

Korpuslinguistik mit Online-Ressourcen

Tutorium der Computerlinguistik-Sektion der DGfS

Stefan Evert, Stefanie Dipper, Heike Zinsmeister

23. Februar 2010

<http://wordspace.collocations.de/doku.php/dgfs2010>

1 Europarl – einfache Anfragen

In der ersten Übung untersuchen Sie das Europarl-Korpus (Debatten des europäischen Parlaments aus den Jahren 1996–2006) mit einem speziell für dieses Korpus entwickelten Webinterface. Sie können dabei die einfache Anfragesprache CEQL verwenden, die auch von einigen anderen Web-Interfaces unterstützt wird.

Geben Sie zunächst in Ihren Browser die Adresse des Europarl-Webinterfaces in Osnabrück ein. Auf der Startseite finden Sie in der Menüzeile unter dem Korpusnamen drei Links auf weiterführende Seiten. Wählen Sie hier die Option `simple query`. Auf der Anfrageseite zur *Simple Query* finden Sie einen Link auf eine Hilfeseite (Help page im Hauptmenü), auf der die Syntax der einfachen Anfragesprache erläutert wird.

- **URL:** <http://www.cogsci.uni-osnabrueck.de/~korpora/ws/CQPdemo/Europarl/>
▶ `simple query`
- **Hilfe zur einfachen Anfragesprache:**
<http://www.cogsci.uni-osnabrueck.de/~korpora/ws/CQPdemo/Europarl/help-simple.html>

1.1 Suche nach einem Einzelwort

Um die Anfragesprache und den Inhalt des Korpus kennenzulernen, suchen Sie zunächst nach einem einfachen Wort. Wählen Sie bei der Spracheinstellung `German` und als Sprachen für das Ausgabe-Display `EN, DE` (Englisch, Deutsch). Tippen Sie dann in das Anfragefenster `deutsch` und klicken Sie auf den Knopf `Run Query`. Welche Treffer erhalten Sie? Betrachten Sie sich den dritten Treffer genauer. Hier finden Sie den seltsamen Ausdruck *Eingeborenenvölker*. Lesen Sie den textuellen Kontext für diesen Treffer (Klick auf `context`). Wie wird *Eingeborenenvölker* auf Englisch übersetzt?

- **Einstellungen:** `lang=German`; `Display: EN, DE`; Rest wie voreingestellt
- **Anfrage:** `deutsch` (▶ `Run Query`) [R: 142]
- **Aufgabe:** Kontextanzeige für den 3. Treffer (▶ `context`)

Eigentlich ist es unerwartet, dass auch Treffer mit großgeschriebenem *Deutsch* gefunden werden. Um das zu verhindern, müssen Sie den Suchmodus auf wortgetreue Suche einstellen (`mode=literal`). Suchen Sie erneut und überprüfen Sie die Suchergebnisse.

- **Einstellung:** `mode=literal`
- **Erneute Anfrage:** `deutsch` (▶ `Run Query`) [R: 37]

1.2 Suche nach einer idiomatischen Wendung und ihrer Übersetzung

Der erste Treffer der letzten Suche weist eine interessante Wendung auf: *auf gut deutsch*. Wie wird diese Wendung übersetzt? Suchen Sie hierfür nach der gesamten Wendung. Wie viele Treffer erhalten Sie? Lassen Sie sich zusätzlich zum Deutschen und Englischen die Treffer in Französisch (`FR`) und Niederländisch (`NL`) anzeigen. Die Originalsprache des Beitrags ist oft in der Kopfzeile nach dem Sprechernamen angegeben, d.h. die Treffer dokumentieren sowohl Übersetzungen aus dem Deutschen als auch Übersetzungen *in* das Deutsche.

- **Einstellung:** lang=German; mode=case-folded; Display: EN, DE, FR, NL;
- **Anfrage:** auf gut deutsch (► Run Query) [R: 8]
- **Aufgabe:** Finden Sie die Übersetzungen für *auf gut deutsch* bzw. Quellen, die ins Deutsche mit *auf gut deutsch* übersetzt wurden.

1.3 Suche nach einer flektierten Wortart: attributives Adjektiv

Auch mit der Einstellung `mode=case-folded` sucht das Europarl-Webinterface nach einer spezifischen Wortform, vgl. die erste Anfrage in Abschnitt 1.1 nach `deutsch`. Ist man speziell an der Verwendung von *deutsch* als attributives Adjektiv interessiert, ist dies nicht hilfreich. Stattdessen verwenden Sie wieder die Einstellung `mode=literal` und erweitern das Suchmuster mit Platzhalteroperatoren (`*`, `+`) für die Flexionsendung des Adjektivs. Betrachten Sie jeweils die Ergebnisse. Lassen Sie sich dann die Treffer alphabetisch sortieren (Suche mit `sort order=ascending`). Was fällt Ihnen auf?

Bei der dritten Anfrageoption erhalten Sie 2890 Treffer. Lassen Sie sich hierfür auch die Häufigkeiten aller gefundenen Wortformen anzeigen (► `Frequencies`).

- **Einstellung:** mode=literal
- **Anfrage 1:** deutsch* (Endung optional) [R: 3046]
- **Aufgabe:** Sortieren Sie Wörter alphabetisch: `sort order=ascending`
- **Anfrage 2:** deutsch+ (Endung obligatorisch) [R: 3009]
- **Anfrage 3:** deutsche* [R: 2890]
- **Aufgabe:** Frequenzauflistung mit ► `Frequencies`
- **Zusatzaufgabe:** Sortieren Sie die Frequenzauflistung zuerst nach Häufigkeit, dann alphabetisch.

1.4 Kollokationen

Mit welchen Nomen kommt das attributive Adjektiv *deutsch* vor? Testen Sie zunächst die einfachen Suchanfragen und sichten Sie die Ergebnisse. Ein einzelnes `+` steht in der CEQL-Syntax für ein beliebiges Wort (warum funktioniert `*` hier nicht?). Bei der dritten Anfrage verlangen Sie zusätzlich, dass das auf *deutsch* folgende Wort ein Substantiv ist. Wortarten (*parts of speech*, POS) können mit der Notation `_{POS}` gesucht werden. Die einfachen POS-Tags, die Sie hier verwenden können, sind auf der Online-Hilfeseite erläutert.

- **Anfrage 1:** deutsche* *; `sort order=unsorted` [R: 2890]
- **Anfrage 2:** deutsche* +; `sort order=unsorted`
- **Variante:** deutsche* +; `sort order=ascending`
- **Anfrage 3:** deutsche* `_{N}` [R: 2640]
- **Aufgabe:** Frequenzen der Wortpaare sichten (► `Frequencies`)

1.5 Komposita

Welche Komposita mit dem Kopf *-gesetz* gibt es? Die beiden Anfragen ergeben sehr ähnliche Resultate. Wo unterscheiden Sie sich?

- **Anfrage 1:** *gesetz; mode=literal [R: 595]
- **Anfrage 2:** +gesetz; mode=case-folded

Ein Nachteil bei diesem Ansatz ist, dass nur die Grundformen der Komposita gefunden werden, nicht aber flektierte Formen wie *Grundgesetzes* oder *EU-Gesetze*. Sie können eine Lemma-Suche durchführen, indem Sie das Suchwort in geschweifte Klammern `{...}` einschließen. Lassen Sie sich anschließend die Häufigkeiten anzeigen (`sort order=unsorted`). Wie werden die Komposita übersetzt? Finden Sie lexikalisierte Übersetzungen, halb lexikalisierte, Spontanbildungen?

- **Anfrage:** {+gesetz}; mode=case-folded [R: 1369]

- **Aufgabe 1:** Häufigkeiten anzeigen lassen (`sort order=unsorted`; ► **Frequencies**)
- **Aufgabe 2:** Übersetzungen sichten (z.B. *Grundgesetz*, *Antiterrorgesetz*, *Religionsgesetz*) — dazu in der Frequenzliste auf das jeweilige Wort klicken
- **Aufgabe 3:** Alternative Sortierung (`sort order=reverse desc.`; ► **Frequencies**)
- **Aufgabe 4:** Übersetzungen sichten

1.6 Wortlisten

Suchen Sie nun nach Gesetzes-Komposita im *Word List*-Modus (Link im Hauptmenü oben). Wie wird *Grundgesetz* übersetzt?

- **Auswahl:** Word List
- **Einstellungen:** Word list for: German, attribute=lemma, sort=by frequency, literal, filter+=gesetz; ► **Make Word List**
- **Aufgabe:** Sichten der Übersetzungen (Klick auf T neben Wort)

1.7 Disjunktion in einfachen Anfragen

Gehen Sie wieder zurück in den *Simple Query*-Modus. Sie wollen wissen, ob im Europarl-Korpus häufiger von *Atomkraft* oder von *Kernkraft* die Rede ist. Hierzu verwenden Sie eine Disjunktion in der Anfrage.

- **Anfrage 1:** `[Kern,Atom]kraft` [R: 329]
- **Anfrage 2:** einschließlich Komposita mit `[Kern,Atom]kraft+` [R: 1220]
- **Aufgabe:** Häufigkeiten anzeigen lassen (`sort order=unsorted`)

1.8 Syntax: Folge von Adjektiven

Welche Folgen mit zwei unmittelbar aufeinander folgenden Adjektiven gibt es im Korpus? Zunächst suchen Sie über Wortartenfilter. Sichten Sie das Ergebnis. Die automatische Wortartenannotation verwechselt oft Adjektive und Adverbien, da sie im Deutschen sehr ähnlich im Satz verteilt sind.

Um die Genauigkeit (*precision*) zu erhöhen verwenden wir den alten Trick, die Flexionsendung direkt anzusprechen. Diesmal sollten Sie über einen Operator jeden beliebigen Stamm erlauben und die Flexionsendungen einzeln auflisten. Kombinieren Sie dieses Muster mit dem POS-Tag für Adjektive.

Da die Trefferzahl immer noch sehr groß ist, suchen Sie anschließend nach Sequenzen mit drei Adjektiven. Um die Anfrage übersichtlicher zu gestalten verwenden Sie einen numerischen Wiederholungsoperator: `_{A} _{A} _{A}` kann auch als `(_{A}){3}` geschrieben werden.

- **Anfrage 1:** `_{A} _{A}` [R: > 100 000]
- **Anfrage 2:** `*[e,er,es,em,en]_{A} *[e,er,es,em,en]_{A}` [R: > 100 000]
- **Anfrage 3:** `(*[e,er,es,em,en]_{A}){3}` [R: 2376]
Sichtung mit `sort=unsorted`
- **Anfrage 4:** `(*[e,er,es,em,en]_{A}){4}` [R:29]
- **Zusatzaufgabe:** Formulieren Sie die Anfrage über ein geeignetes STTS-Tag (vgl. Einführung), das nicht in `{...}` eingeschlossen wird. Müssen Sie auch jetzt noch die Adjektiv-Endungen angeben, um "saubere" Ergebnisse zu erhalten?

1.9 Wie es im Buche steht – Realisierung des Dativ -e

Zunächst verschaffen Sie sich einen ersten Eindruck und testen die spezielle Kombination von *in* + *Buch*.

- **1. Test:** `im Buche; mode=literal` [R: 2]
- **2. Gegenteil:** `im Buch` [R: 6]

Da das Korpus keine morphologische (Kasus-)Information enthält, müssen Sie indirekt Dativ einfordern: z.B. über eindeutige Artikel und Präpositionen.

- **Anfrage 3:** `[einem,dem,am,beim,im,vom,zum] (_{A})* *e_{N};` [R: 31 910]
`mode=case-folded; sort=unsorted`

Die Ergebnisse sehen nicht schlecht aus. Aber die Präzision ist nicht gut: Treffer 1 und 4 z.B. sind *false positives*. **Kommentar:** Bessere Ergebnisse wären zu erwarten, wenn man zusätzlich verlangen könnte, dass das zugehörige Lemma nicht auf *-e* endet. Hierzu benötigen wir eine mächtigere Anfragesprache, die auch einen Negationsoperator enthält: die CQP-Syntax.

Bevor Sie in den *CQP Mode* wechseln, scrollen Sie an das Seitenende. Dort wird Ihnen die Übersetzung Ihrer aktuellen Anfrage in CQP-Syntax angezeigt.

- **Aufgabe:** Kopieren Sie die CQP-Variante Ihrer Suchanfrage.

2 Europarl – CQPweb

Zunächst bleiben Sie im Europarl-Webinterface. Wechseln Sie nur in den CQP-Modus (durch Klick auf *CQP Mode* im Hauptmenü oben).

- **URL:** <http://www.cogsci.uni-osnabrueck.de/~korpora/ws/CQPdemo/Europarl/>
▶ CQP mode
- **Hilfe:** <http://www.cogsci.uni-osnabrueck.de/~korpora/ws/CQPdemo/Europarl/help-cqp.html>

2.1 Realisierung des Dativ -e – Anfragen in CQP-Syntax

Kopieren Sie Ihre letzte Anfrage aus Abschnitt 1.9 in das Suchfenster. Verstehen Sie die einzelnen Komponenten? Ergänzen Sie dann den negativen Filter und überprüfen Sie die Resultate. Was passiert, wenn man den “Dativfilter” weglässt?

- **Anfrage 1:** [R: 31 836]
`[word="(einem|dem|am|beim|im|vom|zum)%"c] ([pos="ADJ."])* [word=".*e" & pos="N.*"]`
- **Anfrage 2:** [R: 18 197]
`[word="(einem|dem|am|beim|im|vom|zum)%"c] ([pos="ADJ."])*`
`[word=".*e" & lemma!=".*e" & pos="N.*"]`
Einstellungen: `lang=German; deaktivieren: EN,FR,ES,IT,NL`
- **Aufgabe:** Häufigkeiten anschauen; ggf. `normalized` deaktivieren
- **1. Zusatzaufgabe:** Was passiert, wenn Sie den “Dativfilter” weglassen?
- **2. Zusatzaufgabe:** Ersetzen Sie den “Dativfilter” durch einen STTS-Wortartenfilter für Artikel und das Muster *-m* für die Wortendung [R: 18 185]
`[pos="ART|APPRART" & word=".*m"] [pos="ADJ."]* [word=".*e" & lemma!=".*e" & pos="N."]`
- **Variante:** eine kompaktere Formulierung ist `[pos=".*ART" & word=".*m"] ...`

Klicken Sie auf *Distribution*, um eine Aufteilung der Treffer nach Jahr und Originalsprache zu sehen. Es fällt dabei auf, dass die Dativformen mit *-e* (fast) ausschließlich in Texten vorkommen, die ins Deutsche übersetzt wurden. **Kommentar:** Der beobachtete Zeitabschnitt ist zu klein, um der Aufteilung nach Jahren aussagekräftige Tendenzen zu entnehmen. In Abschnitt 3 wechseln Sie auf die DWDS-Korpora, die eine Zeitspanne von 100 Jahren umfassen und eine bessere diachrone Darstellung erlauben.

2.2 Gegenseitiges Vertrauen – Wechsel zu CQPweb

Welche Adjektive treten mit dem Nomen *Vertrauen* auf? Wechseln Sie zu dem CQPweb-Interface in Osnabrück. Falls dabei eine Sicherheitswarnung erscheint, bestätigen Sie bitte die Ausnahme und akzeptieren das Sicherheitszertifikat des Servers.

- **URL:** <https://quickie.ikw.uos.de/cqpweb/>
- Zugangsinformationen (Login + Passwort) werden im Tutorium ausgeteilt

Wichtige Anmerkung: Wenn man auf *New Query* klickt, wird der Such-Modus auf *Simple Query* zurückgesetzt. Daher ist es oft bequemer, den “Zurück”-Knopf des Web-Browsers zu verwenden. (Bei *Simple Query* können Sie die gleiche CEQL-Syntax wie im Europarl-Webinterface verwenden. Wir werden hier jedoch ausschließlich mit CQP-Syntax arbeiten.)

Suchen Sie zunächst nach dem Nomen *Vertrauen*. Nutzen sie dann die Kollokations-Suche von CQPweb mit verschiedenen Filteroptionen. Zuerst untersuchen Sie den linken Kontext des Nomens *Vertrauen* (Fenster: 3 Token), in dem Sie relevante Adjektive erwarten. Anschließend untersuchen Sie den rechten Kontext von *Vertrauen*. Hier erwarten Sie typischerweise Nomen, die gehäuft mit *Vertrauen* verwendet werden.

- **Anfrage:** [lemma="Vertrauen"]
Einstellung: Query mode = CQP syntax [R: 5451]
- **Schritt 1:** ► Collocations
Include: Lemma, POS; ► Create collocation database
Collocation based on: Lemma;
Collocation window from: 3 to the left;
Collocation window to: 1 to the left;
► [Submit changed parameters] Go!
- **Schritt 2:** zusätzlich Filter and/or tag: ADJA aktivieren; ► Go!
- **1. Resultate:** Klick auf eines der Adjektive – zeigt detaillierte Statistiken zu dieser Kollokation
- **2. Resultate:** Klick auf Zahl in Spalte *Observed collocate frequency* – zeigt alle Belege

Untersuchung des rechten Kontexts:

- **Schritt 3:** “Zurück”-Knopf (um zur Kollokationsliste zurückzukehren)
Collocation based on: Lemma;
Collocation window from: 1 to the Right;
Collocation window to: 3 to the Right;
► [Submit changed parameters] Go!
- **Schritt 4:** zusätzlich Filter and/or tag: NN aktivieren
- **Aufgabe:** Details für *Verbraucher* anschauen. In welchem Abstand von *Vertrauen* tritt das Wort am häufigsten auf?
- **Zusatzaufgabe:** Suchen Sie nach Kollokationen mit dem Wort *deutsch* (Filter: Adjektive).

2.3 Ohne *Dirigent(en)* – Können “nackte” Substantive flektiert werden?

In diesem Abschnitt untersuchen wir die Gültigkeit der Regel, dass Substantive nicht flektiert werden, wenn sie ohne Artikel oder Adjektiv verwendet werden. Um unnötige *false positives* zu vermeiden, wählen wir den eindeutig erkennbaren Kontext einer Präpositionalphrase. Die erste Anfrage resultiert in viele Pluralnomen, daher ergänzen wir einen Filter, der typische Pluralendungen aussortiert.

- **Anfrage 1:** [pos="APPR"] [pos="NN" & lemma != word] [R: 233 783]
- **Anfrage 2:** [pos="APPR"] [pos="NN" & lemma!=word & word!=".*(n|e|er)"] [R: 16 879]

Die Ergebnisse der letzten Anfrage enthalten immer noch alle dem POS-Tagger unbekanntem Nomen, die in Europarl als <unknown> lemmatisiert sind. Wir schließen diese nun explizit aus (das Flag %c ist wegen Inkonsistenzen in der Europarl-Annotierung erforderlich):

- **Anfrage 2:** [pos="APPR"] [pos="NN" & lemma!=word & lemma!="<unknown>%c & word!=".*(n|e|er)"] [R: 9 390]
- **Aufgabe:** ► Frequency breakdown: Sichten Sie die ersten 80 Fälle

3 DWDS-Suche: ein diachrone Perspektive

Wechseln Sie auf die Seite des Digitalen Wörterbuchs der Deutschen Sprache.

- URL: <http://beta.dwds.de/>
- Hilfe: <http://beta.dwds.de/help/>

3.1 Vorbemerkungen

- Im DWDS wird automatisch nach Lemmata gesucht; eine Wortform muss extra mit @ markiert werden!
- Sobald man *Wildcards* (Platzhalter) nutzt, wird jedoch automatisch nach Wortformen gesucht. Z.B. müsste *blau eigentlich logischerweise @*blau heißen.
- Man kann nicht mit Wildcards nach Lemmata suchen (d.h. unsere Flexionsanfragen werden komplizierter).
- Man kann in jedem Wort nur eine Wildcard verwenden.
- Disjunktion gibt es nur "satzweise", also A || B heißt: A oder B kommen (irgendwo) im Satz vor. Und z.B. "A B (C || D)" kann nur als "A B C" || "A B D" geschrieben werden.
- Wenn man nach Sequenzen sucht (d.h. mehrere Wörter, die unmittelbar aufeinander folgen), muss die Anfrage in doppelte Anführungszeichen eingeschlossen werden.
- Sonstiges:
 - DWDS nutzt auch das STTS-Tagset
 - DWDS bietet verschiedene Korpora (und Wörterbücher) für die Suche an

3.2 Einfache Wortsuche

Testen Sie die Anfragen auf verschiedenen Korpora (Kernkorpus, Zeitungskorpora)

- **Anfrage 1:** deutsch
- **Anfrage 2:** @deutsch
- **Anfrage 3:** "@auf @gut @deutsch"

3.3 Attributives Adjektiv und Kollokationen mit Substantiven

Suchen sie jetzt nur noch auf dem Kernkorpus, da es balanciert ist (Jahre 1900–2000, Genre). Wie suchen Sie nach Belegen für *deutsch* als attributives Adjektiv? Mit welchen Substantiven tritt *deutsch* im Kernkorpus auf?

- **Anfrage 1:** deutsch with \$p=ADJA
- **Anfrage 2:** "deutsch with \$p=ADJA #0 \$p=NN"

3.3.1 Komposita

- **Anfrage 1:** *gesetz
- **Aufgabe 1:** Testen Sie verschiedenen Sortierungen unter *Darstellungsoptionen* (nach Datum, Dokumentlänge, ...)
- **Aufgabe 2:** Testen Sie die *Wortverlaufsstatistik*, die Ihnen die Trefferzahl pro Genre und Jahrzehnt anzeigt.

3.4 Syntax: Sequenz von Adjektiven

- **Anfrage:** "\$p=ADJA \$p=ADJA \$p=ADJA \$p=ADJA"

3.5 Realisierung des Dativ *-e* – im DWDS

Da man im DWDS keine Platzhalter auf Lemmata anwenden kann und auch keine Disjunktion auf Token-Ebene möglich ist, wird die Anfrage etwas komplexer als in CQP.

- **Anfrage:** "@einem #1 @Buche" || "@dem #1 @Buche" || "@am #1 @Buche"
|| "@beim #1 @Buche" || "@im #1 @Buche" || "@vom #1 @Buche" || "@zum #1 @Buche"
- **Aufgabe:** Die Wortverlaufsstatistik anschauen
- **Aufgabe:** Dasselbe für andere Nomen testen.

3.6 *Gegenseitiges Vertrauen* – Kollokationen

- **Anfrage:** Vertrauen
- **Aufgabe:** In der Statistik-Sicht das Fenster DWDS-Wortprofil anschauen: Wortwolke; Tabellenansicht

4 Das STTS-Tagset

| Tag | Beschreibung | Beispiele |
|---------|---|--|
| ADJA | attributives Adjektiv | <i>[das] große [Haus]</i> |
| ADJD | adverbiales oder prädikatives Adjektiv | <i>[er fährt] schnell, [er ist] schnell</i> |
| ADV | Adverb | <i>schon, bald, doch</i> |
| APPR | Präposition; Zirkumposition links | <i>in [der Stadt], ohne [mich]</i> |
| APPRART | Präposition mit Artikel | <i>im [Haus], zur [Sache]</i> |
| APPO | Postposition | <i>[ihm] zufolge, [der Sache] wegen</i> |
| APZR | Zirkumposition rechts | <i>[von jetzt] an</i> |
| ART | bestimmter oder unbestimmter Artikel | <i>der, die, das, ein, eine, ...</i> |
| CARD | Kardinalzahl (Ordinalzahlen sind als ADJA getaggt) | <i>zwei [Männer], [im Jahre] 1994</i> |
| FM | Fremdsprachliches Material | <i>[Er hat das mit "A big fish" übersetzt]</i> |
| ITJ | Interjektion | <i>mhm, ach, tja</i> |
| KOUI | unterordnende Konjunktion mit <i>zu</i> und Infinitiv | <i>um [zu leben], anstatt [zu fragen]</i> |
| KOUS | unterordnende Konjunktion mit Satz | <i>weil, daß, damit, wenn, ob</i> |
| KON | nebenordnende Konjunktion | <i>und, oder, aber</i> |
| KOKOM | Vergleichskonjunktion | <i>als, wie</i> |
| NN | normales Nomen | <i>Tisch, Herr, [das] Reisen</i> |
| NE | Eigennamen | <i>Hans, Hamburg, HSV</i> |
| PDS | substituierendes Demonstrativpronomen | <i>dieser, jener</i> |
| PDAT | attribuierendes Demonstrativpronomen | <i>jener [Mensch]</i> |
| PIS | substituierendes Indefinitpronomen | <i>keiner, viele, man, niemand</i> |
| PIAT | attribuierendes Indefinitpronomen ohne Determiner | <i>kein [Mensch], irgendein [Glas]</i> |
| PIDAT | attribuierendes Indefinitpronomen mit Determiner | <i>[ein] wenig [Wasser], [die] beiden [Brüder]</i> |
| PPER | irreflexives Personalpronomen | <i>ich, er, ihm, mich, dir</i> |
| PPOSS | substituierendes Possessivpronomen | <i>meins, deiner</i> |
| PPOSAT | attribuierendes Possessivpronomen | <i>mein [Buch], deine [Mutter]</i> |
| PRELS | substituierendes Relativpronomen | <i>[der Hund ,] der</i> |
| PRELAT | attribuierendes Relativpronomen | <i>[der Mann ,] dessen [Hund]</i> |
| PRF | reflexives Personalpronomen | <i>sich, einander, dich, mir</i> |
| PWS | substituierendes Interrogativpronomen | <i>wer, was</i> |
| PWAT | attribuierendes Interrogativpronomen | <i>welche [Farbe], wessen [Hut]</i> |
| PWAV | adverbiales Interrogativ- oder Relativpronomen | <i>warum, wo, wann, worüber, wobei</i> |
| PAV | Pronominaladverb | <i>dafür, dabei, deswegen, trotzdem</i> |
| PTKZU | <i>zu</i> vor Infinitiv | <i>zu [gehen]</i> |
| PTKNEG | Negationspartikel | <i>nicht</i> |
| PTKVZ | abgetrennter Verbzusatz | <i>[er kommt] an, [er fährt] rad</i> |
| PTKANT | Antwortpartikel | <i>ja, nein, danke, bitte</i> |
| PTKA | Partikel bei Adjektiv oder Adverb | <i>am [schönsten], zu [schnell]</i> |
| TRUNC | Kompositions-Erstglied | <i>An- [und Abreise]</i> |
| VVFIN | finites Verb, voll | <i>[du] gehst, [wir] kommen [an]</i> |
| VVIMP | Imperativ, voll | <i>komm [!]</i> |
| VVINFIN | Infinitiv, voll | <i>gehen, ankommen</i> |
| VVIZU | Infinitiv mit <i>zu</i> , voll | <i>anzukommen, loszulassen</i> |
| VVPP | Partizip Perfekt, voll | <i>gegangen, angekommen</i> |
| VAFIN | finites Verb, aux | <i>[du] bist, [wir] werden</i> |
| VAIMP | Imperativ, aux | <i>sei [ruhig !]</i> |
| VAINFIN | Infinitiv, aux | <i>werden, sein</i> |
| VAPP | Partizip Perfekt, aux | <i>gewesen</i> |
| VMFIN | finites Verb, modal | <i>dürfen</i> |
| VMINFIN | Infinitiv, modal | <i>wollen</i> |
| VMPP | Partizip Perfekt, modal | <i>gekonnt, [er hat gehen] können</i> |
| XY | Nichtwort, Sonderzeichen enthaltend | <i>3:7, H2O, D2XW3</i> |
| \$, | Komma | <i>,</i> |
| \$. | Satzbeendende Interpunktion | <i>. ? ! ; :</i> |
| \$(| sonstige Satzzeichen; satzintern | <i>- [] ()</i> |