

Unified Linguistic Annotation Transcontinental Perspectives

Koenraad De Smedt

Erhard Hinrichs

Sandra Kübler

Francis Bond

Workshop Co-chairs

State of the Art:

- Richly annotated corpora for a growing number of languages
 - using different encoding standards
 - using different theoretical assumptions
 - leading to a fragmented landscape of resources, often with limited access

Theme and Purpose of the ULA Workshop

- Discussing directions for overcoming this fragmented landscape
- eScience and Grid Computing provide new paradigms for collaborative research.
- They require a new type of infrastructure with clear rules of access and common data exchange formats.

Some Leading Questions

- What do we mean by „Unified Linguistic Annotations“ (ULA)?
- What role do multilinguality and endangered languages play?
- What do we mean by „interoperability“ in the context of linguistic resources?
- What would an infrastructure look like that supports ULA and interoperability?

It's a hot topic

- ACL-2005 Pie in the Sky Workshop
- LREC-2006 Workshop on Merging and Layering Annotation
- NSF Workshop on Unified Linguistic Annotation 2007 (Brandeis)
- NSF Workshop on Treebank Annotation (NAACL) 2007
- ACL 2007 Linguistic Annotation Workshop
- NSF Workshop on Toward Interoperability of Linguistic Resources (TILR, Stanford)

... And it continues to be ...

- First International Conference on Global Interoperability for Language Resources (Hong Kong) 2008
- Special Interest Group on Linguistic Annotation (SIGANN)
- Workshop proposals at LREC 2008
- More?

Workshop Structure

- Wednesday
 - Grammar and more Grammar
- Thursday
 - Resources and annotation
 - Panel on cooperative projects
- Friday–Saturday
 - Treebanks and Linguistic Theories

Wednesday

- Anette Frank:
Networks of Linguistic Annotation: The Linguist's Web
- Jonas Kuhn:
Broad-coverage LFG Parsing: Beyond Parallel Grammar De
- Julia Hockenmaier: What kind of a treebank
do we need for grammar extraction?
- Stephan Oepen:
Beyond the Science of the Wall Street Journal. WeScience
: Registration now Open — RSVP!
- Joakim Nivre:
Dependency Grammar and Dependency Parsing
- **General Discussion**
- **Dinner**

Thursday – 1

- Nancy Ide: LAF and GrAF
: The ISO standard for Linguistic Annotations
- Manfred Stede:
Deep text: Multi-level annotation for discourse research
- Marc Verhagen:
What's in a Link? Task decomposition and layered annotation
TimeML
- Andreas Witt: A Generalised
Architecture for Sustainable Access to Linguistically Annotated Text
- **General Discussion**

Thursday – 2

- **Panel**

An overview of current initiatives to support interoperability of linguistic resources.

- Yoshihiko Hayashi:

 - Toward an Ontology-based Language Service Infrastructure

- D. Terence Langendoen:

 - Opportunities for massive extensions of linguistic annotations

- Erhard Hinrichs:

 - CLARIN: A Pan-European Infrastructure Project

- **General Discussion**

- **Reception**

Why here and now?

- Research tradition in Bergen
- Bergen is hosting TLT'07
- Upcoming CLARIN project
- The Research Council of Norway wants to stimulate infrastructure projects
- Grant awarded by RCN
- Additional grant from the University of Bergen, Humanities Faculty