



Marine Carpuat

(University of Maryland)

Friday, Sept. 28

2:30 pm Talk, Porter Hall 100
4:00 pm Snacks, LTI 5th Floor

SEMANTIC AND STYLE DIVERGENCES IN MACHINE TRANSLATION

Neural machine translation achieves impressive quality when trained on millions of parallel sentences (translations produced by humans). However, there is growing evidence that parallel sentences are not equally useful. While noise and domain mismatch naturally affect their usefulness, we hypothesize that more subtle semantic and style divergences, reflecting choices made by human translators, also matter for machine translation. We first show that mismatches in the meaning of source and target are surprisingly frequent in parallel corpora, and have a substantial impact on neural machine translation at training and decoding time.

Marine Carpuat is an Assistant Professor in Computer Science at the University of Maryland. Her research focuses on multilingual natural language processing and machine translation. Marine is the recipient of an NSF Career award, research awards from Google and Amazon, best paper awards at *SEM and TALN, and an Outstanding Teaching Award. She received her PhD in Computer Science from the Hong Kong University of Science & Technology, a Mphil in Electrical Engineering from the Hong Kong University of Science & Technology and a Diplome d'Ingenieur from the French Grande Ecole Supelec.