Perspectives for Ontologies in Linguistics

Eduard Hovy, Helen Dry, Laurent Romary , Peter Wittenburg

1. Motivation

The usage of ontologies in the linguistic domain becomes increasingly important since linguists will make use of large and/or virtually integrated repositories of language resources created by different groups and individuals. We can expect that these resources (annotations, lexica, etc.) were originally created to describe/analyze specific languages with specific linguistic theories and purposes in mind. Performing operations (searching, comparing, etc) on combinations of such material therefore implies the need to overcome linguistic encoding differences, i.e., ontological knowledge has to be used and eventually be created. First ontologies claiming to include widely agreed concepts have been or are in the process of being created such as GOLD and the ISO TC37/SC4 Data Category Registry. They certainly have the potential of being used to overcome the semantic interoperability gap. On the other hand, however, there may be many objections or practical problems to re-using pre-defined concepts such as changing linguistic theories, the time needed to map legacy data, differences in the definition and value ranges of concepts, and others. Some researchers may, therefore, prefer to rely on simple frameworks that allow them to easily extract – either manually or automatically – concepts and the mappings between them by bottom up processes, thereby leaning towards "practical" ontologies.

Of course, we have to keep in mind different usage scenarios, for example from field and corpus linguists who want to compare two languages, to language engineers trying to optimize natural language processing algorithms. In addition, some ontologies describe a closed domain defined by a set of resources, while others seek to cope with quickly changing domains.

The panel will bring together experts working in the different areas, covering both the "bottom-up" and the "top-down" approaches. With the help of answers of the panelists to a few questions we want to better understand what kind of application scenarios are most suitable for the different approaches, what kind of services and approaches should become available in future to help the individual researcher, and how the current initiatives will develop.

2. Programme

2.1. Introduction Statement prepared by Hovy/Uszkoreit/Wittenburg
- Introduction to the goals of the panel
- Overview of different approaches

2.2. Panel
- Moderator of the discussion: Hans Uszkoreit
- Panelists
  - Laurent Romary (Background: ISO Data Category Registry)
  - Baden Hughes (Background: GOLD Ontology)
  - Adam Saulwick (Background: "manual" bottom-up Ontology Creation)
  - Eduard Hovy (Background: "automatic" bottom-up Ontology Creation)
  - Alexander Geyken (Linguist as potential user)

2.3. Discussion

3. Prepared Questions

3.1. Scenario
- What are the typical scenarios for the chosen approach?
- Which typical scenarios are not covered by the chosen approach?
- How does your approach help the linguist, what does he or she have to provide in addition, and what is the cost/benefit ratio?

3.2. Theoretical Foundation
- Linguistic Interoperability: Is it a matter of deep linguistic theory, logical consistency/completeness, semantic accuracy, or of practical and temporal concerns?
- Are linguistic concepts independent of the languages they describe? Are language specific variations part of the definition?
- Are linguistic concepts subject of changes and do the changes have to be represented?
- Does it make sense to store concepts independent from the relations amongst them, or does one lose relevant information?
- What is the required/allowed granularity of your ontological descriptions?

3.3. Methodology
- What does community agreement mean when speaking about central ontologies? How can agreement be achieved?
- Can "top-down" and "bottom-up" created ontologies be integrated or used together?
- What are most suitable frameworks/tools, given your usage scenarios?