

Was ist COSMAS II?

Eine am IDS konzipierte Datenbank (**Corpus Search, Management and Analysis System**) für die korpusbasierte Erforschung der Sprache

- in umfangreichen Korpora (über 13 Mrd. Wortformen, zur Verfügung gestellt vom Projekt DEReko);
- in linguistisch und strukturell annotierten Korpora; z.B. Wortklassen (über 1,7 Mrd. Nomen), Überschriften etc.;
- in benutzerdefinierten Korpusquerschnitten (anhand von bis zu acht bibliografischen Kriterien);
- in unterschiedlichen Sprachkorpora mit eigenen Tagsets, die in einem grafischen Assistenten eingebettet sind;
- mittels zahlreicher Such-, Abstands- und Bereichsoperatoren, die einfache bis komplexe Sachverhalte bzw. grammatische Muster formulieren können.

Die Ergebnisse werden

- anhand von bibliografischen Kriterien zusammengefasst und sortiert;
- mittels Häufigkeitsmaßen in Bezug auf ihre Verteilung bewertet;
- mittels einer Kookkurrenzanalyse analysiert, sortiert und tabellarisch dargestellt;
- als KWIC und Belege sortiert, analysiert und präsentiert;
- (auf Wunsch) mittels eines Zufallsgenerators auf eine repräsentative überschaubare Menge reduziert.

»In den Weiten der Korpora«

Die Größe der IDS-Korpora erlaubt

- seltene Fälle eines Wortgebrauchs ausfindig zu machen;
- Textbeispiele für Wörter wie *präsumptuos*, die in einem Wörterbuch vielleicht unauffindbar sind, zu erhalten;
- mittels der Kookkurrenzanalyse starke Wortverbindungen, geläufige Assoziationen und syntaktische Muster des Gebrauchs von Wörtern zu identifizieren;
- Schwankungen im Gebrauch eines Wortes entlang verschiedener Dimensionen (Zeit, Quellen, Textsorte etc.) zu erfassen;
- neue Wortschöpfungen oder neue Bedeutungen existierender Wörter (Neologismen) dank der kontinuierlichen Aktualisierung mit neueren Texten aufzuspüren;
- Recherchen auf spezielle Korpusquerschnitte einzugrenzen (Texte nach dem 11. Sept. 2001, Belletristik der 90er Jahre etc.) oder Ergebnisse aus verschiedenen Teilkorpora miteinander zu vergleichen.

Fragestellung

erweiterter Infinitiv

Muster

V + , + ... + zu + V_{inf}

Suchanfrage

```
(MORPH(V) /w0 ,) /+w3:6,s0  
(zu /+w1:1,s0 MORPH(V INF))
```

Die Quellen

Die Zusammensetzung der vordefinierten und dynamisch erstellten Benutzerkorpora kann, nach verschiedenen bibliografischen Metadaten (*Erscheinungszeit* und *-land*, *Quelle*, *Textsorte* und automatisch ermittelte *Thematik*) aufgefächert, eingesehen werden.

Treffer bzw. Belege werden mit einem ausführlichen Quellennachweis angezeigt und exportiert.

Zugang zu COSMAS II

- kostenlos für wissenschaftliche und nicht-kommerzielle Zwecke;
- für die interessierte Öffentlichkeit.

Zur Web-Version:

<http://www.ids-mannheim.de/cosmas2-web/>

Die in COSMAS II ausgeführten Recherchen (9,3 Mio. Zugriffe seit 2003) stammen

- zu ca. 80% von externen Nutzern;
- davon zu ca. 60% aus deutschsprachigen Ländern;
- von Linguisten, von Übersetzern und Sprachlehrern auf der Suche nach Beispielen lebendigen Sprachgebrauchs;
- von interdisziplinären Wissenschaften wie Psychologie, Sprachtherapie, Medienwissenschaften, Statistik etc.

Hinzu kommen jährlich 1 Mio. automatisierte Zugriffe zur Unterstützung rechercheintensiver Projektvorhaben.