + *improving*
- "improv.+"** alle *improve*-Derivate
- "analy(s|z)e"** *analyze*, *analyse*
- "yes{1,}"** *yes*, *yess*, *yesss*, *yessss*...
- [pos="VB.*"]** VB, VBG, VBN, VBD
- [pos="VB.+"]** VB-Tags, außer VB
- [pos="(VB|NN)"]** VB oder NN
- [lemma=".*i(z|s)e" & class="V-*"]**
Verblemmas auf *-ize/-ise*
(also nicht *prize* od. *sunrise*)

CQP-Einstellungen ändern

A.4

Allgemeine Einstellungen sind in der Datei .cqprc im Homeverzeichnis des FU-Userspace gespeichert. Sie lassen sich dort dauerhaft ändern (wer sich mit Linuxbefehlen/-editoren auskennt) oder in CQP für die *aktuelle* Sitzung setzen (wem der erste Teil wenig bis gar nichts sagt):

```
set Context 70           70 Zeichen links & rechts (cq@fu-Default: 30); 0 gibt nur Beleg aus
set Context 5 words     5 Token links & rechts
set Context 1s          1 Satz links & rechts (statt Zeichen)
set LeftContext 70      70 Zeichen links (geht natürlich auch: RightContext)
```

```
set PrintStructures 'text_id'           gibt Korpusdateinamen aus (Inhalt text_id-Tag)
set PrintStructures 'text_id, text_year' gibt Jahr & Dateinamen aus
```

Muster:

```
set PrintStructures 'STRUCTURALATTRIBUTE1, STRUCTURALATTRIBUTE2, STRUCTURALATTRIBUTE3, ... '
```

Achtung: Werte für `set PrintStructures` sind korpusabhängig, weil sie oft unterschiedlich heißen, d.h. sie müssen beim Laden eines anderen Korpus neu gesetzt werden. Welche Strukturattribute für welche Korpora verfügbar sind: `info BROWN` oder `BNC> show cd`. Wichtig: wenn Metainformationen mit in Textdatei exportiert werden sollen, müssen sie vor dem Export in CQP aktiviert werden. Denn ist die .txt-Datei geschrieben, kann diese nicht mehr mit Metainformationen angereichert werden (sondern muss neu abgefragt & exportiert werden).

Frequenzlisten in CQP erstellen

C.1

Aus Treffern eines Suchmusters können Frequenzlisten erstellt werden, gezählt anhand von Positionsattributen (Muster: `count NAMEDESUCHANFRAGE by ATTRIBUT on POSITION`). Bei Mustern mit mehreren Token können Referenzpunkte spezifiziert werden: bei `match[ZAHL]` beginnt CQP am Anfang des Suchmusters zu zählen, mit `matchend[ZAHL]` am Ende. CQP beginnt bei 0 mit dem Zählen, 1 ist eine Position rechts, -1 eine Position links des Referenzpunkts (usw.); es kann also auch außerhalb des Suchmusters im Kontext des Treffers gezählt werden. Eine simple NP zum Ausprobieren: `BROWN> [pos="DT"] [class="SUBST"]:`

```
count Last by word           Frequenzen des Suchmusters der letzten Eingabe
count A by lemma             Wie oben, hier: zählt nach Lemmas der Anfrage A
count Last by pos on match[0] Zählt nach Lemmas des 1. Tokens des Suchmusters
count Last by pos on matchend[0] Zählt nach POS des letzten Tokens des Suchmusters
count Last by pos on matchend[-1] Zählt nach POS des zweitletzten Tokens des Musters
count Last by pos on matchend[1] Zählt nach POS des 1. Tokens rechts des Musters (R1)
count Last by word on match[-1] Zählt nach Wort des 1. Tokens links des Musters (L1)
```

Frequenzlisten können in eine Datei exportiert werden: `count Last by word > "freqlist.txt"`. Dateiname kann frei gewählt werden (Dateiendung ist optional, aber sinnvoll; "" ist Pflicht).

Strukturattribute aufrufen // nach Strukturattributen eingrenzen

D.1

Alle Korpora haben Strukturattribute (=Metainformationen), die in der Konkordanz mit ausgegeben werden können und/oder auf die eine Suchanfrage eingegrenzt werden kann. Verfügbare Attribute sind korpuspezifisch; einsehbar über `show cd` (bei geladenem Korpus). Muster: `CORPUS->[QUERY]::match.ATTRIBUTE="VALUE"`. Diese können beliebig kombiniert und mit regulären Ausdrücken verwendet werden, wie in Suchanfragen. Eine Auswahl:

```
BNC>"fabulous"::match.text_mode="spoken" Findet fabulous in gesprochener Sprache (BNC)
BNC>"lovely"::match.u_sex="(male|female)" Findet lovely im BNC (spoken part!), wenn von Frauen oder Männern geäußert, aber nicht, wenn Geschlecht unbekannt oder nicht verfügbar ist.
```

```
BROWN>"democracy"%c:::match.text_genre="PRESS"
democracy im Genre „Presse“ in BROWN
```

```
PPCEEC>"de[ea]r.*"%c:::match.letter_relToRec="MOTHER" & match.letter_auth_age="(2|3)."
Dear, deer, deare, dearest etc. in Briefen, deren Schreiber/innen zwischen 20–39 Jahre alt sind (und wo das Alter bekannt ist!) und die an ihre Mütter geschickt wurden (wo die Verwandtschaftsbeziehung bekannt ist!).
```

```
set PrintStructures 'text_mode' wenn gesetzt, gibt es text_mode in Konkordanz aus (und wird beim Speichern in Datei (cat Last > "KONKORDANZ.txt") mit exportiert (bzw. alle, die gesetzt sind).
```

Wichtig: Strukturattribute sind nicht selbsterklärend! Verfügbarkeit, Darstellung und Ausprägungen der Werte (inkl. fehlender Werte) müssen bekannt sein und einkalkuliert werden. Hier gibt es lediglich grundlegende Hinweise/Übersichten CQP-Prinzip, um auf diese Informationen zuzugreifen (diese sind auf allen CQP-Systemen gleich). Es muss vor der Arbeit *immer* die Dokumentation des Korpus konsultiert werden! Technische Änderungen und/oder zusätzliche Annotationen gegenüber den offiziellen Releases der Korpora sind in der Infodatei `info CORPUS` vermerkt.

E

Troubleshooting

(im Aufbau)

Feedback: susanne.flach@fu-berlin.de

- ▶ Regel 1: CQP macht, was man ihr sagt.
- ▶ Regel 2: Sag nix falsches!
- ▶ Regel 3: Probleme liegen zu 99% bei Nutzer/innen & Unerfahrenheit.
- ▶ Regel 4: Practice makes (perfect|better).
- ▶ Regel 5: CheatSheet nicht unfehlbar.

Kurzabhilfe:

- **Alles** in CQP ist case-sensitive:
 - Groß-/Kleinschreibung prüfen!
 - Eckige Klammern alle da?
 - Anführungszeichen richtig?
 - Korpus geladen?
 - Richtiges Korpus geladen?
 - Gewünschtes Attribut im Korpus überhaupt vorhanden?
 - Attributnamen korrekt eingegeben?
 - Attributwerte korrekt eingegeben?
 - Reguläre Ausdrücke richtig?
 - Attribute richtig verknüpft?
-
- ▶ Verbindung bricht ab (Windows)?
 - PuTTY auf Desktop speichern und von dort starten oder auf UNIX-System umsteigen.
 - ▶ CQP „hängt“? Manche Prozesse brauchen sehr lange, je nach Komplexität (z.B. `count`)
 - ggf. mit `Ctrl+Strg+C` abbrechen
 - ▶ CQP sagt „Syntax error“?
 - Your query is ill-formed
 - Checklist „Kurzabhilfe“

Penn-Tagset (Extended)

(BROWN, FROWN, LOB, FLOB, CLMET)

POS	CLASS	BESCHREIBUNG	PENN
CC	CONJ	coordinating conjunction	CC
CD	ADJ	cardinal number	CD
DT	ART	determiner	DT
EX	PRON	existential <i>there</i>	EX
FW	UNC	foreign word	FW
IN	PREP	preposition/subord. conj.	IN
IN/that	COMP	complementizer	–
JJ	ADJ	adjective	JJ
JJR	ADJ	adjective, comparative	JJR
JJS	ADJ	adjective, superlative	JJS
MD	VERB	modal	MD
NN	SUBST	noun, singular or mass	NN
NNS	SUBST	noun plural	NNS
NP	SUBST	proper noun, singular	NNP
NPS	SUBST	proper noun, plural	NNPS
PDT	ART	predeterminer	PDT
POS	SUBST	possessive ending	POS
PP	PRON	personal pronoun	PRP
PP\$	PRON	possessive pronoun	PRP\$
RB	ADV	adverb	RB
RBR	ADV	adverb, comparative	RBR
RBS	ADV	adverb, superlative	RBS
RP	ADV	particle	RP
SENT	SENT	end punctuation	SENT
SYM	SYM	symbol	SYM
TO	PREP	to go, to him	TO
UH	INTJ	interjection	UH
VB	VERB	<i>be</i> , base form	VB
VBD	VERB	<i>be</i> , past	VBD
VBG	VERB	<i>be</i> , gerund/participle	VBG
VBN	VERB	<i>be</i> , past participle	VBN
VBZ	VERB	<i>be</i> , pres, 3rd p. sing	VBZ
VBP	VERB	<i>be</i> , pres non-3rd p.	VBP
VD	VERB	<i>do</i> , base form	VD
VDD	VERB	<i>do</i> , past	VDD
VDG	VERB	<i>do</i> gerund/participle	VDG
VDN	VERB	<i>do</i> , past participle	VDN
VDZ	VERB	<i>do</i> , pres, 3rd per. sing	VDZ
VDP	VERB	<i>do</i> , pres, non-3rd per.	VDP
VH	VERB	<i>have</i> , base form	VH
VHD	VERB	<i>have</i> , past	VHD
VHG	VERB	<i>have</i> , gerund/participle	VHG
VHN	VERB	<i>have</i> , past participle	VHN
VHZ	VERB	<i>have</i> , pres 3rd per. sing	VHZ
VHP	VERB	<i>have</i> , pres non-3rd per.	VHP
VV	VERB	verb, base form	VB
VVD	VERB	verb, past tense	VBD
VVG	VERB	verb, gerund/participle	VBG
VVN	VERB	verb, past participle	VBN
VVP	VERB	verb, present, non-3rd p.	VBP
VVZ	VERB	verb, present 3rd p. sing.	VBZ
WDT	ART	wh-determiner	WDT
WP	PRON	wh-pronoun	WP
WP\$	PRON	possessive wh-pronoun	WP\$
WRB	ADV	wh-abverb	WRB
:	SENT	general joiner	:
\$	SYM	currency (not systematic)	\$

Know-your-corpus-Hinweis:

Das Penn-Tagset hat zwei Varianten: (1) eine ausführliche (Spalte 1), die bei cq@fu verwendet wird, sowie (2) die Standardvariante (Spalte 4, PENN), die in vielen anderen Korpusystemen verwendet wird (z.B. COW). Sie unterscheiden sich v.a. in der Detailtiefe im Bereich der Verben.

CLAWS-5 Tagset

(BNC, BNC-BABY)

POS	CLASS	BESCHREIBUNG
AJ0	ADJ	Adjective (general / positive) (e.g. <i>good, old, beautiful</i>)
AJC	ADJ	Comparative adjective (e.g. <i>better, older</i>)
AJS	ADJ	Superlative adjective (e.g. <i>best, oldest</i>)
AT0	ART	Article (e.g. <i>the, a, an, no</i>)
AV0	ADV	General adverb (e.g. <i>often, well, longer (adv.), furthest</i>)
AVP	ADV	Adverb particle (e.g. <i>up, off, out</i>)
AVQ	ADV	Wh-adverb (e.g. <i>when, where, how, why, wherever</i>)
CJC	CONJ	Coordinating conjunction (e.g. <i>and, or, but</i>)
CJS	CONJ	Subordinating conjunction (e.g. <i>although, when</i>)
CJT	CONJ	The subordinating conjunction <i>that</i>
CRD	ADJ	Cardinal number (e.g. <i>one, 3, fifty-five, 3609</i>)
DPS	PRON	Possessive determiner-pronoun (e.g. <i>your, their, his</i>)
DT0	ADJ	General determiner-pronoun (not a DTQ or an AT0)
DTQ	PRON	Wh-determiner-pronoun (e.g. <i>which, what, whichever</i>)
EX0	PRON	Existential <i>there</i> , i.e. <i>there is, there are</i>
ITJ	INTERJ	Interjection/isolate (e.g. <i>oh, yes, mhm, wow</i>)
NN0	SUBST	Common noun, neutral for num (e.g. <i>data, committee</i>)
NN1	SUBST	Singular common noun (e.g. <i>pencil, goose, time</i>)
NN2	SUBST	Plural common noun (e.g. <i>pencils, geese, times</i>)
NP0	SUBST	Proper noun (e.g. <i>London, Michael, Mars, IBM</i>)
ORD	ADJ	Ordinal numeral (e.g. <i>first, sixth, 77th, last</i>) .
PNI	PRON	Indefinite pronoun (e.g. <i>none, everything, nobody</i>)
PNP	PRON	Personal pronoun (e.g. <i>I, you, them, ours</i>)
PNQ	PRON	Wh-pronoun (e.g. <i>who, whoever, whom</i>)
PNX	PRON	Reflexive pronoun (e.g. <i>myself, yourself, itself</i>)
POS	UNC	The possessive or genitive marker 's or '
PRF	PREP	The preposition <i>of</i>
PRP	PREP	Preposition (except <i>of</i>) (e.g. <i>about, at, in, on, with</i>)
PUL	STOP	Punctuation: left bracket, i.e. (or [
PUN	STOP	Punctuation: general separating mark (, , ! : ; – and ?)
PUQ	STOP	Punctuation: quotation mark (' and ")
PUR	STOP	Punctuation: right bracket, i.e.) or]
T00	PREP	Infinitive marker <i>to</i>
UNC	UNC	Unclassified items
VBB	VERB	Present tense forms of BE (except for <i>is</i> and 's)
VBD	VERB	Past tense of BE (<i>was, were</i>)
VBG	VERB	-ing form of BE (<i>being</i>)
VBI	VERB	Infinitive form of BE (<i>be</i>)
VBN	VERB	Past participle of BE (<i>been</i>)
VBZ	VERB	-s form of BE (<i>is, 's</i>)
VDB	VERB	Finite base of DO (<i>do</i>)
VDD	VERB	Past tense of DO (<i>did</i>)
VDG	VERB	-ing form of DO (<i>doing</i>)
VDI	VERB	Infinitive of DO (<i>do</i>)
VDN	VERB	Past participle of DO (<i>done</i>)
VDZ	VERB	-s form of the verb DO (<i>does, 's</i>)
VHB	VERB	Finite base of the verb HAVE: <i>have, 've</i>
VHD	VERB	Past tense of the verb HAVE: <i>had, 'd</i>
VHG	VERB	-ing form of the verb HAVE: <i>having</i>
VHI	VERB	Infinitive of the verb HAVE: <i>have</i>
VHN	VERB	Past participle form of the verb HAVE: <i>had</i>
VHZ	VERB	The -s form of the verb HAVE: <i>has, 's</i>
VM0	VERB	Modal auxiliary verb (e.g. <i>will, would, can, 'll, 'd</i>)
VVB	VERB	Finite base, lexical verbs (e.g. <i>forget, send, live</i>)
VVD	VERB	Past tense, lexical verbs (e.g. <i>forgot, sent, lived</i>)
VVG	VERB	-ing form, lexical verbs (e.g. <i>forgetting, sending</i>)
VVI	VERB	Infinitive, lexical verbs (e.g. <i>forget, send, live, return</i>)
VVN	VERB	Past participle, lexical verbs (e.g. <i>forgotten, lived, sent</i>)
VVZ	VERB	-s form of lexical verbs (e.g. <i>forgets, sends, lives</i>)
XX0	ADV	Negative particle <i>not</i> or <i>n't</i>
ZZ0	SUBST	Alphabetical symbols (e.g. <i>A, a, B, b, c, d</i>)

Know-your-corpus-Hinweis:

Manche Webversionen des BNC verwenden das CLAWS-7-Tagset.

Structural Attributes for BRITISH NATIONAL CORPUS (BNC, BNC-BABY)

ATTRIBUT	WERT	BNC>[word="water"%c]::match.text_genre="W:ac:(medicine nat_science)" BNC>[word="fantastic"]::match.text_mode="spoken" & match.u_sex="female"			
Textinformationen					
text_id, text_title, text_n_words, text_n_token, text_n_w, text_n_s, text_n_c,	Zahlreich. Attribute hauptsächlich für set PrintStructures und/oder count / tabulate-Befehle .				
text_publication_date	1960-1974, 1975-1984, 1985-1993, unknown				
text_text_type	written_books_periodicals, written_misc, written-to-be-spoken, spoken_context, spoken_demographic				
text_context	public, educational, leisure, business, ---				
text_respondent_age	0-14, 15-24, 25-34, 35-44, 45-59, 60+, ---				
text_respondent_class	AB, C1, C2, DE, unknown, ---				
text_respondent_sex	female, male, unknown, ---				
text_interaction_type	dialogue, monologue, ---				
text_region	south, north, midlands, unknown, ---				
text_author_age	0-14, 15-24, 25-34, 35-44, 45-59, 60+, unknown, ---				
text_author_domicile	UK, Commonwealth, USA, Europe, other, unknown, ---				
text_author_sex	female, male, mixed, unknown, ---				
text_author_type	corporate, multiple, sole, unknown, ---				
text_audience_age	any, adult, child, teen, ---				
text_domain	applied_science, arts, belief_thought, commerce_finance, imaginative, leisure, natural_science, social_science, world_affairs, ---				
text_difficulty	medium, high, low, ---				
text_medium	book, periodical, misc_published, misc_unpublished, to-be-spoken, ---				
text_publication_place	UK, UK_midlands, UK_north, UK_south, Ireland, USA, unknown				
text_sampling_type	beginning, middle, end, whole, composite, unknown, ---				
text_circulation	high, medium, low, unknown, ---				
text_audience_sex	female, male, mixed, unknown, ---				
text_availability	cleared, unclaimed				
text_mode	written, spoken				
text_derived_type	prose, misc_published, fiction, academic, newspaper, spoken_other, spoken_conversation, unpublished				
text_genre (spoken)	S:brdcast:discussn S:brdcast:documentary S:brdcast:news	S:lect:commerce S:lect:humanities_arts S:lect:nat_science S:lect:polit_law_edu S:lect:soc_science	S:conv S:sportslive S:tutorial S:consult S:courtroom S:demonstratn	S:unclassified S:meeting S:parliament S:pub_debate S:sermon	S:interview S:interview:oral_history S:speech:scripted S:speech:unscripted
see also: text_mode text_text_type					
text_genre (written)	W:ac:humanities_arts W:ac:medicine W:ac:nat_science W:ac:polit_law_edu W:ac:soc_science W:ac:tech_engin	W:news_script W:newsp:brdsht_nat:arts W:newsp:brdsht_nat:commerce W:newsp:brdsht_nat:editorial W:newsp:brdsht_nat:misc W:newsp:brdsht_nat:report W:newsp:brdsht_nat:science W:newsp:brdsht_nat:social W:newsp:brdsht_nat:sports W:newsp:other:arts W:newsp:other:commerce W:newsp:other:report W:newsp:other:science W:newsp:other:social W:newsp:other:sports W:newsp:tabloid	W:fict:drama W:fict:poetry W:fict:prose	W:pop_lore W:religion	W:admin W:advert W:biography W:commerce W:email W:hansard W:institut_doc W:instructional W:letters:personal W:letters:prof W:misc
see also: text_mode text_text_type	W:non_ac:humanities_arts W:non_ac:medicine W:non_ac:nat_science W:non_ac:polit_law_edu W:non_ac:soc_science W:non_ac:tech_engin W:essay:school W:essay:univ				
Utterance-Informationen (spoken part only)					
u_who, u_age, u_name, u_occupation, u_role, u_n_words, u_n_tokens, u_n_w, u_n_c, u_n_u, u_n_s	Zahlreiche. Entweder exakte Nummern oder unsystematische Werte.				
u_age_group	0-14, 15-24, 25-34, 35-44, 45-59, 60+, unknown, ---				
u_dialect	Home_Counties, Other, London, Scottish, NW_Midlands, central_Northern_England, Welsh, central_Midlands, central_SW_England, East_Anglian, Irish, Lancashire, Midlands, NE_Midlands, Merseyside, NE_England, lower_SW_England, S_Midlands, Northern_England, European, Humberside, upper_SW_England, American, German, West Indian, Indian, Canadian, French, none, ---				
u_education	in_school, 14-, 19+, unknown, ---				
u_first_language	British, American, German, French, Indian, unknown, ---				
u_sex	female, male, unknown, ---				
u_class	AB, C1, C2, DE, unknown, ---				
Weitere strukturelle Attribute (Textdarstellung/-struktur, z.B. Kursiv-/Fettdruck, Kapitelüberschriften etc): siehe show cd und BNC-Handbuch.					

Know-your-corpus-Hinweis: Diese Übersicht ist eine Nutzungsanleitung für das Aufrufen der Informationen in CQP — wofür sie stehen und was sie für korpuslinguistische Arbeit bedeuten bitte unbedingt in offiziellen BNC-Dokumentationen nachschlagen!