Semantic Frame:
A Flexible Interlingua for Machine Translation and Human/Machine Interaction

Young-Suk Lee and Clifford Weinstein
ysl@sst.ll.mit.edu
MIT Lincoln Laboratory

To serve as a practical tool in a machine translation system, an interlingua must be straightforwardly derivable from the analysis module, and represented in a form from which a well formed sentence of any language can be easily generated. In addition, categories/features in an interlingua has to be easily manipulable so that the system developer can add/delete some of the categories until it takes the form of a truly language-independent meaning representation expressed by language universal categories/features.

The English/Korean machine translation system developed by MIT Lincoln Laboratory under DARPA sponsorship, which is based on the understanding/generation (TINA/GENESIS) system developed by MIT Laboratory for Computer Science, produces an interlingua called SEMANTIC FRAME. SEMANTIC FRAME satisfies the above mentioned conditions for an ideal interlingua in a practical machine translation system, and has proven to be effective in multilingual Human/Machine Interaction systems.

A semantic frame is directly derived from the parse tree. All major parse tree constituents (regardless of whether they are semantic or syntactic) are reduced into one of the three language neutral categories in a semantic frame, namely, clause-type, topic and predicate. All clause-level categories such as statement, infinitives are mapped onto "clause." All noun phrase expressions are mapped onto "topic." All modifiers as well as verb phrases are mapped onto "predicate." Reduction of all major parse tree categories into one of the three semantic frame categories enables the generation system to easily produce a syntactically well-formed sentence of any language, especially the well-formed word order of the target language. In principle, however, there is no limit to the kind/number of categories/features which can be expressed in a semantic frame. Linguistic features like 'tense,' 'number,' etc. are easily added or deleted depending on the need.

With SEMANTIC FRAME as an interlingua, the system produces high quality translation output of naval operational messages of a highly telegraphic nature (see Weinstein et al. 1996 for details) as well as other more natural texts of English. In the presentation, we will discuss capabilities and limitations of the system in detail, and some informal ideas about how to overcome the system limitations.

References

James Glass, Joseph Polifroni and Stephanie Seneff. 1994. Multilingual Language Generation Across Multiple Domains. In "Proceedings of the 1994 International Conference on Spoken Language Processing," Yokohama, Japan.

Stephanie Seneff. 1992. TINA: A Natural Language System for Spoken Language Applications. Computational Linguistics, 18-1, pages 61-88.

Clifford Weinstein, Dinesh Tummala, Young-Suk Lee and Stephanie Seneff. 1996. Automatic English-to-Korean Text Translation of Telegraphic Messages in a Limited Domain. In "Proceedings of COLING '96," Copenhagen, Denmark.

Victor Zue. 1996. Research and Development of Multilingual GALAXY: A Status Report. In "Proceedings of C-STAR II 96 ATR International Workshop on Speech Translation," Kyoto, Japan.