Preface

Even if language resources covering English tend to receive most attention in the LT community, recent years have shown an increased interest in developing lexical semantic resources and semantically annotated corpora of also lesser-resourced languages, including the languages in the Nordic and Baltic region. Nevertheless, high-quality semantic resources with sufficient coverage still prove to be a serious bottleneck not only in purely rule-based NLP applications but also in supervised corpus-based approaches. Also in the Digital Humanities there is an increased interest in and need for semantic annotation which would enable more refined search in, and better visualization and analyses of large-scale corpus data.

This workshop focuses in particular on the interplay between lexical-semantic resources as resembled by wordnets, framenets, propbanks, and others and their relation to practical corpus annotation. The workshop – a follow-up on the successful Nodalida 2009 and 2013 workshops on semantic resources – intends to bring together researchers involved in building and integrating semantic resources (lexicons and corpora) as well as researchers who apply these resources for semantic processing. Also researchers who are more theoretically interested in investigating the interplay between lexical semantics, lexicography, corpus linguistics and Digital Humanities are welcome.

For the workshop we invited papers presenting original research relating to semantic resources for NLP and DH on topics such as:

- representation of lexical-semantic knowledge for computational use
- the interplay between lexical-semantic resources and semantically annotated corpora
- corpus-based approaches to lexical-semantic resources
- tools for semantic annotation
- terminology and lexical semantics: concept-based vs lexical semantic approaches
- monolingual vs. multilingual approaches to semantic lexicons and corpora
- quality assessment of lexical-semantic resources: criteria, methods
- applications using lexical-semantic resources (information retrieval, semantic tagging of corpora, MT, Digital Humanities etc.)
- machine-learning techniques to discover semantic structures such as unsupervised learning, distance supervision, or cross-language learning.
- traditional lexicography and NLP lexicons: re-use and differences
- applying semantic resources (lexica, corpora) for semantic processing
- word sense disambiguation based on lexically informed techniques

We received a total of 7 submissions, each of which was reviewed by at least two (anonymous) members of the program committee (see below). On the basis of the reviews, 5 submissions were accepted for presentation at the workshop and inclusion in the workshop proceedings (subject to revisions required by the reviewers).

The workshop was designed to be a highly interactive event. After an invited oral presentation by Johan Bos (University of Groningen): – *Issues in parallel meaning banking* – the other contributions to the workshop were presented during two oral sessions. A general discussion concluded the workshop.

Workshop organizers

- Bolette Sandford Pedersen (University of Copenhagen, Organizing Chair)
- Lars Borin (University of Gothenburg)
- Markus Forsberg (University of Gothenburg)
- Sanni Nimb (Association for Danish Language and Literature)
- Anders Søgaard (University of Copenhagen)
- Pierre Nugues (Lund University)
- Hector Martinez Alonso (University of Copenhagen)
- Anders Johannsen (University of Copenhagen)
- Sussi Olsen (University of Copenhagen)

Program committee

- Lars Borin (University of Gothenburg)
- Normunds Grūzītis (University of Gothenburg)
- Eiríkur Rögnvaldsson (Iceland University)
- Ruth Vatvedt Fjeld (University of Oslo)
- Markus Forsberg (University of Gothenburg)
- Karin Friberg Heppin (University of Gothenburg)
- Richard Johansson (University of Gothenburg)
- Rune Lain Knudsen (University of Oslo)
- Dimitrios Kokkinakis (University of Gothenburg)
- André Lynum (University of Oslo)
- Sanni Nimb (Association for Danish Language and Literature)
- Pierre Nugues (Lund University)
- Krister Lindén (University of Helsinki)
- Jussi Piitulainen (University of Helsinki)
- Bolette Sandford Pedersen (University of Copenhagen)
- Anders Søgaard (University of Copenhagen)
- Joel Priestley (University of Oslo)
- Heili Orav (University of Tartu)
- Esben Alfort (Ankiro, Denmark)
- Jørg Asmussen, (Association for Danish Language and Literature)
- Hector Martinez Alonso, (University of Copenhagen)
- Anders Johannsen, (University of Copenhagen)

Proceedings of the workshop on Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities at NODALIDA 2015

ii