

GAZELLE: Using Statistics to Fill Knowledge Gaps in MT

Kevin Knight, Bonnie Glover-Stalls, Uli Germann, Ulf Hermjakob,
Yaser Al-Onaizan, Philipp Koehn, Irene Langkilde, Kenji Yamada
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey CA 90292
knight@isi.edu
<http://www.isi.edu/natural-language/mt/japangloss.html>

System Demonstration

Knowledge-based machine translation (KBMT) techniques yield high quality in domains with detailed semantic models, limited vocabulary, and controlled input grammar. Scaling up along these dimensions means acquiring large knowledge resources. It also means behaving reasonably when definitive knowledge is not yet available.

This presentation describes GAZELLE, a broad-coverage MT system that uses robust statistical techniques to fill various KBMT knowledge gaps. We currently translate Japanese, Spanish, and Arabic to English. Our approach includes:

- a natural language generation system called Nitrogen, which combines knowledge from a human linguist with that of an automatic “learning by reading” program.
- a robust syntactic analyzer that skips trouble words.
- a large conceptual knowledge base called Sensus, plus large syntactic and semantic knowledge bases.
- a transliterator for unknown names and technical terms (e.g., anjira <-> angela).

We will describe GAZELLE and demonstrate the software within a prototype translating photocopier called AUTOSCRIBE. Selected references:

“Filling Knowledge Gaps in a Broad-Coverage MT System,” (K. Knight, I. Chander, M. Haines, V. Hatzivassiloglou, E. Hovy, M. Iida, S. K. Luk, R. Whitney, and K. Yamada), Proc. Intl. Joint Conference on AI (IJCAI), 1995.

“Machine Transliteration,” (K. Knight and J. Graehl), Proc. Conf. of the Association for Computational Linguistics (ACL), 1995.

“A Controlled Skip Parser,” (K. Yamada), Proc. 2nd Conf. of the Assoc. for Machine Translation in the Americas (AMTA), 1996.

“Two-Level, Many-Paths Generation,” (K. Knight and V. Hatzivassiloglou), Proc. Conference of the Association for Computational Linguistics (ACL), 1995.

“Unification-Based Glossing,” (V. Hatzivassiloglou and K. Knight), Proc. International Joint Conference on Artificial Intelligence (IJCAI), 1995.

“Building a Large-Scale Knowledge Base for Machine Translation,” (K. Knight and S. Luk), Proc. National Conference on Artificial Intelligence (AAAI), 1994.