Interlinguas don't need language-specific information

Lori Levin
Carnegie Mellon University
Lori_Levin@prague.mt.cs.cmu.edu

The C-STAR consortium is currently in the process of designing an Interchange Format (IF). The partners of C-STAR (ATR Japan, ETRI Korea, CMU USA, U. of Karlsruhe Germany, Siemens Germany, IRST Italy, and ?? France) are collaborating on a multi-lingual speech translation demonstration scheduled for 1999. The current semantic domain for C-STAR is meeting scheduling with two dialogue participants. We are now moving on to a more general travel planning domain with multi-party dialogues.

A current point of negotiation among the C-STAR participants is whether the IF should reflect source-language syntax and phrasing. Arguments in favor of retaining source language structure include (1) source language noun phrases must be retained in the IF as antecedents of source language anaphors (I took a bath. It (the bath) was hot. vs. ?I bathed. It was hot.) and (2) it's hard enough to write an analyzer/generator for a language without having to worry about compatibility with other grammars.

My position is: Languages are different and you have to put in some effort to relate them to each other. You can put the effort into transferring between language specific representations or you can put the effort into identifying universal features of an interlingua. Either way you do the same work. In a multi-lingual system, it makes more sense to work with one language independent interlingua. That way each grammar developer has to learn only one system of meaning representation, instead of learning how to relate his/her language to several language-specific pseudo-interlinguas. (In a longer version of this position paper I can address the opposing position point by point, showing that having language-specific features in the IF will not save any time or effort.)

In moving to a larger domain, another issue we will have to deal with is whether to have one uniform interlingua or separate domain-specific interlinguas for the components of travel planning (scheduling, reserving, shopping, etc.) It looks like we are headed for several sub-domain interlinguas. This implies that part of the translation process is classifying utterances into sub-domains.