# Query-focused Summarization Using Text-to-Text Generation: When Information Comes from Multilingual Sources

**Kathleen McKeown**
Department of Computer Science
Columbia University
kathy@cs.columbia.edu

## Abstract

The past five years have seen the emergence of robust, scalable natural language processing systems that can summarize and answer questions about online material. One key to the success of such systems is that they re-use text that appeared in the documents rather than generating new sentences from scratch. Re-using text is absolutely essential for the development of robust systems; full semantic interpretation of unrestricted text is beyond the state of the art. Better summaries and answers can be produced, however, if systems can generate new sentences from the input text, fusing relevant phrases and discarding irrelevant ones. When the underlying sources for summarization come from multiple languages, the need for text-to-text generation is even more pronounced.

In this invited talk I present research on query-focused summarization over a variety of sources, including news, broadcast news, talks shows and blogs. Our research combines approaches from summarization and information extraction to answer open-ended questions. Because our sources include informal genres as well as formal genres and draw from English, Arabic and Chinese, text-to-text generation is critical for improving the intelligibility of responses. In our systems, we exploit information available at question answering time to edit sentences, removing redundant and irrelevant information and correcting errors in translated sentences.



*Proceedings of the 2009 Workshop on Language Generation and Summarisation, ACL-IJCNLP 2009*, page 3,
Suntec, Singapore, 6 August 2009. 