INVITED TALK: How Much Information Does a Human Translator Add to the Original?

Kevin Knight

ISI, University of Southern California, USA

knight@isi.edu

Abstract

It is well-known that natural language has built-in redundancy. By using context, we can often guess the next word or character in a text. Two practical communities have independently exploited this fact. First, automatic speech and translation researchers build language models to distinguish fluent from non-fluent outputs. Second, text compression researchers convert predictions into short encodings, to save disk space and bandwidth. I will explore what these two communities can learn from each others' (interestingly different) solutions. Then I will look at the less-studied question of redundancy in bilingual text, addressing questions like "How well can we predict human translator behavior?" and "How much information does a human translator add to the original?" (This is joint work with Barret Zoph and Marjan Ghazvininejad.)

Bio

Kevin Knight is Director of Natural Language Technologies at the Information Sciences Institute (ISI) of the University of Southern California (USC), and a Professor in the USC Computer Science Department. He received a PhD in computer science from Carnegie Mellon University and a bachelor's degree from Harvard University. Prof. Knight's research interests include machine translation, automata theory, and decipherment of historical manuscripts. Prof. Knight co-wrote the textbook "Artificial Intelligence", served as President of the Association for Computational Linguistics, and was a co-founder of the machine translation company Language Weaver, Inc. He is a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI), the Association for Computational Linguistics (ACL), and the Information Sciences Institute (ISI).