Sato Y. [(*)], Broz F. [(*)] & Nehaniv C. L. [(*)] (2011). Computational simulation of syllabification and syllable based word-discovery. Accepted at the *International Child Phonology Conference*, York, 16-17 June.

[(*)] Adaptive Systems Research Group, University of Hertfordshire

Abstract

Syllables have long been argued to be a potential unit on which children rely to acquire phonological lexicon, but psycholinguistic evidence has been inconclusive. In view of this, we propose to ask the inverse question: if the children *did* use syllables as an intermediate step, would this help them find words more easily than if solely relying on phonemes? To answer this question we test the performance of the two-step methods of syllabification and syllable-based word discovery by means of computer simulations, and compared them with phoneme-only methods. We take syllabification to be an online task: the learner processes a sequence of phonemes as they come in and find syllables on the fly. As syllables are, ulike words, consistent in structure (C*VC*, where C* represents zero or more consonants), syllabification can be rendered a collection of local tasks: it boils down to finding the inter-syllable boundary. Nevertheless when faced with multiple choices (e.g. when a coda cluster is followed by an onset cluster), the task is far from trivial as such choices can cumulatively grow exponentially. However the evidence was famously shown (Saffran et al. 1996) that children could locate the boundaries quickly apparently with transitional probabilities alone. To simulate this learning process, a procedure for online syllabification has been implemented with a statistical learning component. We test two possibilities for this component: a relatively simple learner based on Maximum Likelihood Estimation and a more sophisticated Bayesian non-parametric learner. We then use the same learning mechanism to then find word boundaries, and compare the results with those with the phoneme-only method. We will report on how the eventual performance compares with a phoneme-only method after a number of training sessions. A difference in the learning curve could suggest that the syllable-based method might be a more plausible model.