Moving toward Semantics for Language Processing:Recent Advances in Resource Construction and Application

Eduard Hovy
USC Information Sciences Institute
hovy@isi.edu

Abstact- Research in natural language processing over the past decade has produced amazing results using statistical methods. But increasingly there are signs that continued quality improvement for language processing applications (including QA, summarization, information extraction, and possibly even machine translation) require deeper and richer representations, possibly even (shallow) semantics of text meaning. But how can one define large-scale shallow semantic representations, and create shallow semantic content resources, that are stable and large enough to be adequate for the needs of statistically based NLP applications? In this talk I describe recent work at various locations, focusing on the resource components required (including ISI's symbol definition ontology Omega and the BBN-UPenn-Colorado-ISI large corpus OntoBank of (shallow) meaning representations) and the resources and methods one needs to build them (including existing ontologies, human annotation procedures, and a verification methodology, as embodied in a pilot project called Learning by Reading that involves various prominent researchers). This work is bringing together the NLP and KE/KR communities after a separation of some 40 years, and poses challenges for each side: dealing with large amounts of possibly incorrect knowledge is not something KE/KR has comfortable with, and dealing with requirements that knowledge be formalized and consistent is not something that large-scale NLP has been comfortable with. But should these challenges be resolved, the research discussed in this talk will enable numerous experiments on the wide-scale use of (shallow) semantics, to the benefit of NLP applications of all kinds.

0-7803-9361-9/05/$20.00 ©2005 IEEE