

CORPUS ARCHITECTURE, CORPUS BUILDING AND CORPUS LINGUISTICS

WORKSHOP April 11 – April 15, 2016
Workshop language: English
VENUE: Casa Academiei, Calea 13 Septembrie, No. 13, Bucharest 050711, ROMANIA
Room 3015-3017, 3rd floor, West wing.

Alexander von Humboldt Foundation's *Research Group Linkage Programme* between the University of Bucharest, the Institut für Deutsche Sprache (Mannheim), associated with the Institute for artificial intelligence N. Drăgănescu (Bucharest) and the Institute of Computer Sciences (Iasi) of the Romanian Academy

Monday, April 11 2016 open session

Keywords: project, contrastive linguistics, corpus linguistics

09:00 - 10:45 Presentation of the University of Bucharest, Institut für Deutsche Sprache, Institute for artificial intelligence N. Drăgănescu (Bucharest) and the Institute of Computer Sciences (Iasi)

Presentation of project members and their research focus

10:45 - 11:15 Coffee break

11.15 - 12:00 Towards Contrastive Research based on Comparable Corpora
(Presentation of the project) (Marc Kupietz)

12:00 - 12:30 Corpora for contrastive linguistics (Ruxandra Cosma)

12:30 - 13:00 Corpus Linguistics for Computer Scientists I (Marc Kupietz)

13:00 - 13:30 Lunch break

14.30 - 15:30 Corpus Linguistics for Computer Scientists II (Marc Kupietz)

15:30 - 16:15 Towards Scalable Corpus Query Platforms –
Scientific, Organisational, Economical and Technical Aspects (Marc Kupietz)

16:15- 16:45 Coffee break

16:45 - 18:00 Corpus Query Platforms: KorAP (Nils Diewald)

Discussions - guidelines

Tuesday, April 12 2016 open session

Keywords: DeReKo and CoRoLA, presentation, discussions, decision making

Also skype conference with developer Radu Simionescu (Iasi)

09:00 – 10:00 National Language Institutes, EFNIL and Reference Corpora (Andreas Witt)

- 10:00 - 11:00 Introduction to DeReKo I (Marc Kupietz)
11:00 - 11:30 Coffee break
11.30 - 13:00 Introduction to DeReKo II – Corpus encoding (Marc Kupietz and Andreas Witt)
13:00 - 14:30 Lunch break
14:30 - 16:00 CoRoLa I Bucharest-team

Dan Tufiş: Current status of the CoRoLa corpus

Maria Mitrofan: Medical sub-corpus (to replace most of EMEA content)

Elena Irimia: Metadata schema and metadata generation

Verginica Barbu Mititelu: Romanian tree-bank component

Radu Ion: Natural Language Query Interface; current status and prospects

Ştefan Dumitrescu/Tiberiu Boroş: Speech component (including search)

- 16:00 - 16:30 Coffee break

- 16:30 - 18:00 CoRoLa II Iasi-team

Skype-session with developer Radu Simionescu (Iasi)

Discussions

Wednesday, April 13 2016 open to team members, co-workers and interested computer scientists

Keywords: KoRAP, description, organizational questions

- 09:00 - 11:15 KorAP: The gory details I (Nils Diewald)
11:15 - 11:45 Coffee break
11.45 - 13.15 KorAP: The gory details II (Nils Diewald)
13:15 - 15.00 Lunch break
15:00 - 16:15 Questions to KorAP, questions to queries, questions to harmonization procedures, German vs. Romanian language structure and morphological annotations I
16:15 - 16:45 Coffee break
16:45 - 18:00 Questions to harmonization procedures, German vs Romanian language structure and morphological annotations II

Thursday, April 14 2016 closed session

Keywords: Corpus architecture. Q & A.

- 09:00 - 11:15 Linguistics: research topics and questions

The case of the Romanian supine and of the German infinitive

(Alexandra Cornilescu/Anda Cosma)

- 11:15 - 11:45 Coffee break
- 11:45 - 13:00 How to do contrastive research with comparable corpora? (Brainstorming)
- 13:00 - 14:45 Lunch break
- 14:45 - 16:15 Specification of the Working plan
- 16:15 - 16:45 Coffee break
- 16:45 - 18:00 Discussion II

Friday, April 15 2016 closed session

Keywords: agenda

- 09:00 - 11:00 Discussion of organizational questions, agenda
- 12:00 Andreas Witt - talk at the University of Bucharest: Elektronische Korpora, Kodierung und Annotation (German)
- Faculty of Foreign Languages and Foreign Literatures, Department of Germanic Languages, room Goethe (Str. Pitar Mos 7/13)