A "Triangle Model" of Chinese Reading

Jianfeng Yang (goldswordy@gmail.com)
State Key Laboratory of Cognitive Neuroscience and Learning,
Beijing Normal University, China

Jason D Zevin (jdz2001@med.cornell.edu)
Sackler Institute, Weill Cornell Medical College, USA

Hua Shu (shuh@bnu.edu.cn)
State Key Laboratory of Cognitive Neuroscience and Learning,
Beijing Normal University, China

Bruce D. McCandliss (bdm2001@med.cornell.edu)
Sackler Institute, Weill Cornell Medical College, USA

Ping Li (pli@richmond.edu)
Department of Psychology, University of Richmond, USA

Abstract

In this paper we present two connectionist models of reading, using a parallel distributed processing framework that has been applied to English, to examine the extent to which such models can also account for developmental performance in Chinese. Simulation 1 was trained to map from orthography to phonology for a large corpus of stimuli, and simulates the frequency, regularity, consistency effects and interactions among them. Simulation 2 was trained to map among spelling, meaning, and sound for a smaller set of items, and captures basic effects of orthographic transparency and family size. Although the computational models used here are very similar to the English models, a very different developmental pattern emerged, such that mappings from orthography to semantics were learned more rapidly than mappings from orthography to phonology. The results show how qualitative differences in the development of reading skill across writing systems can arise from the functioning of the same general learning mechanisms.

Introduction

Over the past three decades, computational models have become increasingly sophisticated in accounting for a broad range of phenomena and specifying the hypothetical mechanisms underlying skilled reading and its acquisition. However, because the vast majority of this work has been done in English, questions remain about the generality of these models. For example, the DRC model of reading English contains specific assumptions (e.g., about how conflicting rules at different grapheme sizes are handled) that are unlikely to generalize to transparent languages (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Furthermore, it is not clear how the architecture of that model would handle a language in which there are no regularities that can readily be identified as "rules," such as Chinese. On the other hand, at least one computational model of Chinese reading acquisition involves a set of highly language-specific processes for identifying specific constituents of characters (Xing, Shu, & Li, 2002, 2004).

Our approach is informed by the "triangle" model of reading, developed by Seidenberg & McClelland (1989; see also Plaut, McClelland, Seidenberg, & Patterson, 1996; Harm & Seidenberg, 1999; Harm & Seidenberg, 2004). This series of models treats reading as a constraint satisfaction problem, and provides the basis for a consideration of how insights from models developed to explain reading in English can be brought to bear on a very different reading system – Chinese. Both English and Chinese writing systems contain multiple probabilistic cues to pronunciation that contribute to the learning of spelling-to-sound mappings. Whereas in English, these constraints are relatively few, and highly consistent – so that they are often roughly characterized as mappings from individual letters to individual speech sounds – in Chinese, mappings from spelling to sound tend to be less consistent, and depend on a much larger number of orthographic constituents called "phonetic radicals", part of a character that makes a basic morpheme in this language. In Simulation 1, we examine the effects of regularity, consistency, and frequency in a version of the triangle model adapted to Chinese characters.

An additional factor that distinguishes Chinese from English (and most other writing systems) is that Chinese characters typically contain a "semantic radical" that provides some probabilistic information that aids in the translation from orthography to semantics. In alphabetic languages such as English, monosyllables rarely contain sub-lexical information about meaning. An exception to this is inflected forms, e.g., "walk," "walks," and "walked," which are semantically related in predictable ways. Interestingly, the triangle model picks up on these regularities in English (Harm & Seidenberg, 2004), because the associative learning mechanism used in the triangle model is highly sensitive to similarity. Thus, in a language like Chinese, wherein many characters contain probabilistic cues to meaning, we would

expect effects analogous to consistency effects in mapping from spelling-to-sound to arise in spelling-to-meaning mappings.

Furthermore, because these models are inherently developmental, they provide an opportunity to examine the influence of the statistical regularities in the writing system on learning to read. A number of studies have shown that spelling-to-meaning consistency (or "transparency") plays a role in children's reading from very early on in Chinese (McBride-Chang, Shu, Zhou, Wat, & Wagner, 2003; Shu & Anderson, 1997), and that a factor called "morphological awareness" – essentially the ability to generalize based on sub-lexical semantics – is a strong predictor of learning to read (McBride-Chang et al., 2005; Shu & Anderson, 1997; Shu, Anderson, & Zhang, 1995). In Simulation 2, we explore these phenomena in a model that includes mappings among orthography, phonology, and semantics.

A Brief Primer on Chinese Reading

In modern Chinese, nearly 90% of characters are "phonograms," which consist of two components (radicals). The semantic radical provides information about the meaning of the character, and a phonetic radical indicates the character's pronunciation. These radicals often also appear by themselves as "simple words" (i.e., mono-morphemic words). For example, the word ("皇"huang1) also appears in a "family" of words ("凰, 惶, 煌, 篁"). When a phonogram's pronunciation matches its phonetic radical, it is called "regular." Families in which all the phonograms match in pronunciation are called "consistent."

For most families, however, the pronunciation of at least some of the family members differs from the pronunciation of the phonetic, such as "台/tai/, 殆/dai/, 怡/yi/, 抬/tai/, 答 /chi/". All phonograms in such a family are called "inconsistent," and those particular phonograms that do not match the phonetic are called "irregular."

Research on Chinese has revealed both regularity and consistency effects that interact with frequency. Seidenberg (1985) found that phonograms whose pronunciations differed from the pronunciation of their phonetic radical were named more slowly and less accurately than "regular" characters. Consistency effects have also been observed in single character naming, such that items from inconsistent families are named more slowly than consistent characters (Fang, Horng, & Tzeng, 1986). A number of studies have now compared three types of stimuli: regular-consistent (R-C), regular-inconsistent (R-I) and irregular-inconsistent (I-I), (Hue, 1992; Peng & Yang, 1997; Peng, Yang, & Chen, 1994). These studies found both a consistency effect, and a regularity effect, such that I-I items were read more slowly than either R-C or R-I items – a finding which may provide a challenge to a constraint-based model.

Semantic radicals also appear in "families" that vary both in the number of items in which they occur ("family size") and in the consistency with which they map on to a particular meaning ("transparency"). For example, the radical for water (氵) appears in many phonograms that are related to water (湖, *lake,*河, *river,* 渴, *thirsty,* 游, *swim*),

which are transparent, but it can also appear in phonograms that are not (法, *law,* 冶, *negotiate*), which are opaque. This is a unique feature of the Chinese writing system – it encodes probabilistic information about semantics at the sub-lexical level. In Simulation 2, we will consider how this property of the writing system influences both skilled reading and the acquisition of reading ability.

For more detailed introductions to characters, reading, and reading acquisition in Chinese, consult Honorof and Feldman (2006), Perfetti and Liu (2006), Shu and Wu (2006), and other chapters in Li, Tan, Bates, and Tzeng (2006).

Simulation 1: Mapping from Orthography to Phonology

Architecture

The architecture of the reading model is similar to a specific model previously implemented in English to account for mappings between orthography and phonology model (Harm & Seidenberg, 1999). The input layer consists of 270 orthographic units, fully connected to an intermediate level of 200 hidden units, which in turn were fully connected to output representation, which was composed of 92 phonological units and 50 cleanup units to create a phonological attractor network. The phonological layer was also fully connected to itself.
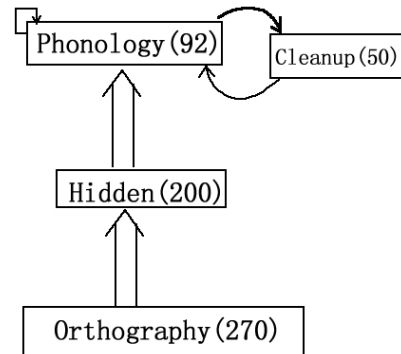

Figure 1: Architecture of Model 1

The orthographic representation is based on a linguistic description of Chinese orthography including radicals, number of strokes and radical position (for details see Xing et al., 2004). The phonological representation includes five slots: one onset slot, three rime slots, and a fifth slot for tone. The rime was divided into a "medial," usually consisting of a glide or approximant, a "nucleus," which is always a vowel, and a "coda" which can either be a nasal (/n/ or /N/) or the second vowel of a diphthong. This slot system captures similarities such as the fact that /tSwaN3/ ("窗") and /ba1/ ("八") share a nuclear vowel, despite having very different syllable structures. Each phoneme was represented using a vector of 22 real-valued units, each of which corresponds to a phonetic feature. Including 4 units for tone, there are in total 92 units for each Chinese syllable representation.

Training

A set of 4468 items from the Modern Chinese Frequency Dictionary (1986) was used to train the model (one character was eliminated because there is no orthographic representation for characters with more than 7 radicals). Frequency counts from the corpus were transformed into a probability of presentation by a square root compression (Plaut, McClelland, Seidenberg, & Patterson, 1996).

Following Harm & Seidenberg (1999), we first pretrained the phonological attractor model, and then trained the full reading model on the mapping from orthography to phonology. The continuous recurrent back-propagation algorithm (Pearlmutter, 1995) was used, with online learning. A learning rate of 0.005 and momentum of 0.9 were used. On each trial, a character was selected and the orthographic units were clamped with the pattern corresponding to the spelling of the word for 12 time ticks. Error computed at the phonological layer was computed from 5-12 time ticks and back propagated to update the connection weights. The model was trained for 3 million trials.

Testing

120 phonograms were used to test the model's performance. Three types of phonograms were selected: Regular Consistent (R-C), Regular Inconsistent (R-I) and Irregular Inconsistent (I-I) items. For each type, we selected 20 high frequency (565 per million) and 20 low frequency (10 per million) items. The number of strokes and radicals was matched across conditions. Naming accuracy and sum squared error were computed to test the model's performance. The model's output was determined by a winner-take-all mechanism based on the Euclidean distance between the model's output and all possible phonemes in the coding scheme. The sum squared error (SSE) was computed from the model's output at the last time tick by taking the square of the difference between the model's phonological output for each unit and the target output.

Results

After 3 millions trials of training, the model can name 86.92% items in the training set accurately. Most of error items (86.3%) are very low frequency characters (no more than 10 per million) and for those items, adults typically have difficulty naming. Out of 120 testing items, 118 (98.33%) items were correctly named. Two low frequency irregular-inconsistent items were named incorrectly. In training, high frequency items were learned most rapidly, and show no effect of regularity or consistency. Among low-frequency items, R-C items were learned faster than R-I items, which were in turn learned faster than I-I items (Figure 2).
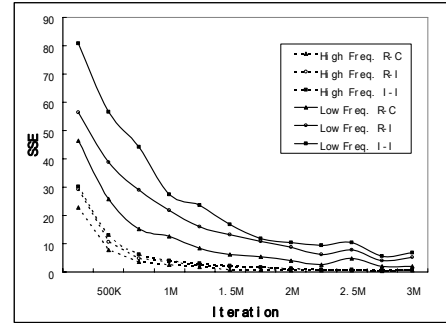

Figure 2: Pattern of SSE over 3 million training trials (Simulation 1)

Analysis of the model at the end of training reveals an interaction between frequency and type typical of the human data (Figure 3, top). High frequency items have a lower SSE than low frequency items, F (1,114) =35.65, p<0.001. There is also a significant effect of the interaction between frequency and type, F (2,114) =3.93, p=0.022. Whereas there was no effect of type for high frequency items, F (2, 115) <1, low-frequency items demonstrated significant type effects, F (2,115) =6.74, p=0.002. R-C items had a lower SSE than R-I items, and R-I items were in turn read more easily than I-I items.
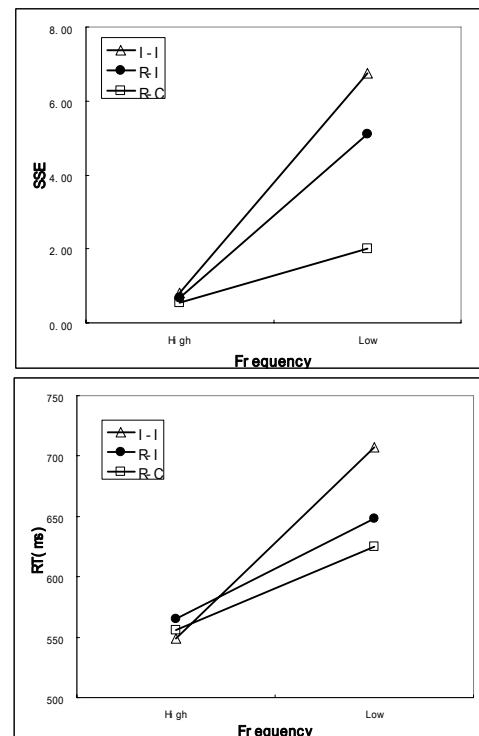

Figure 3: Effects of frequency and type: Simulation 1(top) and adult experiment (down)

Adult Experiment: Naming Latencies

We tested 24 graduate students from Beijing Normal University in a naming task with the same 120 items, using the DMDX software. In each trial, a fixation cross "+" was presented for 300ms, followed by a 300ms blank and then a Chinese character for 400ms presented in black on white

background. An Inter-trail Intervals of 2000ms separated each trial. The results are consistent with previous studies of the impact of type (R-C, R-I, I-I) on naming, and with the predictions of the model, as shown in Figure 3(lower panel). A main effect of frequency was observed, $F_{(1, 19)} = 43.84$, $p<0.001$), as well as an interaction between frequency and type, $F_{(1, 19)} = 12.24$, $p<0.01$.

Discussion

Two critical results emerged from the orthography to phonology model. First, we demonstrate that a model with essentially the same architecture and learning rules used to study reading in English can in fact learn to map from spelling to sound for over 4000 Chinese characters. Previous models (Chen & Peng, 1994; Perfetti, Liu, & Tan, 2005; Perfetti & Liu, 2006; Xing, Shu, & Li, 2004) of Chinese reading have used much smaller training corpora, and did not address consistency effects (Chen & Peng, 1994; Hsiao & Shillcock, 2004, 2005). Our model successfully simulates a central result in the study of Chinese character reading: the interaction of regularity and consistency with frequency. Critically, this was done using items that produce the same pattern of effects in human subjects.

Although the model does not encode any rules, nor indeed does it treat phonetic radicals as at all "special," it does produce regularity and consistency effects, demonstrating that these effects do not depend on literal representations of rules. Finally, the developmental trajectory of the effects of regularity and consistency are correctly simulated. In both humans and the model, both regularity and consistency effects are observed throughout development (Yang & Peng, 1997).

Simulation 2: Triangle reading model

A major feature of the Chinese orthography is that single characters encode probabilistic cues to meaning. In Simulation 2, we add a semantic network to the model and examine the development of mappings among spelling, meaning, and sound.

Architecture

We scaled down the architecture of Simulation 1 (because of the smaller training set, see "Training"), and added a semantic attractor layer, which had full feed-forward connections from orthography (mediated by 100 hidden units), bidirectional connections with phonology (mediated by 100 hidden units) and a cleanup layer of 50 units. Semantic layer consists of 246 units.

Training

The model was trained on 103 phonograms, selected from 6 semantic radical families (3 large and 3 small families), including both transparent and opaque phonograms. The same orthographic and phonological representations were used as in Simulation 1. For semantic representations, we

took the hierarchical feature trees from HowNet (Dong, 2000; Liu & Li, 2002), a Chinese project similar to WordNet (Fellbaum, 1998; Miller, 1990) and "flattened" them into vectors (following Harm & Seidenberg, 2004). Both the semantic and phonological layers were pretrained (both with 100K trials) to simulate early experience, first separately, and then in 400K trails of "speaking" (mapping semantics to phonology) and 400K trails of "hearing" (mapping phonology to semantics) tasks before being trained on the main task of reading aloud. When the reading model was trained, the weights in pretrained task were frozen. Items were presented to the network according to the same online learning scheme as before with the same frequency distributions. Error signals were provided for both the phonological and semantic representations of each item on each trial.

Testing

After 150K trials, testing was carried out as for Simulation 1, with the addition that the "output" of the model was read from both phonology and semantics. From 103 items, we selected 50 items (25 transparent and 25 opaque items) to test the semantic transparency effect. We also selected 46 items (23 from large families and 23 from small families) to test the family size effect. Semantic transparency was defined according to the Elementary School Textbooks Corpus (Hua Shu, Chen, Anderson, Wu, & Xuan, 2003).
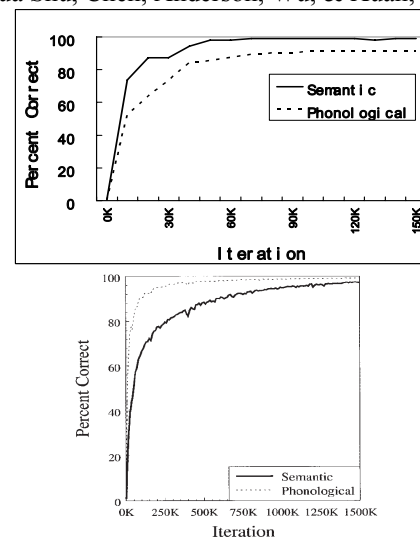

Figure 4: Learning in Simulation 2(top) and in the Harm & Seidenberg (2004) model (bottom).

Results

Figure 4 shows the model's accuracy over time for mapping of orthography onto the semantic and phonological layers. Of particular interest is the comparison of the Chinese model (Figure 4, top) to the English model (Figure 4, bottom): whereas English monosyllables contain very little sub-lexical semantic information, Chinese characters contain probabilistic cues to meaning. As a result, spelling to meaning is learned more rapidly than spelling to sound.
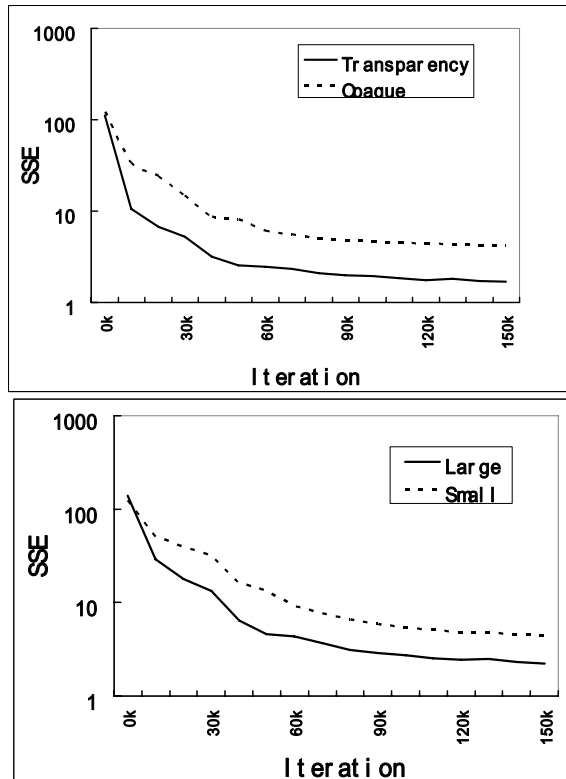
Figure 5: Semantic transparency and Family size effect

Furthermore, as shown in Figure 5, the speed of learning – and efficiency of skilled processing – for this mapping is influenced by both transparency and family size. Items with highly consistent mappings between spelling and meaning are learned more rapidly than items with inconsistent mappings, and items within larger "families" are learned more rapidly than those that share structure with few other items in the training set.

Discussion

Our preliminary results from this model concern the development of mapping from spelling to sound and spelling to meaning over time. Whereas in English, the development of spelling to sound occurs earlier and more rapidly than the development of spelling to meaning, in Chinese, this pattern is reversed, as mappings from spelling to meaning are learned more readily than spelling-to-sound mappings. This cross-linguistic difference, we hypothesize, is due to the language-specific properties of the English versus Chinese orthography-to-meaning relationships: whereas the Chinese orthography contains sub-lexical units that prompt lexical meaning, the English orthography contains no such information.

General Discussion

In this study we present a large-scale connectionist model of Chinese reading based on the triangle model that has been successfully applied to account for English reading. Simulation 1 was trained to map orthography to phonology based on a large corpus, in which effects of frequency, regularity, consistency and their interactions were modeled. Simulation 2 was trained to map spelling, meaning, and sound for a smaller set of items, in which effects of orthographic transparency and family size were modeled. The model can explain a number of interesting empirical phenomena in the acquisition and skilled use of reading ability in Chinese. Furthermore, the same basic computational principles can explain effects in both the mapping from spelling to sound and the mapping from spelling to meaning. These results suggest that the differential development of reading skills across languages may be driven by the statistical regularities particular to each writing system. They demonstrate how qualitative differences in the development of reading skills can arise from the functioning of the same general learning mechanisms.

References

Chen, Y., & Peng, D. (1994). A connectionist model of recognition and naming of Chinese characters. In J.-T. H. H-W. Chang, C-W Hue, & O. Tzeng (Ed.), *Advances in the study of Chinese language processing* (Vol. 1, pp. 211-240). Taipei: National Taiwan University Press.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC : A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud. *Psychological Review, 108*(1), 204-256.

Dong Zhendong(2000), introduction to HowNet, from website http://www.keenage.com/.

Fang, S. P., Horng, R. Y., & Tzeng, O. J. L. (1986). Consistency effects in the Chinese character and pseudo-character naming tasks. In H. S. R. Kao & R. Hoosain (Eds.), *Linguistics, psychology, and the Chinese language* (pp. 11-21). Center of Asian Studies: University of Hong Kong.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, Reading Acquisition, and Dyslexia: Insights From Connectionist Models. *Psychological Review, 106*(3), 491-528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review, 111*(3), 662-720.

Honorof, D.N, & Feldman, L. (2006). The Chinese

character in psycholinguistic research: Form, structure, and the reader. In P. Li, E. Bates, L. H. Tan & O. J. L. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). Cambridge: Cambridge University Pess.

Hsiao, J. H.-W., & Shillcock, R. (2004). Connectionist modelling of Chinese character pronunciation based on foveal splitting. *Proceedings of the Twenty Sixth Annual Conference of the Cognitive Science Society*, 601-606.

Hsiao, J. H.-W., & Shillcock, R. (2005). Differences of split and non-split architectures emerged from modelling Chinese character pronunciation. *Proceedings of the Twenty seventh Annual Conference of the Cognitive Science Society*, 989-994.

Hue, C. (1992). Recognition Processes in Character Naming. In Chen Husanchih & T. O. J.L. (Eds.), *Language Processing in Chinese* (pp. 93-107). North-Holland.

Li, P. (2006). Modeling language acquisition and representation: Connectionist networks. In P. Li, E. Bates, L. H. Tan & O. J. L. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). Cambridge: Cambridge University Pess.

Liu, Q., & Li, S. (2002). Word similarity computing based on How-net. *Computational Linguistics and Chinese Language Processing*, 7, 59-76.

McBride-Chang, C., Cho, J.-R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., et al. (2005). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *Journal of Experimental Child Psychology,* 92(2), 140-160.

McBride-Chang, C., Shu, H., Zhou, A., Wat, C. P., & Wagner, R. K. (2003). Morphological Awareness Uniquely Predicts Young Children's Chinese Character Recognition. *Journal of Educational Psychology, 95*(4), 743-751.

Miller, G. (1990). WordNet: An on-line lexical database.*International Journal of Lexicography, 3*, 235-312.

Modern Chinese frequency Dictionary (1986). Beijing, China: Beijing Language Institute Press.

Pearlmutter, B. A. (1995). Gradient calculations for dynamic recurrent neural networks: A survey. *IEEE Transactions on Neural Networks, 6*, 1212–1228.

Peng, D., & Yang, H. (1997). The phonological processing of Chinese phonograms. *Asia and Pacific Journal of Speech, Language and Hearing, 2*, 177-195.

Peng, D., Yang, H., & Chen, Y. (1994). Consistency and phonetic-independcency effects in naming tasks of Chinese phonograms. In Q. Jing, H. Zhang & D. Peng (Eds.), *Information processing of the Chinese language* (pp. 26-41). Beijing: Beijing Normal University Press.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The Lexical Constituency Model: Some Implications of Research on Chinese for General Theories of Reading. *Psychological Review, 112*(1), 43-59.

Perfetti, C.A., & Liu, Y. (2006). Reading Chinese characters: Orthography, phonology, meaning and the lexical constituency model. In P. Li, E. Bates, L. H. Tan & O. J. L. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). Cambridge: Cambridge University Pess.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-Regular Domains. *Psychological Review, 103*(1), 56-115.

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523-568.

Shu, H., & Anderson, R. C. (1997). The role of radical awareness in learning characters of Chinese children. *Reading Research Quaterly, 32*(1), 78-89.

Shu, H., Anderson, R. C., & Zhang, H. C. (1995). Incidental Learning of Word Meanings while Reading: a Chinese and American Cross-cultural Study. *Reading Research Quaterly, 30*(1), 76-95.

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of School Chinese: Implications for Learning to Read. *Child Development, 74*(1), 27-47.

Shu, H., & Wu, N. (2006). Growth of orthography-phonology knowledge in the Chinese writing system. In P. Li, E. Bates, L. H. Tan & O. J. L. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). Cambridge: Cambridge University Pess.

Xing, H., Shu, H., & Li, P. (2002). A self-organizing connectionist model of character acquisition in Chinese. In W.D. Gray & C.D. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 950-955). Mahwah, NJ: Lawrence Erlbaum.

Xing, H., Shu, H., & Li, P. (2004). The Acquisition of Chinese Characters: Corpus Analyses and Connectionist Simulations. *Journal of Cognitive Science, 5*(1), 1-49.

Yang, H., & Peng, D. (1997). The learning and naming of Chinese characters of elementary school children. In H. C. Chen (Ed.), *Cognitive Processing of Chinese and Related Asian Languages* (pp. 323-346): The Chinese University of Hong Kong Press.