Role of KR in Natural Language Understanding and Synergic KR

Yuliya Lierler
University of Nebraska at Omaha

Research in *natural language understanding* (NLU) focuses on the design of systems that process information expressed in natural language and use it in reasoning. Key components of such systems are responsible for (a) transforming input in natural language (NL) into *logic form* – expression in a formal language capturing semantic information carried by the input; (b) performing required reasoning task on this logic form. Creation of advanced logic forms and methods for their construction given NL input can be seen as common challenges of the NLU and KR fields. Automated reasoning plays an important role in KR as its ultimate goal to advance both accumulation and processing of knowledge. In linguistics the focus is on development of logic forms that capture numerous phenomena displayed by languages, but automation is rarely of concern in this field. Work by Bos and Markert on the system NUTCRACKER for the task of Recognizing Textual Entailment (RTE) is an exceptional example when a logic form language – discourse representation structures (DRS) – is used both for capturing and processing information in NL using FOL theorem provers and model builders. Yet only a subset of the DRS language compatible with FOL is utilized by NUTCRACKER.

There are a couple of challenging questions. How similar logic form languages are to KR languages? Are current KR languages and respective automated reasoning tools applicable to logic form languages or, in other words, can they address the phenomena that natural language expressions encode? Developing systematic ways for constructing logic forms given arbitrary NL input is also a challenge. System BOXER by Bos is a rather unique example of a tool that translates English sentences into FOL. It has limitations. One of them is related to the fact that FOL is inappropriate for capturing NL input. Another peculiar aspect of NL expressions is that each word carries independent information that has to be represented formally in order to process utterances. For example, sentence *John walks* suggests not only that *John* executes an action *walk* but also commonsense knowledge associated with action *walk* – an entity walking changes its location. Designing procedures that combine logic forms stemming from a NL input together with "background knowledge" associated with its words poses another challenge. Lexical databases WORDNET, FRAMENET, VERBNET, PROPBANK, NOMBANK, and commonsense knowledge bases as CONCEPTNET, KNEXT provide wide scope of machine readable knowledge. If and how can they be used to advance NLU and KR? Recent tasks RTE, COPA (Choice of Plausible Alternatives), and Winograd Schema are good examples of problems that require sophisticated logic forms, background knowledge, inference mechanisms. These tasks may serve as an inspiration for the directions of research discussed above.

As KR matures, it offers a variety of languages and reasoning procedures appropriate for different tasks at hand, such as scheduling and planning, to name a few. For many application domains, including NLU, systems that combine multiple KR languages and tools are necessary. Developing systematic means for combining (a) heterogeneous KR languages and (b) various reasoning techniques under one roof by no means is a solved issue. Advances in satisfability modulo theories and constraint answer set programming demonstrate a potential for this direction of research. For instance, constraint programming is known for being an efficient tool for solving scheduling problems, whereas answer set programming is effective technology for addressing elaborate planning domains. Constraint answer set programming that unifies these two KR sub-fields is best for solving problems that require both scheduling and planning capabilities of underlying tools.