



Anleitung CQPWeb

mit Beispielanfragen für die Korpora →GerMov und KLG (bestehend aus Texten der →militanten gruppe und der →Revolutionären Zellen)

4. April 2010

Caveat I: Standardmäßig ist in *CQPWeb* die Abfragesprache *Simple query language* aktiviert. Die vorliegende Anleitung stellt Beispiele für die Syntax des *Corpus Query Processor (CQP)* bereit, daher ist eine Umstellung des *Query mode* auf diese Sprache vor jeder Abfrage notwendig (sofern das Browser-Fenster nicht geöffnet bleibt).

Caveat II: Das Kopieren und Einfügen der Abfragen aus diesem Dokument in *CQPWeb* führt bei einzelnen Zeichen (etwa beim Unterstrich oder bei Umlauten) zu Problemen. Eine Rohtextversion der Abfragen liegt →hier bereit.

Für jedes Token ist dessen Wortform (`word`), Lemma (`lemma_tree`) und Wortart (`pos_tree`) verfügbar. Diese Angaben werden als **positionelle Attribute** bezeichnet. Die Wortarten-Annotationen basieren auf dem Stuttgart-Tübingen-Tagset (STTS), wie es in <http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-1999.pdf> dargestellt ist.

Suche nach positionellen Attributen

Suche nach einzelnen Token

... mittels Angabe der Wortform

Beispiel: Suche nach der Wortform *Politik*

"Politik";

Diese Abfrage ist äquivalent zu [word = "Politik"]. (Die Leerzeichen vor und nach dem Gleichheitszeichen können weggelassen werden.) Das Semikolon am Ende einer Abfrage entspricht der regulären *CQP*-Syntax, muss in *CQPWeb* aber nicht zwingend gesetzt werden.

→Zur Korpusabfrage

Beispiel: Suche nach den Wortformen *du* und *Du*

"(D|d)u";

| ist der Disjunktionsoperator. Er steht für ein „inklusives“ *oder*, d. h. es wird sowohl nach *du* als auch nach *Du* gesucht. Mit der Klammerung wird verhindert, dass nach *D* einerseits und *du* andererseits gesucht wird.

→Zur Korpusabfrage

Beispiel: Suche nach den Wortformen *Haus* und *Hause*

"Hause?";

Das Spezialzeichen ? markiert Optionalität. Das Zeichen, auf das es folgt, darf kein- oder einmal vorkommen.

→Zur Korpusabfrage

Beispiel: Suche nach den Wortformen *ACH*, *ACHH*, *ACHHH* usw.

"ACH+";

Zeichen, die unmittelbar vor dem Spezialzeichen + stehen, müssen mindestens einmal auftreten.

→Zur Korpusabfrage

Beispiel: Suche nach den Wortformen *ACH*, *ACHH*, *ACHHH* usw.

"ACHH*";

Zeichen, die unmittelbar vor dem Spezialzeichen * stehen, dürfen keinmal oder beliebig oft auftreten.

→Zur Korpusabfrage

... mittels Angabe der Wortart

Beispiel: Suche nach den STTS-Tags *ADJA* und *ADJD*

[pos_tree = "ADJ."];

Der Punkt steht für ein beliebiges Zeichen. Die Suche nach *ADJA* und *ADJD* entspricht

einer Suche nach allen Adjektiven.
→Zur Korpusabfrage

Beispiel: Suche nach Token, die keine Adjektive sind

```
[pos_tree != "ADJ."];
```

!= ist der Negationsoperator.
→Zur Korpusabfrage

... mittels Angabe von Lemma und Wortart

Beispiel: Suche nach Verben mit dem Präfix *ver*

```
[(lemma_tree = "ver.+") & (pos_tree = "V.*")];
```

Die Kombination der Spezialzeichen . und + setzt voraus, dass ein beliebiges Zeichen (.) mindestens einmal (+) vorkommt. Mit dem Spezialzeichen & können Bedingungen verknüpft werden.
→Zur Korpusabfrage

Beispiel: Suche nach Verben mit dem Präfix *ver* in allen Formen außer dem Infinitiv

```
[(lemma_tree = "ver.+") & (word != lemma_tree)];
```

lemma_tree entspricht der Grundform eines Tokens; bei Verben ist dies der Infinitiv.
→Zur Korpusabfrage

Suche nach Sequenzen von Token

... mittels Angabe der Wortform

Beispiel: Suche nach der Wortformen-Folge *für die gute Sache*

```
"für" "die" "gute" "Sache";
```

→Zur Korpusabfrage

Beispiel: Suche nach *eine ... Sache* mit null bis zwei Token dazwischen

```
"eine" []{0,2} "Sache";
```

[] steht für ein beliebiges Token.
→Zur Korpusabfrage

Beispiel: Suche nach der Wortformen-Folge *wir alle* und *ihr alle*

```
("wir" "alle")|("ihr" "alle");
```

→Zur Korpusabfrage

... mittels Angabe von Wortform und Wortart

Beispiel: Suche nach der Wortformen-Folge *wir alle* und *ihr alle*

```
"ein" "neuer" [pos_tree = "NN"];
```

→Zur Korpusabfrage