Pangloss: A Machine Translation Project

Sergei Nirenburg

Center for Machine Translation
Carnegie Mellon University
Pittsburgh, PA 15213

The project involves three sites (NMSU, USC, CMU) and is devoted to enhancing the state of the art in machine translation of natural language texts.

Pangloss uses a hybrid, multi-engine approach, though knowledge-based machine translation takes a majority of resources.

Types of work in the *knowledge-based* direction include:

- building high-quality syntactic, semantic and pragmatic analyzers for the source languages (Spanish and Japanese);

- developing a high-quality text planner and generator for the target language (English);

- designing a highly-expressive text meaning representation language (the interlingua) in which the results of analysis and input to generation are recorded;

- putting togethera detailed domain model ("ontology") and, for each of the languages involved, a lexicon which maps lexical units of respective languages into elements of the world model and a grammar.

Other types of work includes

- developing an *example-based* machine translation engine, a first system of this kind to attempt processing of full text rather than specific examples;

- developing a back-up lexical transfer system based on a battery of glossaries, dictionaries, gazetteers, etc. and fortified with morphological analysis and synthesis routines;

- developing a mechanism for integrating the results of the multiple MT engines in a hybrid, *multi-engine* system;

- continuing development, deployment and testing of a human-computer interface, a *translator's workstation*, to support postediting and interactive editing during the operation of the translation system; and

- continuing development of a set of knowledge acquisition tools and utilities to support the above work.