# Examining strains and symptoms of the 'Literacy Virus': The effects of orthographic transparency on phonological processing in a connectionist model of reading

**Alastair C. Smith (alastair.smith@mpi.nl)**
Max Planck Institute for Psycholinguistics, Wundtlaan 1,
Nijmegen, 6525 XD, The Netherlands

**Padraic Monaghan (p.monaghan@lancaster.ac.uk)**
Department of Psychology, Lancaster University,
Lancaster, LA1 4YF, U.K.

**Falk Huettig (falk.huettig@mpi.nl)**
Max Planck Institute for Psycholinguistics, Wundtlaan 1,
Nijmegen, 6525 XD, The Netherlands

**Abstract**

The effect of literacy on phonological processing has been described in terms of a virus that "infects all speech processing" (Frith, 1998). Empirical data has established that literacy leads to changes to the way in which phonological information is processed. Harm & Seidenberg (1999) demonstrated that a connectionist network trained to map between English orthographic and phonological representations display's more componential phonological processing than a network trained only to stably represent the phonological forms of words. Within this study we use a similar model yet manipulate the transparency of orthographic-to-phonological mappings. We observe that networks trained on a transparent orthography are better at restoring phonetic features and phonemes. However, networks trained on non-transparent orthographies are more likely to restore corrupted phonological segments with legal, coarser linguistic units (e.g. onset, coda). Our study therefore provides an explicit description of how differences in orthographic transparency can lead to varying strains and symptoms of the 'literacy virus'.

**Keywords:** Connectionist Modeling, Reading, Literacy, Spoken word recognition.

## Introduction

There is a well-established link between the acquisition of literacy and changes to the manner in which phonological information is processed. Children's awareness of the phonological structure of words has been shown to improve qualitatively and quantitatively when exposed to literacy training (e.g. Alcock, Ngorosho, Deus, & Jukes, 2010; Hulme et al., 2005). Critically, similar improvements have been recorded in adults who receive literacy training later in life (e.g. Morais, 1979), indicating that changes in phonological processing are consequent upon literacy and not due to extended language exposure or other developmental factors.

Literacy training has also been shown to relate to improvements in categorical perception of speech sounds (Serniclaes et al., 2005). Participants who were literate indicated a sharper boundary in judging syllables along a ba-da continuum, compared to illiterates. Further, effects of orthography on speech perception have been reported widely using varying tasks (Taft et al., 2008; Pattamadilok et al., 2009; Peereman et al., 2009), languages (Ventura et al., 2004; Zeigler et al., 2008) and orthographic manipulations (Orthographic Neighborhood: Zeigler et al., 2003; Orthographic Consistency: Zeigler et al., 2004). Many of these previous studies have investigated effects of literacy on phonological processing by manipulating the properties of the phonological form of the word. In addition, recent evidence from studies of language-mediated visual attention has linked literacy to changes in the granularity of online speech processing without requiring explicit operations on the phonological structure of the word (Huettig, Singh & Mishra, 2011; Smith, Monaghan & Huettig, 2013). In eye-tracking behavioral studies, and computational models of the effects, they found that literate, but not illiterate participants, demonstrated phonological cohort effects on visual search of objects with similar names, but equivalent levels of co-activation of visual search of objects with similar meanings. These studies suggest a broader impact of literacy-related changes to phonological processing that extends to language processes that interface with visual systems.

Brain imaging data provides further support for changes to phonological processing as a consequence of literacy training, suggesting both an influence of activation in orthographic processing regions when processing speech (Dehaene et al., 2010) and a restructuring of phonological processing regions (Perre et al., 2009, 2011; Pattamadilok et al., 2010).

Computational modelling studies offer a means by which mechanisms that drive such effects can be isolated. Harm & Seidenberg (1999) captured explicitly the emergent effects of learning orthographic-to-phonological mappings on phonological representations in a connectionist model of reading. They trained a model that learned to stably represent the phonological forms of a large set of English words (the “illiterate” model), and compared this to a model that also learned to map orthographic forms onto the same phonological representations (a “literate” model). They found that the literate model represented phonological forms of words with a higher degree of componential structure than the illiterate model. For instance, the model was tested on its ability to reconstruct individual phonological features within phonemes, and the literate model was able to do this more accurately than the illiterate model, akin to the increased sensitivity of the categorical boundary between similar phonemes in literate participants (Serniclaes et al., 2005). Similarly, the model was also tested on its ability to reconstruct single phonemes that were degraded in the phonological input, the literate model performed with a higher degree of accuracy, indicating greater ability to construct individual phonemes in the spoken form of the word.

The model was analyzed in terms of the connection weights between different regions of the words’ phonological representations. The literate model resulted in stronger connections from each phoneme to itself, and slightly weaker connections to other phonemes within the word. This suggested that the model’s granularity of processing was affected by orthographic training – the literate model processed more phoneme-by-phoneme, whereas there was less componential processing in the illiterate model. Hence, effects of experience of orthography mapping onto phonology qualitatively changes the way in which the phonological forms of words are represented, even when orthography is not directly implicated in the test.

## Effects of different orthographies on speech processing

It is therefore understandable that the analogy of literacy as a “virus” that “infects all speech processing…” (Frith, 1998) has resonated so strongly with researchers within this field. However, evidence from cross-linguistic studies suggests that this virus may have many strains, each giving rise to a different set of symptoms. One factor that seems to play a key role in determining the precise effect that literacy has on speech processing is the orthographic transparency of a language.

Orthographic systems differ radically in how they map onto the speech sounds of the language. English represents a semi-transparent language as individual letters provide a reasonable approximation to the word’s pronunciation, though there are many irregularities present in the set of letter sound mappings (e.g. PINT. YACHT). Languages such as Serbian, however, demonstrate a very shallow or transparent orthography, as each letter corresponds with perfect regularity to a single speech sound. Chinese, at the other extreme, provides an example of a deep orthography (a morphosyllabic language). Chinese characters are each pronounced as syllables, and are composed of two portions: a phonological and a semantic radical. The phonological radical provides a hint about the pronunciation of the syllable, but this is modulated by the semantic radical. The effect is that each character’s pronunciation has to be learned individually, and there is almost no componential structure at the syllable level in terms of relationships between orthography and phonology.

Much of the evidence reported earlier in this section describes effects of literacy observed in transparent or semi-transparent languages. However, recently the body of evidence examining the effects of literacy training on phonological processing in less transparent languages such as Chinese has grown substantially (e.g. Tan et al., 2005; Shu et al., 2008; Brennan et al., 2013).

Shu et al.’s (2008) study supports predictions that changes to the granularity of phonological processing as a consequence of literacy training are modulated by orthographic transparency (Ziegler and Goswami, 2005). They measured the performance on syllable and rime awareness in Chinese children aged 3 to 6 years, and found that this improved gradually with age. However, awareness of phoneme onset and tone awareness only improved above chance when exposed to additional instruction which made explicit the relationship between orthography and phonemes (by training additionally in Pinyin).

Further evidence can be found in Brennan et al.’s (2013) study that contrasted the effects of literacy training in a transparent language (English) with training in a non-transparent language (Chinese), on phonological processing. They compared neural activity in Chinese and English children (8-12 years of age) and adults when making rhyming judgments to pairs of spoken words. Differences over the course of development in activity in brain regions associated with phonological processing (superior temporal gyrus) were only observed between English speaking adults and children, further this difference was more pronounced for words with conflicting orthography (e.g. PINT – MINT). These findings provide evidence that phonological processing regions can be effected by literacy training and that orthographic transparency can influence the nature of this effect.

Together this evidence supports the position that the consequences of literacy training on phonological processing differ as a consequence of orthographic transparency. The current study aims to capture these differences explicitly in a computational model based on Harm & Seidenberg (1999). Within this model we manipulated the transparency of the orthography on which the model is trained. We predicted that the type of literacy training would have an effect on the componential structure of phonological processing in the speech processing of the model. In particular, we examined the effect of transparent versus nontransparent literacy training on the model’s ability to reconstruct phonological features of phonemes, and also the model’s ability to reconstruct individual phonemes in words.

## Method

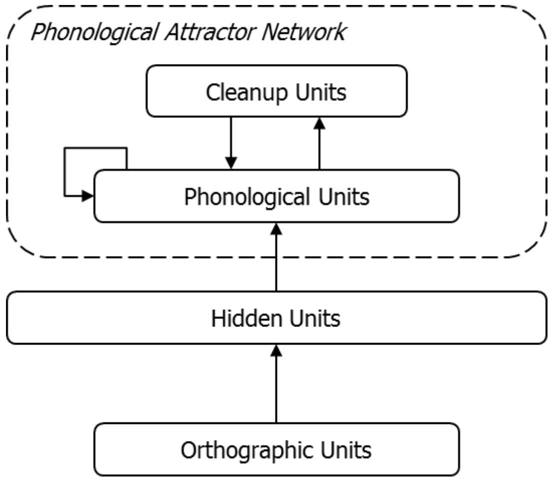


Figure 1: Network Architecture (Harm & Seidenberg, 1999).

## Architecture

The model used within this study was based on Harm & Seidenberg's (1999) connectionist model of reading (see Figure 1).

The model contained a phonological attractor network which was trained to develop stable representations of spoken forms of words. This network consisted of a phonological layer of 200 units which were fully self-connected and fully connected to and from a set of 100 clean up units. The phonological layer comprised 8 phoneme slots of 25 units representing the presence or absence of a specific phonological feature. Speech input was simulated as input activations to the phonological layer, and phonological production was simulated as output activation across the same set of units in the phonological layer.

Written input was presented at an orthographic layer consisting of 260 units. The orthographic layer was composed of 10 orthographic slots, each consisting of 26 units, with each unit corresponding to a distinct letter. The orthographic layer was fully connected to a hidden layer of 1000 units, which was fully connected to the phonological layer.

The model differed from that presented in Harm & Seidenberg (1999) as additional resources were provided in the form of extra processing units in the cleanup layer and hidden layer to enable the model to learn the complex non-transparent mapping between orthography and phonology. Pilot studies demonstrated that the model failed to learn accurate and stable representations without these additional resources.

## Representations

A corpus of 6,188 monosyllabic English words, each between 1-10 letters in length was used to train and test the transparent orthography model**.** Each word comprised an orthographic and phonological representation. Orthographic representations were 260 unit binary vectors, with 26 units encoding each letter slot. A single unit indicated the presence or absence of a given letter in each letter slot. Phonological representations were 200 unit binary vectors, with 25 units encoding each of 8 phoneme slots. Phonemes were defined in terms of 25 phonological features as implemented in Harm & Seidenberg (2004). Orthographic and phonological representations were vowel centered to control for variation in vowel positions.

The non-transparent orthography model was trained using the same set of orthographic and phonological representations, but with orthographic representations randomly reassigned to phonological representations (non-transparent corpus). Thus, the model had to learn the relationship between the whole word and its pronunciation without recourse to regularities at a finer grain level, for example between particular letters and phonemes, as was available for the transparent orthography model. The two orthographies were therefore controlled in terms of the set of inputs and outputs, but what differed was the extent to which the mapping between them was componential.

## Training

All connections within the network were initialized with random weights ($\mu = 0$, $\sigma = 1$), except for the self-connections in the phonological layer which passed activation back to the same unit. The weights on these connections were fixed to 0.75 and were therefore not adjusted during training to ensure that input to the phonological network decayed over time, thus forcing the phonological network to developed phonological attractors.

Training of the model consisted of two stages, similar to Harm and Seidenberg's (1999) model of reading: a pre-literate stage, which simulated developing experience with listening and speaking the phonological forms of words, and a literate stage, simulating learning to map written onto spoken forms of words.

The pre-literate stage, the first 1 million training trials, involved training the phonological attractor to maintain a stable phonological representation at the phonological layer over time. For each word, the phonological representation was clamped at the phonological layer from time step 0 until time step 2. The model then cycled activity for a further 5 time steps (ts). At time steps 5-7 the target phonological representation was presented to the phonological layer, error was computed as the Euclidean distance between actual and target activation at the phonological layer and then error was back-propagated and connection weights updated within the phonological attractor network.

The second stage of model training included literacy training trials, in which the model was required to map from orthography to phonology. The orthographic representation of a word was clamped to the orthographic layer for the entire training trial. At time steps 5-7 the word's phonological representation was presented to the phonological layer as a target, with error (distance between target and actual phonological layer activations) back-propagated and weights updated throughout the entire network.

The model was trained on a further 9 million training trials with orthography to phonology (reading trials: p = 0.8) and phonology to phonology trials (listening trials: p = 0.2) randomly interleaved, although reading trials were four times more likely to occur than listening trials.

Words were sampled randomly according to their log-compressed frequency, with minimum frequency set at 0.05. Four simulations each were trained for the transparent and the non-transparent models, Each simulation run differed in terms of the random initial starting weights, and the random selection of words. Each instantiation of the non-transparent model was trained on a different corpus each with a new random reassignment of orthography-phonology mappings. Networks were trained with a learning rate of 0.05.

## Results

### Pre-Test

Once trained networks were tested on their ability to perform each training task for each word in the training corpus to ensure sufficient training had been provided for all models to learn the requisite mappings. To test listening performance the phoneme closest in euclidean distance to the activation of units in each phoneme slot at the end of a listening trial was computed, providing the phonological representation produced by the model. This was compared to the phonological representation of the word presented at the beginning of the test trial. If all phonemes within the two representations matched, then the trial was registered as successful. Networks trained on a transparent corpus successfully performed the listening task for all words in the corpus. Networks trained on a non-transparent corpus performed the listening task successfully for 99.9% (SD = 0.0003) of words in the corpus. A similar procedure was applied to examine performance on reading tasks. A reading trial following the training procedure was conducted for each word in the corpus. The phoneme closest to the activation of units in each phoneme slot at the end of the reading trial was computed. If this matched the phonological representation of the word presented to the orthographic layer then the model was registered as having correctly read the word. Networks trained on a transparent corpus read all words within the training corpus successfully. Networks trained on a non-transparent corpus successfully read 98.8% (SD = 0.0009) of words within the training corpus.

To examine differences in the phonological processing of networks trained on a transparent corpus and those trained on a non-transparent corpus we tested networks on their ability to reconstruct phonological features of phonemes, and to reconstruct individual phonemes in words when corrupted by noise.

### Pattern Completion

Pattern completion tests examined the model's ability to restore each active phonological feature, in each phoneme, of each word. The phonological representation of a given word was clamped to the phonological layer for ts 0-2 yet with a single active phonological feature switched off. The network was then free to cycle until ts 7 at which point the activity of the unit representing the corrupted phonological feature was recorded. The distance between its level of activation at the end of the test trial and its correct value (i.e. 1) was recorded. Figure 2 presents the mean sum squared error calculated over all features, phonemes and words for networks trained on a transparent corpus and networks trained on a non-transparent corpus. Networks trained on a transparent corpus re-activated corrupted phonological features more accurately than networks trained on non-transparent corpora [$\mu = -0.013$, $\sigma = 0.005$, $t(3) = -4.747$, $p = 0.018$].

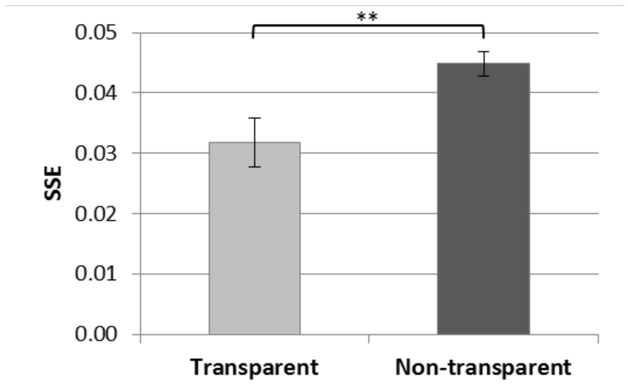


Figure 2: Mean sum squared error (SSE) calculated between correct feature and restored feature

### Segment Restoration

Networks were also tested on their ability to restore entire phoneme segments. On segment restoration trials a single phoneme was replaced with random noise. Features within the corrupted segment were assigned values between [0 – 0.1] with p = 0.8 and values between [0 – 0.5] with p = 0.2. Networks were tested on their ability to restore each phoneme within each word. During ts 0-2 the corrupted phonological representation was clamped to the phonological layer, the network was then allowed to cycle freely until ts 7. At ts 7 activation in the phonological layer was recorded. The nearest neighbor phoneme (euclidean distance) was identified given the activation recorded in each phoneme slot.

**Phoneme Restoration** For each network, the average (euclidean) distance between the pattern of activation in the phoneme slot corresponding to the location of the corrupted phoneme and its nearest neighbor phoneme was calculated. This provides a measure of how well a network's phonological attractor component is able to restore a corrupted phoneme segment. Figure 3 displays the mean performance of networks trained on a transparent corpus and networks trained on a non-transparent corpus. Networks trained on a transparent corpus restored corrupted phonemes marginally better than those trained on a non-transparent corpus [$\mu = -0.045$, $\sigma = 0.029$, $t(3) = -3.054$, $p = 0.055$].

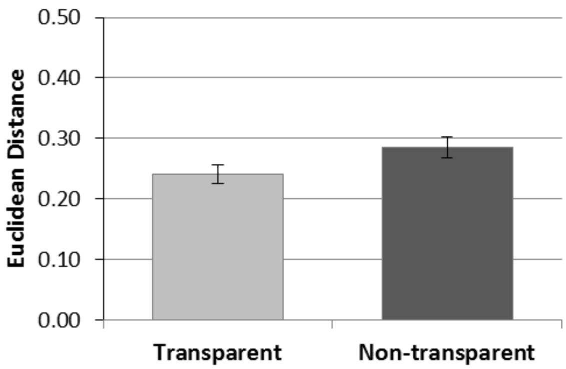


Figure 3: Mean Euclidean distance calculated between nearest neighbor phoneme and restored phoneme

**Validity of Segment Restorations** Networks were also tested on the validity of the segment restored. The restored sub-syllabic segment (onset, vowel or coda) in which the corrupted phoneme was embedded was examined and recorded as a legal restoration if the segment existed in the training corpus. Figure 4 displays the proportion of illegal restorations made by networks trained on a transparent corpus and networks trained on a non-transparent corpus. Networks trained on a transparent corpus made more illegal segment restorations than those trained on a non-transparent corpus [$\mu = -0.102$, $\sigma = 0.012$, $t(3) = -16.492$, $p < 0.001$].

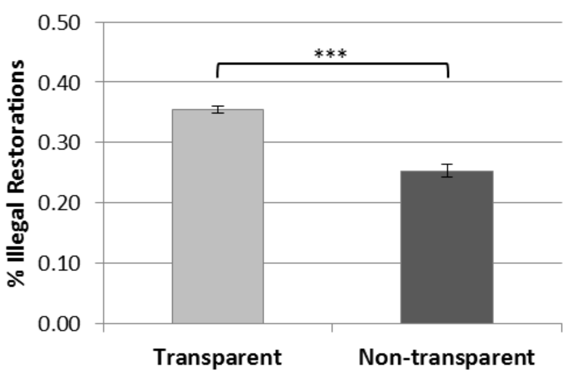


Figure 4: Proportion of illegal segment restorations

## Discussion

Harm & Seidenberg (1999) demonstrated that literate networks trained on a transparent orthography restored both phonetic features and phonological segments (onset, vowel, coda) more accurately than illiterate models. Our simulations show that networks trained on a transparent orthography also outperform models trained on a non-transparent orthography in restoring phonetic features and phonemes. However, models trained on a non-transparent orthography were more likely than models trained on a transparent corpus to restore phonemes to form valid segments at a coarser grain size (onset, vowel, coda).

These findings suggest that transparent models processing of phonological information is more componential (at the phoneme level), and subsequently models trained on a non-transparent orthography are more likely to process phonological information more coarsely. Importantly each gains contrasting benefits in phonological processing as a consequence of the mappings on which they are trained. This is an important extension of the findings of Harm & Seidenberg (1999). In our simulations we have shown that literacy training has an effect on phonological processing both in the case of transparent and non-transparent orthographies and further that transparency defines the nature of the effect.

Both models benefit from the regularities of the orthographic training. In the transparent model increases in finer grain phonological processing arises as a consequence of the network learning from consistencies at the phoneme level between orthography and phonology. It is pushed towards processing words phoneme by phoneme. Therefore, when noise was applied to a phoneme segment the transparent model was more likely to restore the distorted segment with a segment that was closer to a valid phoneme.

In contrast, in the non-transparent model, correspondence between orthography and phonology only exists at the word level. The model is therefore more likely to develop phonological attractors at a coarser grain size. For this reason corrupted segments are increasingly pushed toward valid segments at the level of onsets, vowels or codas.

The finding that transparent networks processing of phonological information is more componential than non-transparent networks connects with established empirical findings such as the results of Shu et al., (2008) in which phoneme awareness increased in Chinese children when exposed to training on a transparent orthographic system. Further this study’s findings provide additional support for theoretical models that argue that the granularity of phonological processing developed is dependent on the transparency of the orthography on which an individual is trained (Zeigler & Goswami, 2005).

Our results also provide predictions for future empirical studies. Transparent networks outperformed non-transparent networks on phonetic feature restoration. This predicts that individuals trained on deep orthographic systems such as Chinese will perform worse than individuals trained on shallower orthographies such as English in identifying fine phonetic contrasts. A second prediction follows from the finding that non-transparent models were more likely to generate valid restorations of coarser grain units (e.g. onset, coda). Should this finding represent the development of stronger phonological attractors at onset, coda and word levels, then individuals trained on non-transparent orthographies should be more resilient to corrupted speech at these grain sizes than individuals trained on transparent orthographies.

Studies that examine the impact of literacy training on phonological processing regions within the brain have largely focused on effects in individuals trained on transparent or semi-transparent languages. Our findings suggest that phonological restructuring is likely to occur for both populations trained on transparent and non-transparent languages yet the nature of this restructuring and therefore its neural and behavioral consequences will differ.

Previous modelling of the effects of literacy on phonological processing provided an explicit description of how literacy leads to improved phonological processing. Within the current paper we extend this previous work by describing how characteristics of the language in which training is received can have significant consequences for the nature of the effects observed.

Behavioral and brain imaging studies are beginning to define various strains and symptoms of the literacy virus. Computational modelling studies such as the one presented here provide a means of identifying the factors that categorize these strains. Such models can then be used to raise predictions of their resulting symptoms that can then be tested empirically.

## References


Alcock, K. J., Ngorosho, D., Deus, C., & Jukes, M. C. H. (2010). We don't have language at our house: disentangling the relationship between phonological awareness, schooling, and literacy. *British Journal of Educational Psychology, 80(1)*, 55-76.

Brennan, C., Cao, F., Pedroarena-Leal, N., McNorgan, C. and Booth, J. R. (2013), Reading acquisition reorganizes the phonological awareness network only in alphabetic writing systems. *Human Brain Mapping, 34*: 3354–3368.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science, 330(6009)*, 1359-1364.

U. Frith (1998) Literally changing the brain. *Brain, 121,* 1011–1012

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological review, 106(3)*, 491.

Huettig, F., Singh, N., & Mishra, R. K. (2011). Language-mediated visual orienting behavior in low and high literates. *Frontiers in psychology, 2*.

Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles and Coltheart. *Scientific Studies of Reading, 9(4)*, 351-365.

Loureiro, C. D. S., Willadino Braga, L., Souza, L. D. N., Queiroz, E., & Dellatolas, G. (2004). Degree of illiteracy and phonological and metaphonological skills in unschooled adults. *Brain and language, 89(3)*, 499-502.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7(4)*, 323-331.

Pattamadilok, C., Knierim, I. N., Duncan, K. J. K., & Devlin, J. T. (2010). How does learning to read affect speech perception?. *The Journal of Neuroscience*, 30(25), 8435-8444.

Pattamadilok C., Perre L., Dufau S., Ziegler J. C. (2009). On-line orthographic influences on spoken language in a semantic task. *Journal of Cognitive Neuroscience, 21*, 169–179.

Perre, L., Bertrand, D., & Ziegler, J. C. (2011). Literacy affects spoken language in a non-linguistic task: an ERP study. *Frontiers in psychology, 2*.

Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring?. *Brain research, 1275*, 73-80.

Peereman R., Dufour S., Burt J. S. (2009). Orthographic influences in spoken word recognition: the consistency effect in semantic and gender categorization tasks. *Psychonomic Bulletin & Review, 16*, 363–368.

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition, 98(2)*, B35-B44.

Shu, H., Peng, H., & McBride-Chang, C. (2008). Phonological awareness in young Chinese children. *Developmental Science, 11(1)*, 171-181.

Smith, A., Monaghan, P., & Huettig, F. (2013). Modelling the effects of formal literacy training on language-mediated visual attention. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 3420-3425). Austin, TX.

Taft M., Castles A., Davis C., Lazendic G., Nguyen-Hoan M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory & Language, 58,* 366–379.

Tan, L. H., Laird, A. R., Li, K., & Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human brain mapping, 25(1)*, 83-91.

Ventura P., Morais J., Pattamadilok C., Kolinsky R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes, 19*, 57–95.

Ziegler J. C., Ferrand L., Montant M. (2004). Visual phonology: the effects of orthographic consistency on different auditory word recognition tasks. *Memory & Cognition 32*, 732–741

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin, 131(1)*, 3.

Ziegler J. C., Muneaux M., Grainger J. (2003). Neighborhood effects in auditory word recognition: phonological competition and orthographic facilitation. *Journal of Memory & Language, 48*, 779–793.

Ziegler J. C., Petrova A., Ferrand L. (2008). Feedback consistency effects in visual and auditory word recognition: where do we stand after more than a decade? *Journal of Experimental Psychology Learning, Memory, and Cognition. 34,* 643–661