



Perspectives for Ontologies in Linguistics - introduction to the LREC Panel -

Peter Wittenburg

MPI for Psycholinguistics

Organizers:

Laurent Romary, Helen Dry, Ed Hovy, Peter Wittenburg



The Problem

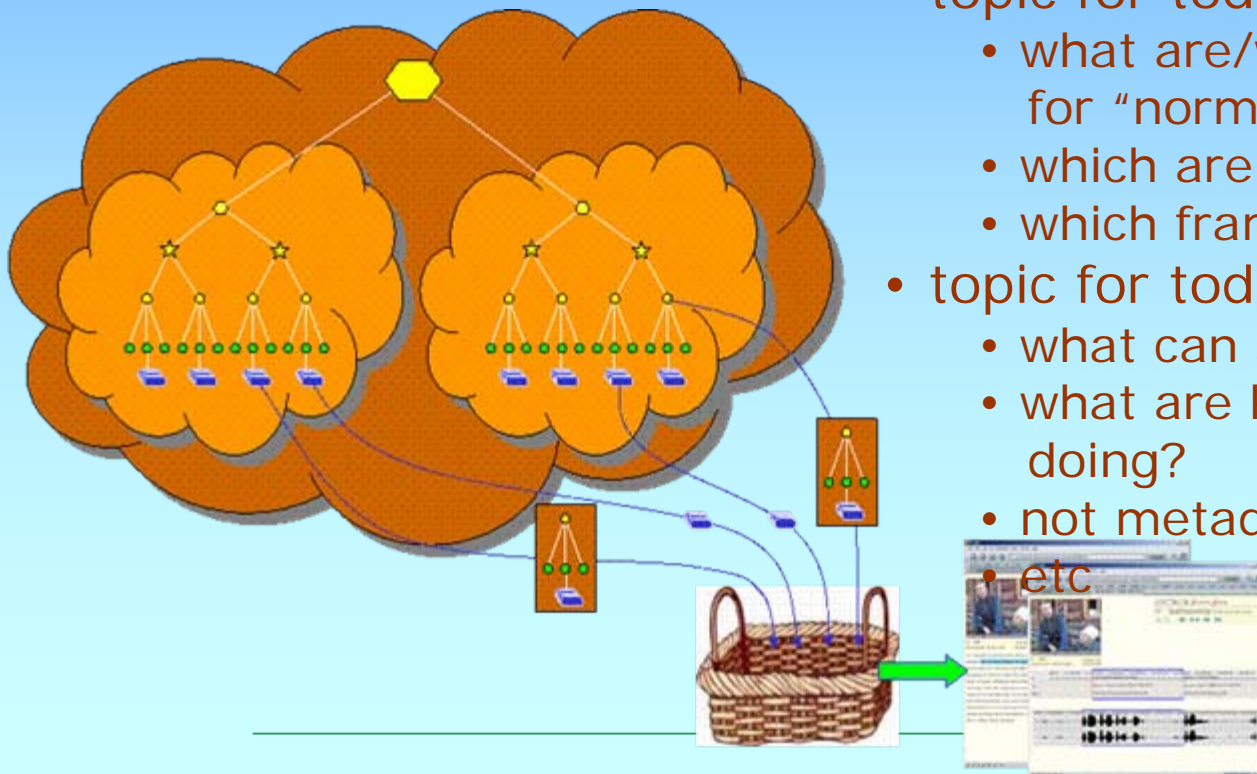
- all share the view that solving the semantic interop problems is one of the big challenges in our domain
 - at tag level
 - at value level in case of vocabularies
 - for concepts used in prose text

- topic for today is:

- what are/will be the options for "normal" linguists
- which are the practical problems
- which frameworks do we need

- topic for today is not:

- what can specialists all do?
- what are knowledge engineers doing?
- not metadata interop (IMDI, OLAC)
- etc





Current Activities

- much important work is being done
 - people have worked out formal frameworks to represent semantics (RDF, RDF-S, OWL, ...)
 - people work on “central/reference” ontologies such as
 - ISO DCR – flat registry of linguistic concepts grouped into profiles
 - GOLD Ontology – covering linguistic concepts including upper-level ontologies
 - Wordnets – grouped into synsets including different languages
 - what else?
 - creating open registries with standardized APIs
 - many bottom-up data driven approaches
 - is this all of use now and in future



Practical Aspects

- just a few practical aspects
 - used linguistic concepts are not (well) described
 - fuzzy type of relations dominant (“mapsTo/relatedTo”)
 - in a DOBES annotation resource about 200 concepts are used and there is no reference to any ontology and people will continue to do so:
how to manage if the researcher simply wants to execute a search including a few categories
 - central ontologies are covering a large amount of concepts – do we expect normal linguists to go through all of them
 - can others re-use the data driven knowledge – practically no if we have a dynamic domain



Egocentric Motivation

- at MPI now an archive with about 50 Mio annotations (utterances) in many different corpora from many different contributors covering a large number of languages
- except for some strange formats format interoperability
- create big indexes so that people can execute combined metadata and content searches via web-based frameworks (ANNEX, LEXUS, IMDI)
- working at Grid LR archive integration (DAM-LR project)
- what is it what we have to tell our linguists?
- participate in ISO TC37/SC4 -> DCR will solve all problems?
- what kind of knowledge resources do we have to build/include?
- how can we combine them?
- what kind of tools should we provide?



• • • Questions to the Panel

- **Scenario**

- what are the typical scenarios for the chosen approach?
- Which scenarios are not covered?
- What does it help/How much to be done/What is the cost/benefit ratio?

- **Theoretical Foundation**

- Is linguistic interop a matter of deep theory/logical consistency&completeness, semantic accuracy or of practical and temporal concerns?
- are linguistic concepts independent of the languages/are language specific variations part of the definition?
- are linguistic concepts subject of changes and do they have to be repressed?
- should relations be separated from the concept defs
- what is the required/allowed granularity of the definitions

- **Methodology**

- what is community agreement for "central" ontologies
- can we integrate "top-down" and "bottom-up" sources
- what are the most suitable frameworks/tools



Panel Setup

- sent these questions to a number of “experts” and asked for 7 minutes statements
- Laurent Romary guilty for ISO DCR
- Baden Hughes involved in GOLD initiative
- Adam Saulwick bottom-up approach
- Ed Hovy engineering solutions
- Alex Geyken critical linguist/potential user
- then open discussion hope that there many more experts
- Peter Wittenburg moderator of panel and discussion