



Parallele Textcorpora

Prof. Dr. Dmitrij Dobrovolskij

AAC / ÖAW

Texts & Files SS 2007

Universität Wien



Alignment

- 4 Arten von Alignment:
 1. auf der Textebene
 2. auf der Absatzebene
 3. auf der Satzebene
 4. auf der Wortebene
- Möglichkeiten und Grenzen der Automatisierung



Alignment auf der Absatzebene

- Probleme:
 - Die resultierenden Textausschnitte sind oft zu lang.
 - Die Grenzen zwischen den Absätzen stimmen nicht immer überein.



Alignment auf der Satzebene

- Probleme:
 - “Chunking” notwendig
 - Die Grenzen zwischen den Sätzen stimmen oft nicht überein.