Preface

Sergei Nirenburg University of Maryland, Baltimore County

Most, if not all, high-end NLP applications—from the earliest, MT, to the latest, question answering and text summarization—stand to benefit from being able to use text meaning in their processing. It might be argued that without this ability one cannot expect real breakthroughs in NLP, to bring it to the level of public expectations and broad utility. At the same time, the bulk of work in the field has not, over the years, pertained to treatment of meaning. The main reason given is the complexity of the task of comprehensive meaning analysis.

This opinion was first formulated by Bar-Hillel over forty years ago on the strength of his famous *box in the pen* example. It concerned the following text: *Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy.* Bar Hillel wrote: "Why is it that a machine ... is ... powerless to determine the meaning of *pen* in our sample sentence within the given paragraph? The explanation is extremely simple ... What makes an intelligent human reader grasp this meaning so unhesitatingly is ... his knowledge [of] the relative sizes of pens, in the sense of writing implements, toy boxes and pens in the sense of playpens Whenever I offered this argument to one of my colleagues working on MT, his first reaction was: 'But why not envisage a system which will put this knowledge at the disposal of the translation machine?' Understandable as this reaction is, it is very easy to show its futility. What such a suggestion amounts to, if taken seriously, is the requirement that a translation machine should not only be supplied with a dictionary but also with a universal encyclopedia. This is surely utterly chimerical and hardly deserves any further discussion."

Using—for lack of space—a sweeping generalization, we can say that several generations of NLP workers agreed and proceeded to work on partial tasks (e.g., computational syntax), building applications that were feasible within the rather low bounds of attainable NLP quality that Bar Hillel's conclusion implies or looking for ways of bypassing the need of treating meaning (which partly accounts for the many "knowledge-lean", largely corpus-based endeavors that instead of treating meaning heads-on largely rely either on a variety of surface distance measures between input texts and texts that are known to be meaningful, or on complex counts of frequency of occurrence of strings in various contexts).

Our field, of course, has never been entirely uninterested in meaning. The tradition of formal semantics has been continuously maintained for many years. Philosophy of language and formal semantics has studied meaning—albeit predominantly that of closed-class lexical items, such as quantifiers—using truth values and versions of lambda calculus as the principal tools. Knowledge representation inside AI has come up with a large number of proposals concerning the metalanguages that could be used to formally represent text meaning. A variety of general and special (e.g., space- or time-related) logical and common-sense reasoning systems have been developed that facilitate inference making on the basis of the representation of "literal" meaning obtained from text. Lexical semantics has made significant progress theoretically (e.g., in the area of generative lexicon), descriptively (e.g., WordNet and the many satellite projects that it engendered) and processing-wise (as witnessed, e.g., by the SEMEVAL experiment). Large-scale descriptive work toward building encyclopedias of the kind alluded to by Bar Hillel has been going on for over 15 years (as exemplified, e.g., by CYC) and has recently become prominent as work on ontologies. Formal aspects of ontology work have also been studied. A novel and forward-looking application, the Semantic Web, has further popularized the need for automatic extraction, representation and manipulation of text meaning: for the Semantic Web to really succeed, capability of automatically marking text for content is essential, and

¹Bar-Hillel, Yehoshua. (1960) "The present status of automatic translation of languages," in Franz L. Alt (ed.), *Advances in Computers*, Vol. 1, Academic Press, New York.

this, arguably, cannot be attained reliably using only knowledge-lean, semantics-poor methods.

A third tradition has been devoted to the core problem of computational semantics-developing practical, increasingly broad-coverage meaning-oriented analysis and synthesis systems. Analysis systems take text, possibly, in a variety of languages, as input and, after having resolved a variety of types of ambiguity output expressions that represent, in a formal metalanguage, the meaning of the input text; synthesis systems take statements in a metalanguage and render them in one or more of natural languages. The meaning that such systems pursue typically concentrates on intended meaning but, whenever feasible, includes pragmatic, interpersonal, stylistic and even literal facets of text meaning, on the assumption that any of the above can be used to drive reasoning in a particular application. This tradition ascends roughly to the work of the conceptual dependency school of Schank and his associates and the procedural semantics of Winograd, Woods, and others, with influences from Katz and Fodor's linguistic semantics and the Meaning-Text model of Mel'čuk and his associates. Broad-coverage approaches within this tradition include, to name a few, preference semantics introduced by Wilks and relevant work by Hobbs, Hirst, and others. That work peaked in the late 1980s with the CMU KBMT-89 project. At this time, the field is experiencing a renaissance, as witnessed in the projects such as LaSIE at Sheffield and Mikrokosmos at NMSU, and the emergence of theories such as Nirenburg and Raskin's ontological semantics. The salient points of this knowledge-based approach include its empirical, heuristics-oriented character, its accent on content (rather than format) and its goal of attaining practical, broad-coverage results—as opposed to emphasis on formalism and partial tasks in formal semantics or acceptance of knowledge-lean approaches and, therefore, low upper bounds on eventual results in statistics-based approaches.

Recently, there has been a flurry of specialized meetings devoted to formal semantics (e.g., the traditional KR conferences, the SIGSEM conferences, the generative lexicon workshops), lexical semantics (WordNet, EAGLES/ISLE workshops), Semantic Web, formal ontology (FOIS), and others. The number of meetings devoted to knowledge-based text meaning processing has been much smaller even though this area, in some sense, provides the ultimate raison d'etre for most of the above fields and a conditio sine qua non for most advanced NLP applications.

The main goal of this workshop is to re-establish the research community of knowledge-based meaning processing and to help to explicate the currently implicit treatments of meaning in knowledge-lean approaches and how the advances in the latter and in formal semantics should influence the task.

We see this workshop not only as a forum for presenting complete work with tangible results (even though this will be encouraged) but also as an opportunity to (a) take stock of the developments in the field, (b) assess the nature of the most pressing extant problems and reasons for current lack of satisfactory solutions, (c) re-assess the potential contributions from developments outside the field (e.g., work on formal ontologies and corpus-based methods), and, especially, (d) to coordinate and plan future work.

Organizers

SERGEI NIRENBURG, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County. *sergei@umbc.edu*

GRAEME HIRST, Department of Computer Science, University of Toronto. gh@cs.toronto.edu

Program Committee

STEPHEN BEALE, University of Maryland, Baltimore County LYNN CARLSON, U.S. Department of Defense SANDA HARABAGIU, University of Texas at Dallas JERRY HOBBS, Information Sciences Institute, University of Southern California NANCY IDE, Vassar College RICHARD KITTREDGE, University of Montreal TANYA KORELSKY, CoGenTex, Inc MARJORIE MCSHANE, University of Maryland, Baltimore County DAN MOLDOVAN, University of Texas at Dallas MARTHA PALMER, University of Pennsylvania JAMES PUSTEJOVSKY, Brandeis University VICTOR RASKIN, Purdue University YORICK WILKS, Sheffield University