Finding Meaning: The Missing Link Between Natural Language and Formal Ontology

Brandon BENNETT [a,1]

[a] *Institute for Artificial Intelligence and Biological Systems*
School of Computing, University of Leeds, UK

Abstract. The obvious starting point for investigating meaning is the analysis of natural languages and their usage. Even from ancient times, examination of reasoning expressed in natural language revealed that patterns of valid inference often follow certain patterns. Moreover, these can be expressed as formal rules that are idealisations of actual natural language arguments (e.g. Aristotle's syllogisms). Subsequent development of artificial symbolic languages and denotational semantics have lead to formal representations that are completely precise in terms of their truth conditions and rules of valid inference.

In parallel to this mathematical approach to semantics, linguists and cognitive scientists have continued to be interested in actual language usage. Empirical investigations strongly indicate that natural languages operate in ways that are far less precise and less constrained than formal languages. In particular, natural language terminology is very often vague and ambiguous, and its interpretation often highly dependent on contextual factors. Hence, many linguists prefer to describe language, in terms of statistical properties of linguistic tokens occurring within corpora of language use examples, rather than in terms of a formal representation.

More recently the requirements of information technology have lead to the construction of "formal ontologies", which seek to capture the meanings of linguistic terms by means of axioms and/or definitions expressed using a formal language. In its most naive form, an ontology simply identifies each of a set of natural language words with a symbol in a formal language. But of course this is problematic, because the formal language is not equipped to deal with the ambiguity, vagueness or context sensitivity that may affect the interpretation of the term.

My talk will explore ways that formal ontology can be reunited with empirical investigations of language usage. I will suggest that establishing a relationship between natural language terminology and a formalised representation requires several aspects of semantic variability to be taken into account. This requires both extension of the formal apparatus associated with an ontology and also empirical, statistical information about the usage of linguistic terms. I outline my own approach which using my "Statistical Standpoint Semantics" to incorporate variability into a formal language and "Corpus Guided Sense Cluster Analysis" to establish a mapping between linguistic terms and formal symbols.

---

[1] e-mail: brandon@comp.leeds.ac.uk