Simulated learning of phonotactics via syllabification

Yo Sato, Adaptive Systems Group, University of Hertfordshire

It has been observed that infants are sensitive at an early age to the statistical patterns of phoneme strings they are exposed to and may use it for segmenting continuous speech (Saffran et al. 1996). One of the proposed mechanisms for learning patterns of phoneme concatenation, or phonotactics, include transitional probability between phonemes, or phoneme bigram, i.e. the probability of one phoneme being followed by another (Daland and Pierrehumbert 2010). In this work we propose that the same mechanism may be used for online syllabification, and assess its plausibility by computer simulation. Our proposal is that syllabification and phonotactic learning go hand in hand: the learner uses two kinds of knowledge, rudimentary syllable structure, i.e. C^*VC^* ('*' denotes zero or more occurrences), and the statistical knowledge induced from the data.

The statistical learning method we employ is the Maximum Likelihood Estimation, in which the learner chooses, amongst the possible hypotheses, the one that would make the observed data the most probable. In our case, given a phoneme sequence, the learner chooses, amongst the possible ways to cut it up into syllables, the one with the greatest probability based on the phoneme bigram statistics. Thus the phonotactic learning and syllabification hypotheses are rendered interdependent.

In this simulation, the primary aim of which is modelling a possible manner of phonotactics learning with the minimum prior knowledge, we pay due attention to simulating the learning situation appropriately. We first ensure that the procedure works incrementally in an online manner: the learner processes the phoneme sequence on the fly, although long pauses are considered to delimit utterances. We use child-directed speech data (CHILDES, MacWhinney 2000), where utterances and words are relatively short, so the intractability of the hypthesis space is not so much of a problem. We also test two types of add-on parameters in the simulation, partial (self-)supervision (the utterance consisting of a single monosyllabic word gives guidance to subsequent syllabifications) and phonemic biases (use of sonority scale and articulatory proximity). We evaluated the results against two standards, the correct syllabification (Standard A) and the less stringent, English phonotactic rule compliance (Standard B). The totally unsupervised and unbiased version produced results which, though mostly significantly better than the random baseline (about 30% precision), are rather disappointing in precision at 50-60% against Standard A, but encouraging 60-65% against B. Good results (80-85% against A, nearly 90% against B) are obtained for the supervised and biased version, with intermediate versions (parameters on and off) showing figures inbetween, suggesting the sufficiency of bigrams in combination of these parameters for phonotactics learning.

There remain some elements that prove difficult to learn, which themselves are instructive. The bigrams of frequent syllable-ending and frequent syllable-starting consonants, such as /dð/ or /ŋt/ are hard to rule out without sufficient supervision. Furthermore, a big gap between the performances against Standards A and B mainly comes from the fact that a long series of consonants tends to be divided at some mid-point in it, even where a syllable should end or start with a vowel: e.g. /nɛks-taɪl/ for 'next aisle'. These results can be taken to indicate the possible pre-requisites and tendencies of infants' phonotactic learning, which we discuss at the end of the presentation.