Phrase Sense Disambiguation for context-rich lexical choice in statistical machine translation

Abstract:

While there has been considerable progress in Statistical Machine Translation (SMT) in recent years, standard SMT architectures still rely on surprisingly little context to translate source words in the target language. In contrast, rich context features have proved very informative for closely related tasks in lexical semantics, such as word sense disambiguation. In this talk, I will describe my work on Phrase Disambiguation (PSD), which aims to leverage insights from sense disambiguation modeling in SMT. I will show the impact of source sentence context on translation quality. I will then present a study of translation lexical choice at the document level, and show that while SMT typically models translation at the sentence level, translation disambiguation benefits from wider document context.

Bio:

Marine Carpuat is a Research Associate at the National Research Council of Canada, where she works on natural language processing, statistical machine translation, and machine learning for multilingual text and document processing. She received a PhD in Computer Science from the Hong Kong University of Science and Technology (HKUST) in 2009, an MPhil in Electrical Engineering from HKUST in 2003, and an engineering degree from the French Grande Ecole Supélec. She was a postdoctoral researcher at Columbia University from 2008-2010.