STRAINS AND SYMPTOMS OF THE ‘LITERACY VIRUS’: MODELLING THE EFFECTS OF ORTHOGRAPHIC TRANSPARENCY ON PHONOLOGICAL PROCESSING

Alastair Smith (Max Planck Institute for Psycholinguistics), Padraic Monaghan (Lancaster University) & Falk Huettig (Max Planck Institute for Psycholinguistics).
alastair.smith@mpi.nl

The effect of literacy on phonological processing has been described as analogous to a virus that “infects all speech processing” (Frith, 1998). Behavioural data has shown that exposure to literacy training in alphabetic languages coincides with both qualitative and quantitative improvements in the awareness of the phonological structure of words. Recent brain imaging data indicates that such behaviour may result from a restructuring of phonological processing networks as a consequence of learning to read an alphabetic language. Harm & Seidenberg (1999) present a computational model of reading that provides an explicit description of how such phonological restructuring may occur. They showed that connectionist networks trained to map between English orthographic and phonological representations displayed more componential phonological processing than networks trained only to stably represent the phonological forms of English words. This model and processing level models such as Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005) predict that systems trained on alphabetic languages will develop more componential processing due to the systematic relationships that exist between the letters of the language and corresponding speech sounds.

We trained a model similar to that described in Harm & Seidenberg (1999) to capture explicitly how the effects of literacy training on phonological processing may differ as a consequence of orthographic transparency. We developed two models, a transparent literate model and a non-transparent literate model, both of which were trained on a corpus of 6,188 English words with orthographic and phonological representations. For the transparent model the mapping between orthographic and phonological representations corresponded to the English mappings, however in the corpus used to train the non-transparent literate model orthographic representations were randomly assigned to phonological representations (e.g. Transparent: cake -> keɪk, fort -> fɔːt; Non-transparent: cake -> fɔːt, fort -> keɪk). Thus the non-transparent model had to learn the relationship between the whole word and its pronunciation without recourse to regularities at a finer grain-size. This arrangement ensured that the two models were controlled in terms of the set of inputs and outputs but differed only in the extent to which the mapping was transparent.

We observed that networks trained on a transparent orthography were better at restoring phonetic features and phonemes, reflecting componential phonological processing. However, networks trained on a non-transparent orthography were more likely to restore corrupted phonological segments with legal, coarser linguistic units (e. g. onset, coda). Our results connect with the growing body of work that describe differences in phonological processing between Chinese (a morphosyllabic language) and English literate populations and provides an explicit description of, and predictions for, how differences in orthographic transparency can determine varying strains and symptoms of the ‘literacy virus’.