

3. Arbeitstreffen des wissenschaftlichen Netzwerks  
„Internetlexikografie“ (gefördert durch die Deutsche Forschungsgemeinschaft)

3<sup>rd</sup> workshop of the academic network on “Internet Lexicography”  
(supported by the Deutsche Forschungsgemeinschaft)

## **Thema „Automatische Gewinnung von lexikografischen Angaben“**

3. – 4. Mai 2012

Europäische Akademie Bozen (EURAC) (Drususallee 1, 39100 Bozen)

---

Die Verfügbarkeit großer elektronischer Textkorpora hat die Lexikografie generell stark verändert. Ein- und mehrsprachige Wörterbücher werden zunehmend korpusgestützt erarbeitet, und das unabhängig vom jeweiligen Publikationsmedium. Zugleich haben Korpuslinguisten reiche Korpusrecherche- und -analysetools entwickelt, während Computerlexikografen und -linguisten z. B. zur automatischen Lesartendisambiguierung oder zur Entwicklung lexikalisch-semantischer Ressourcen (z. B. Ontologien) beigetragen haben. Indem Lexikografen solche Tools anwenden, hat sich die inhaltliche Qualität der Wörterbücher verbessert, allerdings wurde bei gedruckten Wörterbüchern das Angabeninventar als solches wenig verändert.

Da Internetwörterbücher keinen Platzbeschränkungen unterliegen, können hier aber sehr detaillierte und ganz neue Arten von lexikografischen Angaben erscheinen. Welche dies sein können (z. B. Frequenzangaben, Kollokationen, feste Wortverbindungen) und wie lexikografisch erarbeitete und automatisch erstellte Angaben in Internetwörterbüchern kombiniert werden können, wird anhand der Vorstellung entsprechender Tools und Wörterbücher bei dem Arbeitstreffen deutlich werden. Dabei sind auch Antworten auf die Frage der direkten Rückbindung an das zugrunde liegende Online-Korpus wie der Nutzbarmachung von Wortnetzen in Internetwörterbüchern zu suchen.

## **Topic “Automatic extraction of lexicographical information”**

3 – 4 May 2012

European Academy Bolzano/Bozen (EURAC) (Viale Druso/Drususallee 1, 39100 Bozen/Bolzano)

---

The availability of large electronic corpora has changed lexicography dramatically (Teubert 1999). Monolingual and multilingual dictionaries are increasingly corpus-based, and this is happening irrespective of publication medium. At the same time, corpus linguists have developed effective tools for searching and analysing corpora, while computational lexicographers and linguists have

contributed to, for example, automatic disambiguation and the development of lexical semantic resources (e.g. ontologies). By using such tools, lexicographers have improved the quality of the content of dictionaries, but the inventory of information in printed dictionaries is little changed.

Since online dictionaries are not subject to space restrictions, they can contain very detailed and completely new kinds of lexicographical information (Haß 2005, Klosa 2010). What information this could be (e.g. frequency, collocations, set phrases) and how lexicographically acquired and automatically compiled information in online dictionaries can be combined, will become clear at the workshop through the introduction of appropriate tools and dictionaries. Thereby, the aim will also be to answer the question of direct connection to the underlying online corpus, as well as the utilisation of word-nets in online dictionaries.

---

**Mittwoch, 2. Mai 2012**

**Wednesday, 2 May 2012**

---

19:00

**Warm-up**

Restaurant „Walthers“, Waltherplatz 6, Bozen

**Donnerstag, 3. Mai 2012** (Seminar 2-3) (öffentlich)

**Thursday, 3 May 2012** (Seminar 2-3) (public)

---

08:45-09:00      Ankunft / Arrival

09:00-09:45      **Begrüßung und Präsentation EURAC & Institut für Fachkommunikation und Mehrsprachigkeit / Welcome and presentation of EURAC & the Institute for Specialised Communication and Multilingualism**

*Andrea Abel* (Europäische Akademie Bozen)

09:45-10:30      **You can tell a word by the (visual) company it keeps: extracting lexico-semantic information from text and images**

*Marco Baroni* (Università degli Studi di Trento - Center for Mind/Brain Sciences (CIMeC))

10:30-11:00      Kaffeepause / Coffee break

11:00-11:45      **Do we need lexicographers? Prospects for automatic lexicography**

*Adam Kilgarriff* (Lexical Computing Ltd Brighton)

11:45-12:30      **From DWDS-corpora to a German Word Profile - methodological problems and solutions**

*Jörg Didakowski, Alexander Geyken* (Berlin-Brandenburgische Akademie der Wissenschaften)

12:30-14:00      Mittagspause / Lunch break

14:00-14:30      **Führung EURAC / Guided tour through EURAC**

14:30-15:15      **Towards a Firthian notion of collocation**

*Stefan Evert, Sabine Bartsch* (Technische Universität Darmstadt)

- 15:15-16:00 **Exploration of quantitative phenomena for internet dictionaries, using the example of E-VALBU and KoGra-DB**  
*Roman Schneider* (Institut für Deutsche Sprache, Mannheim)
- 16:00-16:30 Kaffeepause / Coffee break
- 16:30-17:15 **Dictionary-cum-corpus-query-tools: A step towards more customisation in pedagogical lexicography?**  
*Magali Paquot* (Université Catholique de Louvain)
- 17:15-18:00 **Abschlussdiskussion / Final discussion**  
(Leitung / Chair: *Andrea Rapp*, Technische Universität Darmstadt)
- 19:00 **Gemeinsames Abendessen / Social dinner**  
Restaurant „Vögele“, Goethestr. 3, Bozen

**Freitag, 4. Mai 2012** (Seminar 2-3) (intern)

**Friday, May 4 2012** (Seminar 2-3) (closed)

---

- 08:45-09:00 Ankunft / Arrival
- 09:00-09:45 **Impulsreferate / Short presentations**
- 1) **Über visuelle Wege zur Korpusexploration / On visual approaches towards corpus exploration**  
*Egon Stemle/Verena Lyding/Lionel Nicholas* (EURAC Bozen)
  - 2) **Benutzungsstudien zu automatisch generierten Angaben / Usage studies on automatically generated lexicographic information**  
*Carolin Müller-Spitzer* (IDS Mannheim)
  - 3) **Automatisch generierte Angabe für (Meta-)Lexikografen / Automatically generated information for lexicographers**  
*Annette Klosa* (IDS Mannheim)
- 09:45-11:00 **Arbeitsgruppe I / Working group I – Automatisch generierte Angaben für Lexikografen / Automatically generated information for lexicographers**  
(Leitung / Chair: *Andrea Abel*, EURAC Bozen)  
**Arbeitsgruppe II / Working group II – Automatisch generierte Angaben und ihre Benutzer / Automatically generated information for users**  
(Leitung / Chair: *Stefan Engelberg*, IDS Mannheim)
- 11:00-11:30 Kaffeepause / Coffee break
- 11:30-13:00 **Präsentation der Ergebnisse aus den Arbeitsgruppen und Diskussion**  
(Diskussionsleitung / Moderator: *Angelika Storrer*, Technische Universität Dortmund)
- 13:00-14:00 Mittagspause / Lunch break
- 14:00-15:30 **Interne Besprechung und Fazit / Internal discussion and conclusion**  
(Leitung / Chair: *Annette Klosa*, IDS Mannheim)