Can We (and Should We) Make Formal Sense of General Knowledge Expressed in Ordinary Language?

Lenhart Schubert
Computer Science Department
University of Rochester
Rochester, NY 14627-0226
schubert@cs.rochester.edu

Abstract
It has generally been assumed that the knowledge employed by an AI reasoning system needs to be in an unambiguous, formally interpretable form. From that perspective, general knowledge expressed in ordinary language (e.g., “dogs bark”) is unacceptably ambiguous and incomplete. However, we can achieve at least a partial transformation of such knowledge into formal, generically quantified sentences by taking account of properties of words and phrases such as the aspectual category, tense, Levin class, and presuppositions of verbs, or the classification of predicates (adjectival, nominal, verbal) as applicable to objects or kinds of objects.

Among the questions that arise are (1) to what extent word sense disambiguation is necessary for enabling useful inference, (2) how definites in the consequent of an if-then sentence, referring back to an indefinite in the if-part (“donkey anaphora”), should be represented, and (3) what general and particular inferences are enabled by (more or less formal) generic sentences. I will provide some preliminary illustrative results, and speculate about the possibility of a compromise between the “crisp” inference methods of traditional KR&R and the looser methods emerging from statistical entailment and natural logic.

Biographical Statement
Lenhart Schubert is a professor of computer science at the University of Rochester, with interests in knowledge representation and acquisition, language understanding, and self-aware agents. He is a Fellow of the AAAI, has served as program chair of several major conferences in NLP and knowledge representation, and has numerous publications in those areas.