

„Schnittstellen und Konkordanzen“
5. Arbeitstreffen deutschsprachiger Akademie-Wörterbücher
Wien, 8.–10. Juni 2006

WORKSHOPS

WORKSHOP 1: *Vernetzte Strukturen: Wörterbücher und Corpora*

Veranstalter: AAC – Austrian Academy Corpus
Leitung: Hanno Biber, Stephen Miller
Referate: Peter Ďurčo (Univerzita Sv. Cyrila a Metoda v Trnave, Slowakei)
 Truus Kruyt (Instituut voor Nederlandse Lexicologie, Leiden, Niederlande)

Im Workshop werden Möglichkeiten erörtert, vorhandene Daten aus den Bereichen Lexikographie und Corpusforschung sinnvoll zu vernetzen, um künftig komplex strukturierte Datenbestände zu sichern und lexikographisch erzeugtes Wissen für viele Interessierte möglichst nutzerfreundlich verfügbar zu machen. Die Referate sollen – bezogen auf die Situation der deutschen Sprache – durch Vergleich und Außenperspektive konstruktive Ansätze liefern und Diskussionen zu folgenden Fragen im ersten Teil des Workshops provozieren:

1. Welche Daten lassen sich aus den bereits existierenden digitalen Lexika, digitalen Grammatiken, digitalen Archiven und speziellen Korpora effizient für die Verbesserung der lexikographischen Arbeiten nutzen (Bearbeitung, Auswertung und Export der Daten)?
2. Wie zuverlässig und repräsentativ sind diese Daten?
3. Welche Methoden der Datenerhebung in technologisch und qualitativ sehr heterogenen Quellen sind sinnvoll anwendbar?
4. Was ist zu tun mit nichtannotierten digitalisierten Daten? Eignen sich auch solche „rohe“ Daten für die Lexikographie bzw. wie werden solche Datenbestände selbst zum Objekt der lexikographischen Bearbeitung?
5. Welche Aussagekraft und welchen praktischen Nutzen für die Lexikographie haben die quantitativen Methoden, wie z. B. Frequenzanalysen, Distributionsanalysen, Kollokations- und Kookkurrenzanalysen, word-sketches?
6. Wie effektiv und zuverlässig sind die in den Corpusbrowsern implementierten formalen Sprachen bei der Suche nach Entitäten der natürlichen Sprache?
7. Lassen sich die Suchprogramme, basiert auf der Logik der formalen Algebra (vgl. z. B. *regular expressions*), effizient zum Modellieren sowohl von einfachen als auch von komplexen Suchmustern real einsetzen?
8. Wie kann man das aus dem Technologiebereich und aus der Lexikographie erzeugte Wissen in der Zukunft zur Entwicklung einer Plattform für „lexicographic stations“ nutzen?

Im zweiten Teil des Workshops wird die am Institut für Niederländische Lexikologie in Leiden in Arbeit befindliche Plattform: "Geïntegreerde Taalbank (GTB)" [The Integrated Language Database of 8th-21st-Century Dutch] von Truus KRUYT präsentiert und diskutiert:

www.inl.nl/images/stories/taalbank/publicaties/lrec2004_kruyt.pdf
www.inl.nl/index.php?option=com_content&task=view&id=46&Itemid=80&lang=nl

WORKSHOP 2: *Grenze der Vernetzung: ß̃ñđę Zeichen ?*

Veranstalter: I DINAMLEX – Institut für Österreichische Dialekt- und Namenlexika

Leitung: Eveline Wandl-Vogt, Hubert Bergmann

Impulsreferat: Marc Wilhelm Küster (FH Worms, University of Applied Sciences, Fachbereich Informatik / Telekommunikation, Mitglied der W2 und d. DIN N129-1 *Kodierte Zeichensätze*; <http://people.fh-worms.de/~kuester>)

Titel: *Das Universal Character Set (UCS) und seine Bedeutung für wissenschaftliche Editionen im Zeitalter der EDV*

Inhalt: Das Universal Character Set (UCS) und Unicode / ISO IEC 10646, Einführung in Problemstellung, generelle Überlegungen, Bedeutung, Implementierungen; das UCS und das WWW; der organisatorische Rahmen und das Prozedere bei der Standardisierung neuer Schriftzeichen

Sonderzeichen stellen bei der Digitalisierung oder elektronischen Edition wissenschaftlicher Werke häufig einen Knackpunkt dar.

Ziele des Workshops:

- Information über die Problematik der Sonderzeichen im digitalen Umfeld;
- Zusammenstellung einer konkreten (Wunsch-)Liste von Sonderzeichen, die die Workshopteilnehmer(innen) gemeinsam erarbeiten und die an die entsprechenden Normierungsgremien (ISO/IEC JTC1/SC2/WG2 und das Unicode Consortium) weitergegeben werden soll;
- bei Interesse: Bildung einer Arbeitsgruppe für Sonderzeichen (um ein gemeinsam definiertes Projekt umzusetzen, z.B. Weiterentwicklung einer entsprechenden Font).

Vorbereitung durch die Workshopteilnehmer(innen): Erstellung einer (digitale) Liste mit Sonderzeichen, im Sinne von *besonderen* Zeichen, die für digitale Umsetzung problematisch erscheinen bzw. die für die Publikation des/der jeweils edierten Werke/s von Relevanz sind.