
Martha Palmer
University of Pennsylvania
mpalmer@linc.cis.upenn.edu

Moving towards applications by augmenting verb classes with information structure

How, apart from their role in support of machine translation, might interlinguas be applied to various other information processing tasks (e.g., text summarization, information extraction, query systems, information retrieval, tutoring, multimodal communication, and the like)?

This depends very much on what the "interlingua" is. If it is a canonical "English" semantic network representation that includes coreference and links to the domain model(GRAIL <example>,

Mikrokosmos <example>), then it functions very much like a standard deep semantic/pragmatic representation (a la Pundit <example>). As such it would provide an appropriate basis for performing the tasks that comprise MUCK evaluations: named entity, template elements, coreference, and finally scenario templates <example>. But that does not necessarily mean that it does a good job of capturing cross-linguistic generalizations.

Is it possible to include the hooks for cross-linguistic generalizations in the canonical "English" semantic network? Yes, Mikrokosmos, also Pundit used LCSs, and Dorr bases an interlingua on LCSs, so it has to be possible - the primitive LCS predicates can correspond to supertypes of the verbs (and nouns). But that means that these supertypes must be carefully chosen to be universal or at least multilingual. Then they could be represented as either predicates or features that can map onto both source and target languages.

If the interlingua focusses more on cross-linguistic generalizations, (a la Dorr's LCSs <example>, a la verb classes in STAGs <example>) then it would need to be augmented with an information structure (a la Doran and Stone <example>), that would include the semantic and pragmatic information necessary for building an application. Does this then look any different from the canonical network mentioned above? Not necessarily, although it could allow two different languages to have different underlying predicate argument structures (a la STAG).

The two languages would need to share a discourse model and a domain model, and a set of common verb and noun supertypes that could be co-indexed in order to capture cross-linguistic generalizations. As long as the entities that are referred to by the arguments can be co-indexed by the source language rep and the target language rep, and the important cross-linguistic supertypes can be shared, then the representation can function as both an interlingua and a basis for applications requiring semantics/pragmatics.
