

White Paper Proposal
11-734 Advanced MT Seminar
Aaron B. Phillips

Overview

I propose investigating the role of context in machine translation. Context is an obvious source of information important in translation. However, what contextual features to use, how to score them, and how to integrate them into machine translation remains unclear. The work by Carpuat and Carbonell, while both claiming to perform context-dependent translation, employ very different strategies. Carpuat starts with a static phrase table learned from a bilingual corpus, but dynamically rescores the entries based on the input. This process uses context as a “feature” within a traditional SMT framework. Carbonell takes an entirely different approach and uses a small, high-precision phrase table. The decoder is then responsible for finding very long ngrams from a monolingual corpus that include the entries from the phrase table (possibly reordered and including other words). The “context” here is the fact that extremely long phrases are being used by the decoder, which in theory preserves the original context of the phrase. Carpuat focuses on preserving context on the source side of the translation while Carbonell focuses on preserving context on the target side of the translation. Furthermore, there is a connection between some work in context and adaptation (specifically translation model adaptation—not language model adaptation). An example of this would be the work by Hildebrand where a phrase table is built from documents with the best TF-IDF scores. TF-IDF is a feature of the document context. While Carpuat uses context features to dynamically rescore a traditional phrase table, while Hildebrand uses context to select a dynamic sample of the data from which she builds a traditional phrase table. This goes to show how the line at times may blur. I plan to focus on methods for incorporating context within a traditional framework (Carpuat), but as time permits I will also look at more exotic approaches (Carbonell) and the interrelatedness to translation-model adaptation (Hildebrand).

Suggested Reading

Marine Carpuat and Dekai Wu. [Context-Dependent Phrasal Translation Lexicons for Statistical Machine Translation](#). Machine Translation Summit XI. Copenhagen: September 2007.

Jaime Carbonell, Steve Klein, David Miller, Mike Steinbaum, Tomer Grassiany and Jochen Frei. [Context-Based Machine Translation](#). Proceedings of the 7th Conference of the Association for Machine Translation of the Americas. Cambridge, Massachusetts: August 2006.

Almut Silja Hildebrand, Matthias Eck, Stephan Vogel, and Alex Waibel. [Adaptation of the Translation Model for Statistical Machine Translation based on Information Retrieval](#). Proceedings of EAMT 2005. Budapest, Hungary: May 2005.

Preferred Presentation Dates

- March 5th
- February 27th
- March 19th