

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

- ▶ Windows: PuTTY öffnen (http://www.zedat.fu-berlin.de/tip4u_03.pdf)
(Tipp: 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. `LOKALERNUTZERNAME$ 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]
(auf Leerzeichen achten — no error, all good!)
- ▶ 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 (betritt CQP-Raum)*
- ▶ `[no corpus]> show corpora` *listet Korpora auf (optional)*
- ▶ `[no corpus]> BNC` *lädt BNC-Korpus*
- ▶ `BNC> BROWN` *wechselt zum 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)
- info KORPUSNAME** Korpusinfos (Tags, Inhalt, Doku etc.)
- show** Zeigt Korpora/gespeicherte Konkordanzen
- show cd** Zeigt verfügbare Attribute/Annotation
- cat Last** Letzte Konkordanz wieder ausgeben
- A = Last** Speichert letzte Konkordanz als „A“
(neue Suche wird als **Last** gespeichert)
- cat A > "DATEINAME.txt"**
Speichert **A** (od. **Last**) als Datei im Userspace
www.zedat.fu-berlin.de → Datenablage
- exit** CQP-Raum verlassen (Server verlassen: 2x)

Positionsattribute in cq@fu A.3

- (Verfügbare Attribute: **info KORPUSNAME** oder **show cd**)
- word** ‚as is‘, d.h. wie im Korpus/Text
 - lemma** Grundform (Achtung: im BNC **hw**, für *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. Muster:

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

Beispielhaft am BROWN-Korpus (Penn-Tagset, siehe jeweilige Infodatei `info KORPUSNAME`):

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

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

- | | |
|--|---|
| <code>BROWN> [lemma="the"] [class="SUBST"]</code> | <i>the</i> gefolgt von N (sg./pl./Eigennamen) |
| <code>BROWN> [lemma="the"] [pos="NNS"]</code> | <i>the</i> gefolgt von N im Plural (NNS) |
| <code>BROWN> [class="ADV"] [pos="JJ"]</code> | alle ADV-ADJ-Sequenzen („Bigramme“) |
| <code>BROWN> [pos="JJ"]{2}</code> | alle ADJ-ADJ-Sequenzen („Bigramme“) |
| <code>BROWN> [word="the"] []{0,2} [pos="N.*"]</code> | <i>the</i> , gefolgt von 0–2 Token, gefolgt von N |
| <code>BROWN> [] "and" []</code> | <i>and</i> ; + 1 Treffer links (L1) & rechts (R1) |
| <code>BROWN> [pos="N.*"] "and"%c [pos="N.*"]</code> | <i>N-and-N</i> -Koordination (%c: <i>and, And, 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 (bei fast allen Korpora; abhängig von Tagsets). 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 KORPUSNAME`.

Tipp: Navigieren im Verlauf der Suchanfragen spart Zeit-/Tipparbeit (mit Pfeiltasten ↑↓)

Reguläre Ausdrücke & Operatoren (Auswahl) B.2

- `.` irgendein Zeichen
- Wiederholungsoperatoren**
(für voriges Zeichen oder 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[ia]ng"` *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, yesssss...*
- `[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 z.B. nicht *prize* oder *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 Editoren wenig bis gar nichts sagen):

```
set Context 70          70 Zeichen links & rechts (cq@fu-Default: 30)
set Context 5 words    5 Token links & rechts
set LeftContext 70     70 Zeichen links (rechts: set RightContext)

set PrintStructures 'text_id'          zeigt Korpusdateinamen an (Inhalt text_id-Tag)
set PrintStructures 'text_id, text_year' Korpusdateinamen & Jahr (wo vorhanden!)
```

Muster:

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

Achtung: Werte für `set PrintStructures` sind korpusabhängig, weil sie meist unterschiedlich heißen; d.h. beim Laden eines anderen Korpus neu setzen. Welche Strukturattribute je Korpus verfügbar sind: `info KORPUSNAME` oder `KORPUSNAME> show cd` (siehe auch D.1). Wichtig: wenn Metainformationen mit exportiert werden sollen, müssen sie vor der Abfrage in COP aktiviert werden.

Konkordanz exportieren, aufräumen & lokal weiterarbeiten

A.5

1. Im CQP-Raum (auf Anführungszeichen achten):

```
KORPUSNAME> cat Last > "meineKonkordanz.txt"  schreibt Datei in Userspace-Raum
KORPUSNAME> exit                             verlässt CQP-Raum, betritt Userspace-Raum
```

2. Im Userspace-Raum Aufräumskript ausführen (hier keine Anführungszeichen):

```
USER@login~$ perl CWBclean.pl meineKonkordanz.txt > meineKonkordanz_clean.txt
Tabsepariert Konkordanz; Dateidownload über ZEDAT-Portal → Datenablage
```

3. Import in Spreadsheetprogramme (empfohlen wird Open/LibreOffice)

Libre/OpenOffice Neues Tabellendokument → „Einfügen“ → „Tabelle aus Datei“; im Dialog auf UTF-8-Kodierung achten; „Tabulator“ wählen; bei Texttrenner “ löschen

MS-Excel Im Explorer Rechtsklick auf Datei → „Öffnen mit“ → „Excel“

(Bei Excel empfohlen: Datei in gutem Texteditor (nicht Word!) öffnen, Strg+A, Strg+C & mit Strg+V in Excel kopieren.)

Frequenzlisten in CQP erstellen

C.1

Frequenzlisten nach Positionsattributen: `count QUERYNAME by ATTRIBUT on POSITION`. Für Muster mit mehreren Token spezifiziert man Referenzpunkte: bei `match[ZAHL]` zählt CQP ab Anfang des Suchmusters, mit `matchend[ZAHL]` ab Ende. CQP beginnt bei 0 mit dem Zählen, 1 ist eine Position rechts, -1 eine Position links des Referenzpunkts (usw.). (Analog: `sort`-Befehl; siehe auch C.2) NP zum Ausprobieren: `BROWN> [pos="DT"] [class="SUBST"]`:

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

count-Output als Frequenzliste speichern: `count Last by word on match[-1] > "DATEINAME.txt"`

Übersicht erweiterte CWB-Befehle

C.2

Sortieren & Zählen

Muster: `sort QUERYNAME by ATTRIBUT on POSITION`

→ `sort Last by pos on match[-1]`

Syntax für `sort` & `count` identisch (→ C.1).

Randomisieren & randomisiertes Sampling

```
sort Last randomize  randomisiert Last
sort Last           Last zurücksetzen
reduce A to 15%     15%-Sample aus A
reduce B to 200     200-Belege-Sample aus B
```

(Vor `reduce` die Konkordanz besser einer neuen Variable zuweisen, d.h. `A = Last`; `reduce` ändert Inhalt.)

Kreuztabellen

Muster: `tabulate QUERY match ATTR1, match ATTR2`

```
→ group Last match text_mode
→ tabulate A match u_sex, match u_class >
   "|sort |uniq -c|sort -nr "
```

Output kann in Datei umgeleitet werden:

```
group B match ATTRIBUTE > "GROUPRESULT.txt"
tabulate Last match u_sex, match u_age >
   "|sort |uniq -c|sort -nr > TABRESULT.txt"
```

Strukturattribute aufrufen, anzeigen & Suchen einschränken

D.1

Korpora haben Strukturattribute (=Metainfos), die mit ausgegeben werden oder Suchanfragen eingrenzen können (beliebig kombiniert, auch mit RegEx). Attribute & Werte sind *korpuspezifisch* (im Muster in KAPITÄLCHEN); siehe `info KORPUSNAME`. Muster: `KORPUSNAME>[TOKENQUERY]::match.ATTRIBUTE="VALUE"`.

Suchanfragen nach Strukturattributen eingrenzen (Auswahl!):

```
BNC>[hw="tradition"]::match.text_mode="spoken"  tradition, BNC spoken-part
BNC>[hw="evidence"]::match.text_genre="W:ac.*"  evidence, BNC, „Academic“
BROWN>"democracy"%c::match.text_genre="PRESS"  democracy, BROWN, „Presse“
BNC>"lovely"::match.u_sex="(male|female)" & match.text_mode="spoken"
lovely im BNC-spoken; Frau oder Mann (aber nicht, wo u_sex unknown/na ist)
```

Anzeigen lassen von Strukturattributen (Auswahl!):

```
KORPUSNAME>set PrintStructures 'text_genre'
KORPUSNAME>set PrintStructures 'text_mode, u_sex, u_class'
gibt METAINFO in Konkordanz mit aus, auch im Export (cat A > "KONKORDANZ.txt")
```

Wichtig: Strukturattribute sind nicht selbsterklärend! Hintergründe & Eigenschaften müssen bekannt sein und kritisch einkalkuliert werden. CheatSheet gibt nur grundlegende Übersichten zum Zugriffsprinzip in CQP. Bitte vor/bei der Arbeit immer die offizielle Dokumentation des Korpus konsultieren! Technische Änderungen sowie zusätzliche Annotationen gegenüber den offiziellen Korpusreleases sind in der Infodatei `info KORPUSNAME` vermerkt.

E

Troubleshooting

Feedback: susanne.flach@fu-berlin.de

- ▶ Regel 1: CQP macht, was man sagt.
- ▶ Regel 2: Sag nix falsches!
- ▶ Regel 3: Probleme liegen zu 99,9% bei Unerfahrenheit — deshalb:
- ▶ 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 & richtig?
 - 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
 - Unter „Connections“ *Seconds between keepalives* auf 120 setzen, unter ‚Session‘ *Default Setting* anklicken und ‚Save‘, um Einstellung zu speichern.
 - ▶ **CQP „hängt“?**
 - Einige Prozesse brauchen sehr lange, je nach Menge/Korpusgröße (z.B. `count`)
 - ggf. mit `Ctrl/Strg+C` abbrechen
 - ▶ **CQP sagt „Syntax error“?**
 - Your query is ill-formed
 - Checklist „Kurzabhilfe“
 - Meist fehlt ein [oder ein "
 - ▶ **CQP liefert keine Ergebnisliste?**
 - Dann gibt es kein Ergebnis
 - Wahl/Eingabe der Attribute prüfen

Penn-Tagset

B

(BROWN, FROWN, LOB, FLOB, CLMET)

POS	CLASS	BESCHREIBUNG	BEISPIEL
CC	CONJ	coord. conjunction	<i>and, or</i>
CD	ADJ	cardinal number	<i>one, two</i>
DT	ART	determiner	<i>the, all, those</i>
EX	PRON	existential <i>there</i>	
FW	UNC	foreign word	
IN	PREP	prep/subord. conj.	<i>in, of, over</i>
JJ	ADJ	adjective	<i>great, nice</i>
JJR	ADJ	adjective, comparative	<i>greater</i>
JJS	ADJ	adjective, superlative	<i>niciest</i>
MD	VERB	modal	<i>will, could</i>
NN	SUBST	noun, singular or mass	<i>body, world</i>
NNS	SUBST	noun plural	<i>bodies</i>
NP	SUBST	proper noun, singular	<i>Paul, Mike</i>
NPS	SUBST	proper noun, plural	
PDT	ART	predeterminer	<i>all, such</i>
POS	SUBST	possessive ending	<i>'s</i>
PP	PRON	personal pronoun	<i>I, you, he</i>
PP\$	PRON	possessive pronoun	<i>its, their</i>
RB	ADV	adverb	<i>though, now</i>
RBR	ADV	adverb, comparative	<i>more, better</i>
RBS	ADV	adverb, superlative	<i>most, best</i>
RP	ADV	particle	<i>out, upon</i>
SENT	PUNC	end punctuation	<i>!?.</i>
SYM	SYM	symbol	
TO	PREP	to go, to him	
UH	INTJ	interjection	<i>Oh, Please</i>
VB	VERB	verb, base form	<i>be, jump</i>
VBD	VERB	verb, past	<i>was, jumped</i>
VBG	VERB	verb, gerund/participle	<i>jumping</i>
VBN	VERB	verb, past participle	<i>been, jumped</i>
VBZ	VERB	verb, pres, 3rd p. sing	<i>is, jumps</i>
VBP	VERB	verb, pres non-3rd p.	<i>am, jump</i>
WDT	ART	wh-determiner	<i>which, that</i>
WP	PRON	wh-pronoun	<i>what, whom</i>
WP\$	PRON	possessive wh-pronoun	<i>whose</i>
WRB	ADV	wh-abverb	<i>how, where</i>
:	SENT	general joiner	<i>:</i>
\$	SYM	currency (unsystematic)	<i>\$</i>
XX0	ADV	not, n't (not in PENN)	

Vorsichtig verwenden:

PUN	PUNC	punctuation not SENT	<i>;;,-</i>
LQUO	PUNC	opening quotes	
RQUO	PUNC	closing quotes	
BRL	PUUNC	left brackets	
BRR	PUNC	right brackets	
LS	PUNC	list item	

Know-your-corpus-Hinweis:

POS-Tags immer vorsichtig verwenden. Da sie automatisch von einem Tagger vergeben werden, sind sie oft nicht zuverlässig. Eine Kombination aus unterschiedlichen Strategien sollte immer angewandt werden.

In CQP kann man sich die POS-Tags und/oder Klassen zu einer Suche mit ausgeben lassen und auch exportieren. Dazu: **show +pos +class** vor der Abfrage eingeben (zum Abschalten: **show -pos**).

CLAWS-5 Tagset

B

(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</i> (adv.), <i>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]
TO0	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)

D

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					
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!