

NJII_1369 Korpuslinguistik – 5. Sitzung – IdS Mannheim – 4.11.2010

1. Ankündigungen

- 6. November: Korpuslinguistik-Konferenz an der PedF. Programm auf Papier.
- 4.-9.11. – Queer-Film-Festival mezipatra.
<http://www.mezipatra.cz/2010/program/>
Zwei deutsche Beiträge:
7.11., 18 Uhr – Saša (Kino Scala)
7.11., 18 Uhr – New York Memories (Art)
- 11.11. 2010 ab 19 Uhr – 17. Lange Nacht der kurzen Texte
Feste feiern. Eine Lange Nacht zum 20. Geburtstag der Österreich-Bibliothek.
Anmeldung: lange-nacht@web.de

2. Die Korpora des IdS Mannheim

- gern auch (fälschlich) als IdS-Korpus bezeichnet oder als Cosmas-Korpus.
- Richtig ist: Das IdS hat eine Sammlung von sehr vielen Korpora.
- Viele davon sind linguistisch annotiert.
- Bei keinem der Korpora handelt es sich um ein ausgewogenes Korpus, das für die dt. Sprache repräsentativ wäre. Meistens Zeitungskorpora.
- Übersicht über die unterschiedlichen Korpora (was für Texte stecken in welchem Korpus): <http://www.ids-mannheim.de/cosmas2/uebersicht.html> - dort auf **Korpora** klicken
- auch interessant: **Dokumente und Texte**. Für ein virtuelles Korpus (maßgeschneidert für Ihre Ziele) können Sie einzelne Dokumente zusammenstellen.
- Achten Sie bei der Korpusauswahl darauf, welche Texte Sie wollen (Textsorte – auch Wortzahl der versch. Dokumente, Zeit).

3. Recherche mit COSMAS II, der Suchmaschine für die Korpora des IdS

Startseite: www.ids-mannheim.de

- allgemeine Informationen
- dann rechts: Cosmas II: <http://www.ids-mannheim.de/cosmas2/uebersicht.html>
- Es gibt mehrere Möglichkeiten. Die beste ist Cosmas_{Win} – aber die können wir nicht nutzen, weil man dazu einen Client runterladen muss, wofür man Administratorenrechte braucht. → Machen Sie das zuhause.
- Registrieren muss man sich in jedem Fall (kostenlos).
- Cosmas_{Web}: <http://www.ids-mannheim.de/cosmas2/web-app/> (= Info)

- kurz zeigen: Online-Hilfe → Einführung → Quicktour (wichtig wegen Terminologie)
- dann Applikation starten, z.B. links über Navigation.
- dann oben einen Reiter weiterklicken: Anmeldung. Links: Login
- nach erfolgreicher Anmeldung weiter zum nächsten Reiter: Recherche

- Archiv: W – Archiv der geschriebenen Sprache
- Korpora auswählen: z.B. alles Österreichische. Dazu auf Korpusverwaltung (links) klicken, dann rechts unten auf „definieren“. Dann kann man einzelne Korpora

auswählen, dann nach rechts verschieben. Denken Sie immer an die vier Kriterien bei der Korpusauswahl (→ Repräsentativität, Größe, Zeit, Inhalt)

- Wenn man seine Auswahl getroffen hat (im rechten Fenster): Name vergeben, „übernehmen“ klicken.
- Hilfedatei zur Formulierung der Suchanfrage: <http://www.ids-mannheim.de/cosmas2/web-app/hilfe/seite/suchanfrage/beispiele.html>
- Einzelne Operatoren erklären:

Platzhalteroperatoren

- * beliebig viele Zeichen (irgendwelche)
- ? genau ein beliebiges Zeichen
- ?? genau zwei beliebige Zeichen (können unterschiedlich sein)
- + null oder ein Zeichen
- \? Backslash vor dem Platzhalteroperator macht diesen zu einem normalen Zeichen (wenn man z.B. nach einem Fragezeichen suchen möchte)

Verknüpfungsoperatoren

- und sucht beide Wörter im selben Text
- oder sucht eins der beiden Wörter im selben Text
- nicht sucht nur das linke Wort und nur in Texten, in denen das rechte nicht auftritt
- /+w1 das rechts folgende Wort = direkt benachbarte Wörter in genau dieser Reihenfolge

Grundformoperator

- & Sucht das ganze Lemma. Vorsicht: sehr weit gefasster Lemmabegriff. Umfasst Flexionsformen, aber auch Derivate und Komposita!

3.1 Suche nach einem bestimmten Wort

- in den Kasten Suchanfrage eingeben: grün
- 4 Wortformen → anzeigen
- Auswahl treffen (Häkchen)
- Wenn man auf „Ergebnisse“ klickt, sieht man, wie viele Treffer in welchen Korpora (brauchen wir nicht) = Korpusansicht
- Jahrzehntansicht/Jahresansicht ist NICHT aussagekräftig, weil es keine vergleichbaren Korpora für jedes Jahrzehnt gibt. Nicht vergleichbar mit den Daten des DWDS!
- Gesamt-KWIC: alle Treffer, Reihenfolge wählbar (links): unsortiert, alphabetisch, chronologisch
- Es wird nicht lemmatisiert.
- Kookkurrenzanalyse: hier Einstellungen ändern: 10 li, 10 re, nur 1 Satz
Lemmatisierung verwenden
→ *Zweig* ist nicht dabei, weil nur die Kookkurrenzen lemmatisiert werden, nicht aber das Suchwort „grün“.
- Exportieren (wichtig für Dokumentation, am besten als CD der Arbeit beilegen)

3.2 Suche nach einem Lemma

- in den Kasten Suchanfrage eingeben: &grün
- davon kann man wieder Kookkurrenzen ermitteln, dann ist *Zweig* auch dabei.

3.3 Kollokation (auf (k)einen grünen Zweig kommen)

- in den Kasten Suchanfrage eingeben: grün?? Zweig
- Kookkurrenzanalyse zeigt, dass das Phrasem auch modifiziert wird, z.B.
finde zu keinem grünen Zweig
nicht nur *auf keinen grünen Zweig kommen*, sondern auch positiv
auf einen grünen Zweig

3.4 Bestimmte Wortart (Adjektive auf –bar)

- nach Wortarten (oder komplexeren syntaktischen Phänomenen) kann man nur im Archiv „TAGGED“ suchen ☹
- Aber auch hier kann man ein eigenes, virtuelles Korpus zusammenstellen (machen wir jetzt nicht).
- Suche nach Adjektiven auf –bar in allen Flexionsformen:
&-bar MORPH(ADJ)
Lemmasuche, Endung –bar, nur Adjektive (und deren Derivate und Komposita)
- anderes Tagset als bei DWDS, aber Morph-Assistent, der hilft

3.5 Wortliste (DeReWo)

- verfügbar unter <http://www.ids-mannheim.de/kl/projekte/methoden/derewo.html>
- angeblich kann man das auch aus virtuellen Korpora erzeugen (und dann Unterschiede feststellen). Wie – keine Ahnung.

4. Projektgruppen

- Gruppen bilden, über Ideen einig werden
- eintragen in die Liste

5. HA

- Lesen zur Vorbereitung der nächsten Woche (Umgang mit Zahlen): Scherer 32-40
- außerdem arbeiten wir nächste Woche an den Projekten (d.h. SIE)