SEMANET: BUILDING AND USING SEMANTIC NETWORKS

INVITED SPEAKER:
Eduard Hovy (University of Southern California/Information Sciences Institute)

ORGANIZING COMMITTEE:
Grace Ngai[1]
Pascale Fung[1,2]
Kenneth W. Church[3]

PROGRAM COMMITTEE:
Eneko Agirre (University of the Basque Country)
Key-Sun Choi (NHK Labs)
Christiane Fellbaum (Princeton University)
Radu Florian (Johns Hopkins University)
Marc Light (MITRE)
Dekang Lin (University of Alberta)
Bernardo Magnini (ITC-irst)
Rada Mihalcea (University of Texas at Dallas)
Philip Resnik (University of Maryland)
German Rigau (Universitat Politècnica de Catalunya)
Charles Schafer (Johns Hopkins University)
Pavel Smrž (Masaryk University)
Keh-Yih Su (Behaviour Design Corporation)
Benjamin Tsou (City University of Hong Kong)
Piek Vossen (Irion Technologies)
Richard Wicentowski (Swarthmore College)
Dekai Wu (Hong Kong University of Science and Technology)

[1] Intendi Inc.,
HONG KONG

[2] Human Language Technology Centre
Hong Kong University of Science and Technology
Clear Water Bay, HONG KONG

[3] AT&T Labs
Florham Park
NJ07932, USA

The organizers would also like to thank Marine Carpuat (Hong Kong University of Science and Technology) and Pingwai Wong (Intendi Inc.), for their time and effort with putting together the proceedings.

PREFACE

WordNet (Miller et al., 1990) is an amazing resource. WordNet brings together a number of very attractive ideas: ontology, dictionaries, thesauruses, lexicography, very large corpora, comprehensive coverage and more from a highly interdisciplinary range of fields including psychology, linguistics, statistics and computer science. In his preface to Fellbaum (1998), George Miller credits a wide number of influences including Roger Schank (Artificial Intelligence scripts), Henry Kučera (the Brown Corpus), Robert Chapman (Roget's thesaurus), Philip Johnson-Laird and Mike Lesk. The basic idea of ontology is so natural and so attractive that it is hard to imagine how we could do anything in computational linguistics without one.

But as George Miller points out, much of the work on ontology was limited in scale. WordNet, at least at first, was no exception. The first WordNet demo in 1984 consisted of just 45 nouns. "An author might propose a semantic theory and illustrate it with some 20 or 50 English words (usually nouns), leaving the other 100,000 words of English as an exercise for the reader" (Fellbaum, 1998, p. xv). "Roger Schank and his colleagues were building language-processing systems having small vocabularies for well-defined topics ... but it was already becoming clear even in 1985 that this approach would have trouble scaling up" (Fellbaum, 1998, p. xvi). What makes WordNet truly interesting is what George Miller calls the comprehensiveness hypothesis. It is hard to remember, but coverage wasn't such a priority in the 1980s when many researchers were still working on "toy" sub-domains. CYC (Lenat, 1995) was a notable exception in that at least it set ambitious goals. What makes WordNet so remarkable is just how much coverage was actually achieved. Miller himself expresses dismay that when he first compared WordNet to COMLEX (www.cs.nyu.edu/cs/faculty/grishman/comlex.html) (Grishman et al., 1994), WordNet covered just 74% of COMLEX (Fellbaum, 1998, p. xix). While 74% may not be very good, or good enough, it is much more than could have ever been imagined back in the 1980s. Now that coverage is a priority, it just continues to improve.

At this point, a decade and a half after that first 45-noun demo, the success of WordNet is an accomplished fact. The number of citations is extremely convincing. I found nearly 1500 citations of WordNet at citeseer.nj.nec.com, far more than I could find for any other work on ontology including all of the influences mentioned above. While this is remarkable, citeseer.nj.nec.com does not properly reflect the fact that Chapman's thesaurus is a best seller and the Brown Corpus was widely cited in its day especially in the humanities. Nevertheless it would not be an exaggeration to say that WordNet is one of the most influential works ever in the area of ontology.

Admittedly, even according to citeseer.nj.nec.com, WordNet does not (yet) have a total monopoly on the concept of ontology — there are still more citations for "ontology" than for "WordNet" — but "WordNet" is catching up fast. It may not be long until the word "WordNet" replaces the word "ontology."

In addition to the large number of citations, it is remarkable just how many new projects have been inspired by WordNet (www.cogsci.princeton.edu/~wn): FrameNet (www.icsi.berkeley.edu/~framenet) (Baker et al, 1998; Fillmore, forthcoming), HowNet (www.keenage.com) (Dong, 2000) and EuroWordNet (www.hum.uva.nl/~ewn) (Vossen, 1998). One of the truly exciting developments in the WordNet area is all the work going on in languages other than English. It is also important for the long-term success of WordNet to draw attention to applications including the use of wordnets for sense disambiguation, machine translation and information retrieval.

This workshop will emphasize both the range of languages as well as the range of applications. It would not have been possible, of course, if WordNet had not become the de facto standard that it has become.

Kenneth W. Church
June 2002

References

1. Baker, C., Fillmore, C., and Lowe, J. (1998) *The Berkeley FrameNet project.* In Proceedings of the COLING-ACL, Montreal, Canada.
2. Dong, Z. (2000) *HowNet*, www.keenage.com.
3. Fellbaum, G., editor. (1998) *WordNet: An Electronic Lexical Database.* The MIT Press.
4. Fillmore, C. (forthcoming) *Linking sense to syntax in FrameNet*, keynote speech, Coling-2002.
5. Grishman, R., Macleod, C., and Meyers, A. *Comlex Syntax: Building a Computational Lexicon*, Coling 1994, Kyoto.
6. Kučera, H. and Francis, W. (1967). *Computational Analysis of Present-Day American English.* Brown University Press, Providence R.I.
7. Lenat, D. B. (1995). *CYC: A large-scale investment in knowledge infrastructure.* Communications of the ACM, 38(11).
8. Miller, G., et al. (1990). *Five Papers on WordNet*, Technical Report, No. 43, *Cognitive Science Laboratory*, Princeton Univ.
9. Vossen, P., editor. (1998), *EuroWordNet: A Multilingual Database with Lexical Semantic Networks.* Kluwer Academic publishers.