The KANTOO MT System: Controlled Language Checker and Knowledge Maintenance Tool

NAACL 2001 Demonstration

Teruko Mitamura, Eric Nyberg, Kathy Baker,
David Svoboda, Enrique Torrejon and Michael Duggan

Language Technologies Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
teruko@cs.cmu.edu

1 Technical Abstract

We will present the KANTOO machine translation environment, a set of software services and tools for multilingual document production. KANTOO includes modules for source language analysis, target language generation, source terminology management, target terminology management, and knowledge source development (see Figure 1).

KANTOO is a knowledge-based, interlingual machine translation system for multilingual document production. KANTOO includes: a) an MT engine, the result of fundamental redesign and reimplementation of the core algorithms of the KANT system [2, 4]; and b) a set of off-line tools that support the creation and update of terminology and other knowledge resources for different MT applications. The demonstration will focus on two of the newer capabilities in KANTOO:

- **Controlled Language Checker (CLC).** The CLC is a thin Java client which supports interactive editing and automatic checking of XML documents. This tool performs vocabulary and grammar checking on each sentence in a document. The checker accesses the KANTOO Analyzer (running as a separate network service), which performs tokenization, morphological processing, lexical lookup, syntactic parsing and semantic interpretation. If a sentence does not pass the check, then a diagnostic message is produced for the user to resolve.
- **Knowledge Maintenance Tool (KMT).**
  A variety of rule-based knowledge sources must be maintained in the KANTOO system. Chief among them are the syntactic grammars for the source and target languages. The biggest challenges for updating rule-based knowledge sources effectively rest in the potential complexity of the debug/test cycle. Changing a particular rule might result in widespread changes in grammar coverage, or regressive failures. The Knowledge Maintenance Tool (KMT) is used by the developer to test individual updates, with recourse to

full regression testing on various reference corpora. All changes to the knowledge are managed under explicit version control, so that it is straightforward to synchronize the knowledge sources for different releases. The KMT also includes an interactive tracing and debugging environment which utilizes the KANTOO Analyzer and Generator servers.

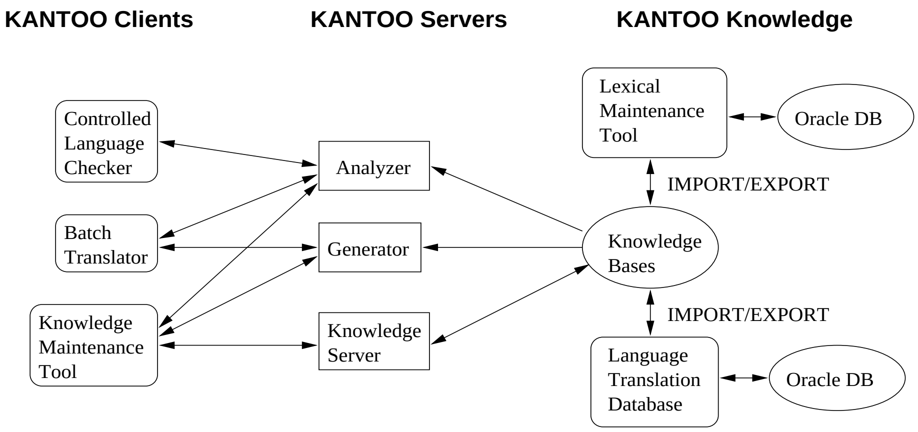

Fig. 1. KANTOO System Components.

For more information about the KANTOO technology, visit the KANT Project home page: `http://www.lti.cs.cmu.edu/Research/Kant`

References

1. Kamprath, C., Adolphson, E., Mitamura, T. and Nyberg, E.: Controlled Language for Multilingual Document Production: Experience with Caterpillar Technical English. In: Proceedings of the Second International Workshop on Controlled Language Applications (1998)
2. Mitamura, T., Nyberg, E. and Carbonell, J.: An Efficient Interlingua Translation System for Multi-lingual Document Production. In: Proceedings of the Third Machine Translation Summit (1991)
3. Mitamura, T. and Nyberg, E.: Controlled English for Knowledge-Based MT: Experience with the KANT System. In: Proceedings of TMI-95 (1995)
4. Nyberg, E. and Mitamura, T.: The KANT System: Fast, Accurate, High-Quality Translation in Practical Domains. In: Proceedings of COLING-92 (1992)