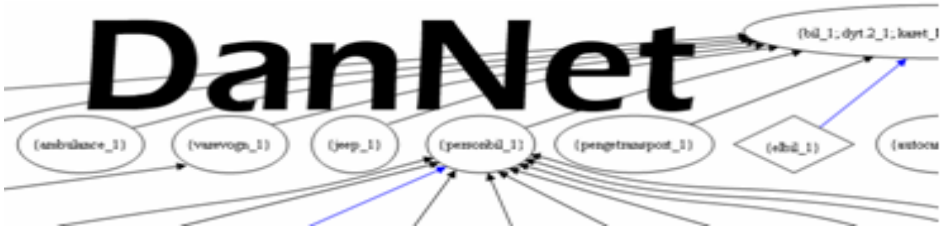

DanNet Symposium

March 12, 2009, University of Copenhagen, Njalsgade 126, room 23.0.49

Programme

9.30 – 10.00: Dean Kirsten Refsing: *Welcome*
Bolette S. Pedersen: *Introduction to DanNet*

10.00 – 11.00: Christiane Fellbaum: *Deep Text Understanding with WordNet*

11.00 – 11.15: *Coffee*

11.15 – 11.45: Eckhard Bick: *The Use of Semantic Prototypes in Semantic Role Annotation*

11.45 – 12.15: Henrik Legind Larsen: *Modelling and Utilizing Uncertainty in Ontology-based Semantic Search*

12.15 – 13.15: *Lunch*

13.15 – 14.15: Tony Veale: *Growing Finely-Discriminating Taxonomies from Seeds of Varying Quality and Size*

14.15 – 14.45: Sanni Nimb: *Linguistic Challenges in DanNet*

14.45 – 15.00: *Coffee*

15.00 – 15.30: Patrizia Paggio: *DanNet – Teaching and Research Perspectives at Center for Language Technology*

15.30 – 16.00: Lars Trap-Jensen: *Using DanNet for Better Dictionaries*

16.00 – 16.15: Bolette S. Pedersen: *Closing*

Registration: Please send email to bspedersen@hum.ku.dk before March 9.

Abstracts:

DanNet talks

Bolette Sandford Pedersen, University of Copenhagen,
Sanni Nimb, Lars Trap-Jensen, Society for Danish Language and Literature

The first version of DanNet is being released as an open-source resource in March 2009. This lexical semantic resource has been developed as a collaborative project between a research institution, Centre for Language Technology, University of Copenhagen, and a literary and linguistic society, Society for Danish Language and Literature under the Danish Ministry of Culture. The WordNet has been semi-automatically compiled on the basis of a traditional dictionary, Den Danske Ordbog (DDO), and a pilot version of a computational semantic resource built on ontological grounds (SIMPLE-DK). In the three talks we present the resource and discuss some of the linguistic challenges we have met in the process. In the case of intentional definitions, many aspects of the DDO definition are easily translated into the appropriate DanNet relations. Other cases are more complicated, due to different definitional styles and practices, e.g. by means of paraphrase. Even more challenging is the problem of 'missing' information in the dictionary, a dictionary definition leaning heavily on the user's ability to make assumptions. We will discuss some cases where it is important to encode explicitly the knowledge that a human user has about a concept. Finally, we discuss the online version of the DDO, which is currently being developed, and where DanNet data are used to improve the data structure of the dictionary to suit the electronic medium. In this project, an onomasiological query option is being offered, thereby giving more direct access to the WordNet data through a thesaurus-like interface.

Deep Text Understanding with WordNet

Christiane Fellbaum, Princeton University
(joint work with Peter Clark and Jerry Hobbs)

The biggest challenge for automatic text understanding is arguably presented by the vast amount of implicit information that is not expressed on the surface but that is a prerequisite for a coherent meaning representation. Language processing systems require a significant amount of lexical, encyclopedic and common sense knowledge, and creating resources meeting these needs remains a fundamental challenge. We have augmented WordNet as a knowledge resource for language understanding in several ways: adding formal versions of its word sense definitions (glosses); typing the morphosemantic links between nouns and verbs; and encoding a small number of "core theories" about WordNet's most commonly used terms. We describe the application of these enhancements in the context of a task focusing on the recognition of textual entailment. Although the improvement in performance is modest, we argue for the value of a linguistic-symbolic approach towards deep understanding of language.

The use of Semantic Prototypes in Semantic Role Annotation

Eckhard Bick, University of Southern Denmark

One effective approach to rule-based grammatical corpus annotation is progressive annotation, where grammar modules step-by-step produce different layers of morphological, syntactic and semantic annotation. My talk will focus on the combined use of lexical-ontological and syntactic information to produce automatic semantic role annotation for Danish text. Among other things, the DanGram-parser's (a) semantic prototype ontology and (b) semantic role inventory are presented, as well as a Constraint Grammar-based syntactic framework for extrapolating (b) from (a).

Modelling and Utilizing Uncertainty in Ontology-based Semantic Search

Henrik Legind Larsen, Aalborg Universty

Uncertainties of several kinds occur when combining and utilizing semantic resources for extracting information from document bases. In particular, in investigative business intelligence, domain ontologies and other semantic resources may be utilized in recognizing entities and establish links across documents. In such applications, it is important that no potentially relevant connection is missed. However, due to the nature of the unstructured data and the diverse semantic resources, the recognition and matching of name entities, as well as concepts, has to deal with imperfections of several kinds, such as imprecision, uncertainty, and inconsistency. I present an approach, in the frameworks of fuzzy logic and possibility theory, for handling such problems through utilization relevant context and semantic resources.

Growing Finely-Discriminating Taxonomies from Seeds of Varying Quality and Size

Tony Veale, University College Dublin

Concept taxonomies offer a powerful means for organizing knowledge, but this organization must allow for many overlapping and fine-grained perspectives if a general-purpose taxonomy is to reflect concepts as they are actually employed and reasoned about in everyday usage. For instance, people do not simply think of Tofu as a food, but as a soft food, a white food, a vegetarian food, a bland food, and so on. Many fine-grained perspectives are idiosyncratic and dynamic rather than universal and static. Nonetheless, general-purpose computational taxonomies will be called upon to perform both coarse-grained and fine-grained judgments. In NLP, for instance, the semantics of "eat" requires just enough knowledge to discriminate foods like tofu and cheese from non-foods like wool and steel, while specific applications in the domain of cooking and recipes (e.g., Hammond's CHEF system) require enough discrimination to know that tofu can be replaced with clotted cheese in many recipes because each is a soft, white and bland food.

In this talk I describe a means by which finely-discriminating taxonomies can be grown from a variety of different knowledge seeds. These fine-grained taxonomies comprise composite categories that can be lexicalized as phrases of the form "ADJ NOUN", such as Sharp-Instrument, which will serve to taxonomically organize the set of all instruments that are typically considered sharp, such as knives, scissors, chisels and can-openers. I will present empirical results from experiments using a variety of different starting points, or seeds, of differing quality and size, that are acquired from three different sources: WordNet, ConceptNet and the web at large.

DanNet – Teaching and Research Perspectives at Center for Language Technology

Patrizia Paggio, University of Copenhagen

My talk deals with current and future uses of DanNet at the Centre for Language Technology. Concerning teaching, DanNet is used in our course on Information Search to illustrate how wordnets are built, and the students are asked to employ domain-relevant fragments in a search engine implementation. From the point of view of research, we are interested in how to combine word sense tagging by means of DanNet with syntactic tagging and chunking, how to use parallel texts for word sense disambiguation, and how to use sense-tagged text in several tasks, e.g. for the extraction of frame semantic structures.