Cognitive Limitations Impose Advantageous Constraints on Word Segmentation

Katarzyna Hitczenko
Yale University, New Haven, CT, USA

Gaja Jarosz
Yale University, New Haven, CT, USA

Abstract: This paper analyzes the learning constraints imposed by cognitive limitations on memory and processing in the domain of word segmentation. We analyze the properties of three learning algorithms that rely on the same probability model: an incremental, constrained learner proposed by Venkataraman (2001), a batch version of Venkataraman's learner, and a variant of the batch model that imposes memory constraints mimicking incremental processing. We show that while Venkataraman's original incremental learner performs well, the batch version predicts massive undersegmentation. We identify the crucial properties of the probability model responsible for undersegmentation and show how memory constraints impose an opposing pressure favoring segmentation. Our results indicate that cognitive limitations exert pressures during learning that shape learning outcomes in ways that can be advantageous. Specifically, we show that the inclusion of memory constraints actually favors segmentation and that cognitive limitations may themselves play an important role in the word segmentation task.