A Self-Organizing Connectionist Model of Early Word Production

Xiaowei Zhao (xzhao2@richmond.edu)
Department of Psychology, University of Richmond
Richmond, VA 23173 USA

Ping Li (pli@richmond.edu)
Department of Psychology, University of Richmond
Richmond, VA 23173 USA

Abstract

In this paper we present DevLex-II, a self-organizing neural network model of early word production. It consists of three self-organizing feature maps (a semantic layer, a phonological layer and a phonemic layer) that are connected via associative links trained by Hebbian learning. We use this model to simulate the early stages of lexical acquisition in children. The simulating results indicate a number of important effects in determining the timing and function of children's word production, such as word frequency and word length effects. In addition, results from lesioned models indicate developmental plasticity in the network's recovery from damage. Plasticity occurs at early stages, and changes with time in a non-monotonic and nonlinear fashion. These simulated patterns are due to the nonlinear dynamic properties of the network and match up with data from empirical studies of children.

Introduction

While many previous models of language acquisition have used back-propagation as the standard algorithm, in our work we have explored self-organizing neural network as a cognitively and neurally plausible model of language acquisition (Li, 2003; Li, Farkas, & MacWhinney, 2004). In this paper we present DevLex-II, an extension of the DevLex model as discussed in Li et al. (2004). The DevLex model has been applied to account for phenomena in early lexical development, including category formation, lexical confusion, and age of acquisition. The extended DevLex-II has been applied to simulate 'vocabulary spurt', a sudden and rapid increase in children's early productive vocabulary around 18 months of age (Li, Zhao, & MacWhinney, 2005). We have been able to model the emergence of vocabulary spurt as a function of a number of parameters, including verbal short-term memory and associative capacity. In this paper, we follow up on the initial findings to further examine patterns of early word production, in three respects: (a) word-frequency and word-length effects, (b) children's early pronunciation errors, and (c) the recovery from brain injury for early word learning.

First, word frequency and word length are two important variables that have been extensively studied in psycholinguistic research. Empirical studies show that token frequency and length of words determine the latency of a variety of tasks such as naming and lexical decision in adult language (Jescheniak and Levelt, 1994). Recent research indicates that these lexical properties may also affect lexical development in early child language: in particular, the age of acquisition of words may be correlated with the frequency of words in parental input (Storkel, 2004). In this study, we attempt to identify through modeling how frequency and length of words affect the time course of lexical acquisition.

Second, in the course of modeling early word production, we compare the performance of the model with empirical data in terms of word pronunciation. It is well known that, compared with the adult lexical forms, children's early speech involves omission, substitution, addition or reduplication of certain sounds or syllables. These patterns often show great individual variations and reflect children's lack of full mastery of articulatory programs at early stages of learning (Menn & Stoel-Gammon, 1993). Study of these errors at different ages can provide us with insights into the development of children's phonological abilities. One goal of the current study is to see if our model displays error patterns similar to those observed with children and to identify cognitive mechanisms underlying the errors.

Third, significant progresses have been made in the understanding of language development and its neurological underpinnings through research with children who suffer from brain injuries that are different in size, location, onset age, and so on (Bates 1999; Bates & Roe, 2001; Vargha-Khadem, Isaacs, & Muter 1994). Researchers have also attempted to simulate developmental language disorders using connectionist nets (Marchman 1993; Thomas & Karmiloff-Smith, 2002). Empirical studies show that, children with early brain injury can go on to acquire linguistic abilities within normal range, whereas similar lesions in adults produce dramatic patterns of aphasia (Bates & Roe, 2001). These data provide evidence supporting a general view that great plasticity is an early privilege (Thomas, 2003). However, how plasticity changes with time is a complex problem and is still unclear. Bates and colleagues have argued that the shape of the function is not monotonic (Bates, 1999; Bates & Roe, 2001). In this study, we hope to shed some light on the issue by observing network performance and its ability to recover under various lesion conditions.

The Model

A Sketch of the Model

The DevLex-II model is based on the DevLex (Li, Farkas, & MacWhinney, 2004) and the DISLEX model (Miikkulainen, 1997). Figure 1 presents a diagrammatic sketch of the model. The model has three basic levels for the representation and organization of linguistic information: Phonemic sequence, phonology, and semantics of the lexicon. At the core of the model is a self-organizing, topography-preserving, feature map (Kohonen, 2001), which processes semantic information of words (meaning). This feature map is connected to two other feature maps, one for the processing of the sound structure of words (phonology), and another for the phonemic sequences of words (phonemes).
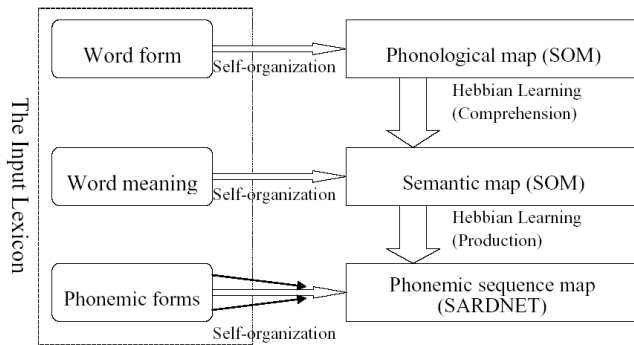

Figure 1: The DevLex-II model of lexical development.

Upon training of the network, the meaning, phonology, and phonemic sequence of a word are presented to and processed by the network. This process can be analogous to the child's analysis of a word's semantic, phonological, and phonemic information upon hearing a word. On the semantic and phonological levels, the network forms representational patterns of activation according to standard self-organizing map algorithms (Kohonen, 2001). Here, a SOM is a two-dimensional square lattice with a set of neurons, and each neuron k on the level has an input weight vector m_k associate with it. Given a stimulus $\mathbf{x}$, the localized output response of neuron k is computed as：

$$\alpha_k = \begin{cases} 1 - \dfrac{\|x - m_k\| - d_{\min}}{d_{\max} - d_{\min}}, & \text{if } k \in N_c \\ 0 & , \text{otherwise} \end{cases}$$

Where N_c is the set of neighbors of winner c ($a_c = \max_k\{a_k\}$), d_{min} and d_{max} are the smallest and the largest Euclidean distances of $\mathbf{x}$ to node's weight vectors within N_c.

The phonemic level works in a slightly different way from the other two levels. The addition of this level is inspired by models of word learning based on temporal sequence acquisition (e.g. Gupta & MacWhinney 1997). It is designed to simulate the challenge that children face during the second year when they need to develop better articulatory control of phonemic sequences of words. Just like the learning of auditory sequences requires the mediation of memory systems, the learning of articulatory sequences may involve support from the articulatory loop of the working memory (e.g. immediate serial recall; Gupta & MacWhinney, 1997). In our implementation of this idea, the activation pattern corresponding to the phonemic sequence of a word is formed according to the algorithms of SARDNET (James & Miikkulainen, 1995). At each training epoch, phonemes of a word are input into the map sequentially according to their order of occurrence in the word. The winner of each phoneme is found, and the responses of nodes in its neighborhood are adjusted. Once a node becomes the winner of an input, it is made ineligible to respond to the subsequent inputs in the sequence. This way, same phonemes in different locations of a word will be mapped to different nodes in the map. In addition, when the output status of the current winner and its neighbors is adjusted, the winners responding to previous phonemes before the current phoneme will be affected by a factor γ^d, where d is the distance between the location of the current phoneme and the previous phoneme that occurred in the word. This process can be used to represent the effect of short-term verbal memory during the learning of articulatory sequences. The factor should be less than 1 (0.8 in our case), as the effect should decay with time. One can consider this decaying effect as reflecting the decay of strength in the phonological memory of phonemes in children's word learning. So for a word with l phonemes, the output of the winner responding to the jth phoneme will be $1 + \gamma + \gamma^2 +\ldots\ldots+ \gamma^{l-j}$, which is a geometric progression, and can be written as:

$$(1-\gamma^{l-j+1})/(1-\gamma)$$

According to this equation, when all phonemes' representations of a word are sent to the phonemic map, the activation of some nodes (e.g., the first winner) will be larger than 1, so they need to be normalized between 0-1. With the identification of winners on each SOM level, weights of nodes around these winners are updated (self-organized) as:

$$m_{kj}(t+1) = m_{kj}(t) + \alpha(t) \cdot [x_j - m_{kj}(t)] \text{ for all } k \in N_c$$

Here, $\alpha(t)$ is learning rate, which changes with time.

In DevLex-II, the activation of a word form can evoke the activation of word meaning via form-to-meaning links (to model word comprehension) and the activation of word meaning can cause the formation of phonemic sequence via meaning-to-phoneme links (to model word production). Simultaneously with input weight change, the weights of associative links between the features maps are trained by Hebbian Learning, and the associative weight vectors are then normalized.

Input Representations

To model early lexical acquisition by children, we used as our basis the vocabulary from CDI, the MacArthur-Bates Communicative Development Inventories (Dale & Fenson, 1996). From the Toddler's List, we extracted 591 words (the original Toddler's List contains 680 words; we excluded the homographs and homophones, word phrases, and onomatopoeias from our analysis).

The phonological input representations of the 591 words were generated by PatPho, a generic phonological pattern generator for neural networks (Li & MacWhinney, 2002). A left-justified template with 114 dimensional binary encoding was adopted. The semantic representations of these same words were generated by WCD, a word co-occurrence detector that learns the lexical co-occurrence constraints of words. It reads through a stream of input sentences (one word at a time) and learns the transitional probabilities between words which it represents as a matrix of weights. The input sentences were from the parental input of the CHILDES corpus, which contains the speech transcripts from child-directed adult speech in CHILDES (MacWhinney, 2000). Finally, like in PatPho, we represented the 38 phonemes by vectors based on articulatory features of the phonemes.

Simulation parameters

In DevLex-II, the phonological map or the semantic map each consists of 60 x 50 nodes, and the phonemic map consists of 15 x10 nodes. These numbers were chosen to be large enough to discriminate among the words and phonemes in lexicon, while keeping the computation of the network tractable. The same learning rate $\alpha(t)$ and the same radius of winner's neighborhood were used for all feature maps, and they change with time. The training process had two phases: the ordering phase and the convergence phase. Learning rate $\alpha(t)$ was initially set as 0.4, then linearly decreased to 0.05 during the first 50 epochs (ordering phase). In the next 50 epochs (convergence phase), it remained at 0.05. At the same time, the neighborhood radius reduced from 3 to 0 and then remained at 0 until the end of training. Learning rate β for associative links between levels was kept constant at 0.1 during the whole training process.

At each epoch, words from the training lexicon were presented to the network one by one. To simulate the effect of word frequency in early child language, the network chose a word each time according to its frequency of occurrence in the parental CHILDES corpus. Since word frequency distributions follow Zipf's law, we calculated the logarithms of the frequencies to force a more even distribution of words in the input.

An additional parameter 'connection probability' was introduced to the model to simulate individual differences in the development of associative abilities. Here, initially, two feature maps are not fully but only partially connected by associative links. The ratio of the number of connected links to the number of all possible links between two maps is defined as connection probability. The connection

probability was set to linearly increase with time from a low threshold θ (<1.0) to full connectivity (1.0), as opposed to full connectivity throughout in the unmodified model.

Results and Discussion

Word-frequency and Word-length effects

In our model, word frequency of the training vocabulary is determined by how frequently the words occur in the CHILDES transcripts. Word length is determined by the number of phonemes a word has. We divided frequency into three ranges, low (<10 times in the 2.7 million word corpus), medium (10-10000 times), and high (>10000 times), and word length into short (<=3 phonemes), medium (4-5 phonemes), and long (>7 phonemes). The short words include mainly monosyllables, while the medium and long words are made up of two to three syllables.

First we recorded the AoA of each word. AoA is defined in the model as the time (training epoch) at which a word is learned. We say that a word is learned in production, when a node in the semantic map can consistently activate a set of phonemes in sequence as winners of the input word in the phonemic map via the meaning-to-phoneme associative links. Then, we calculated the percentage of words acquired for each frequency or length level at each given epoch of training. The results are shown in Fig. 2. Clearly, acquired words of all frequency and length types show a rapid increase in vocabulary size around epoch 40. This vocabulary spurt phenomenon has been captured by the DevLex-II model and has been discussed by Li, Zhao & MacWhinney (2005). In Figure 2, we can see that the spurt curve is significantly dampened for low-frequency and long-phoneme words, especially toward the mid-to-late stages of training. This shows that in our network short and high-frequency words were learned more easily than long and low-frequency words.

These findings suggest that in children's early productive vocabulary short and high-frequency words are more likely to be acquired or will occur earlier than long and low-frequency words. Although there has not been much empirical work on word frequency, length, and AoA in young children (in contrast to adult psycholinguistics work), a recent analysis by Storkel (2004) confirms the patterns in our model. Storkel made a linear regression analysis of nouns obtained from two databases, CDI and adult self-ratings of AoA. She found that AoA of words in children's early vocabulary are negatively correlated with word-frequency, but positively correlated with word-length, such that children's early acquired words are "higher in word frequency, and shorter in length than late acquired words" (Storkel, 2004).

It is worth noting here that, when encountering long and low-frequency words, our network tended to produce wrong sequences or omit phonemes. Such patterns parallel children's early speech errors and reflect the system's poor short-term memory or lack of full articulatory control, which brings us to the next section.
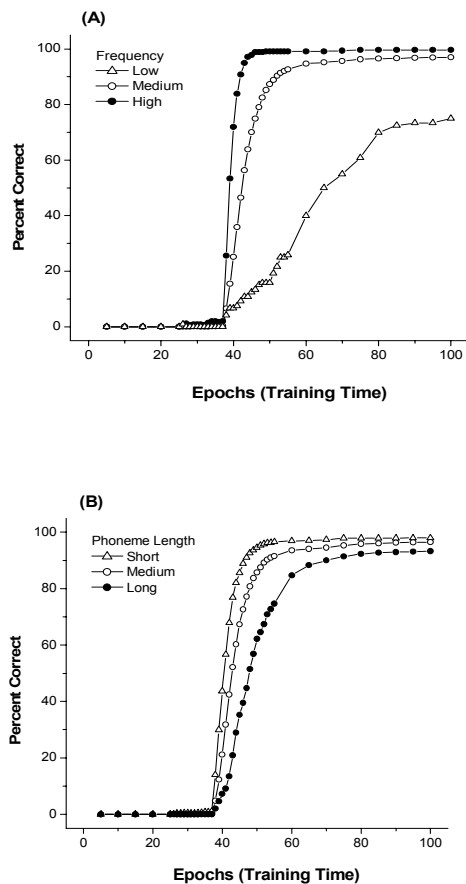
Figure 2: Effects of (A) word frequency and (B) word length, based on 10 simulation runs.

Word Production: Error Analyses

Table 1 presents a list of typical examples from our network's word productions at different training times. These errors parallel children's early word pronunciations, such as omission of consonants at the end of a word (e.g., output to 'bib' at epochs 50, 60); deletion of a consonant in consonant clusters (e.g., outputs to 'smile' and 'glue' at epochs 60, 80 and 100); substitution of consonants with similar phonemes (e.g., /d/ in 'bird' is pronounced as /b/). These errors were due to (a) incomplete meaning-to-phoneme links, and (b) incomplete sequence learning of phonemes. The similarity of the errors between our model and real children suggests that incompletely developed associative links and poor working memory for phonemic sequence may explain children's failure to produce the correct sounds of words.

Table 1 also shows other interesting results. For example, in two different simulation trials, responding to the word 'sock', the system gave two different patterns of production error, the deletion of consonant /k/, and the substitution of it with /t/. Given that the simulation trials had the same parameters, this difference reflects individual differences in

phonological development within and across children (Menn & Stoel-Gammon, 1993). We can also track the development of the sound patterns from Table 1. At early stage of learning, our net's productions were simple and often very different from the words' real pronunciations, similar to children's simplified patterns. During the middle and late stages of learning, our model's output becomes more like real language with correct pronunciations. The amount of correct productions gradually increased. Although there were still production errors, they were closer to the target pronunciations and had typical error patterns as discussed above. The coexistence of correct and incorrect word pronunciations correspond to empirical patterns in children's phonological development from babbling to word production (Menn & Stoel-Gammon, 1993).

To summarize, at the beginning, our model can only pronounce simple, blurry sounds. With the emergence of self-organized structure on every layer, especially the phonemic layer, and the developing associative links, the system's output resembles real words. The transition from wrong sequence, substitution, and omission of phonemes to correct pronunciation indicates that our model is able to capture developmental patterns in phonological acquisition with simple self-organizing principles.

Effect of lesion and developmental plasticity

To model the role of lesion in early lexical production, we added noise to the input connections of a chosen layer at a given training epoch. In particular, each input link's weight has a certain probability to be multiplied by a random number uniformly distributed between 0 and 1. The probability level determines the size and severity of the lesion. For example, if it is 0.6, then approximate 60 percent of the input connections of a layer are damaged by stochastic noises. To simplify discussion, here we show only the results with the probability value at 1.

Figure 3 presents vocabulary development at epoch 55 (mid-stage). Both word comprehension and production rates decreased when lesion was introduced to the semantic layer. Some degree of recovery appeared in the model, but it was obvious that the network's learning was delayed, as the final vocabulary size could not reach a normal level within the learning window. Similar results were obtained with other damaged layers, but only comprehension was affected when the phonological layer was lesioned, and only production was affected when the phonemic layer was lesioned. Our results are consistent with empirical studies that linguistic abilities may be delayed following brain injury (Bates & Roe, 2001). It also shows that the ability to organize semantic information is very important to the vocabulary development process. Without a well-structured semantic representation, the perceived phonological information of a word cannot be correctly projected to its semantic target; a jumbled semantic representation also cannot trigger proper word production.

Young children with brain injury often recover well, but this plasticity changes with time. To investigate the develop-

Table 1 Sample production errors from 10 simulation trials

Words	Epochs (Training Time)					
	30	40	50	60	80	100
Dog (/dɔːg/)	-	/d/	/d/	/dɔː/	/dgɔ/	/dɔːg/
Sock (/sɑːk/)	/ʃd/	/suː/	/suː/	/sɑː/	/sɑː/	/sɑː/
Bib (/bɪb/)	/ɑː/	/br/	/bɪ/	/bɪ/	/bɪb/	/bɪb/
Cat (/cæt/)	/ɑ/	/cæ/	/cæ/	/cæ/	/cæt/	/cæt/
Brush (/brəʃ/)	/n/	/bən/	/bən/	/bəʃ/	/bəʃr/	/brəʃ/
Smile (/smaɪl/)	/pəliː/	/Ima/	/Ima/	/maɪl/	/maɪl/	/maɪl/
Glue (/gluː/)	/I/	/g/	/g/	/guː/	/guː/	/guː/
Sock (/sɑːk/)	/ɑ/	/sɑːt/	/sɑːt/	/sɑːt/	/sɑːt/	/sɑːt/
Bird (/bɜːd/)	-	/bb/	/bɜːb/	/bɜːb/	/bɜːb/	/bɜːb/

Note: '-' indicates no unique output of the word since the word is confused with other words on the semantic map at the current time.
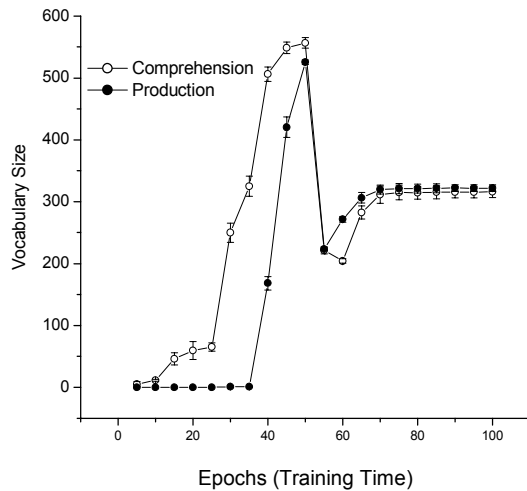

Figure 3: Vocabulary development with semantic layer damaged.

Figure 4: Mean final size of productive vocabulary as a function of the onset time of lesion, based on five simulation runs.

mental plasticity, we introduce lesions at different times to the network, and then calculate the final acquired vocabulary size. The results are shown in Figure 4.

Generally, the final acquired vocabulary size of our network was larger when lesion occurred earlier than it was later. For both word comprehension and production the network recovered more easily from early damage than from late lesion. This pattern is consistent with the general pattern of developmental plasticity (Bates, 1999; Tomas, 2003). However, our results suggest that it is not a simple monotonic decrease. The worst outcome for the final vocabulary size was not when damage occurred the latest (epoch=80), but when it occurred midway (epoch=40). This pattern resembles a kind of U-shaped change and is consistent with empirical evidence of children's recovery from focal brain injury as discussed in Bates (1999).
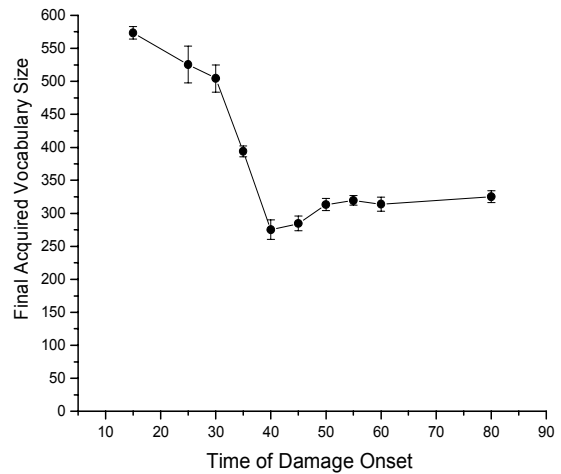
The developmental plasticity shown in our model is related to the nonlinear dynamic properties of the network. In particular, early on, on each layer, structures of different linguistic information have not been organized completely, and the associations between the layers are not strong enough to form fixed patterns. Thus, the whole system is in a dynamical unstable state. It is sensitive to small external changes and can adjust weights more easily, hence its ability to recover from damage. At later stages, the system reaches a dynamical stable state because clear patterns have formed on each layer and for associations between layers. The system is now robust to small external changes and becomes harder to adjust weights, and so if lesion occurs, complete recovery is more unlikely (see also Elman et al., 1996).

With regard to the U-shaped pattern, the nonlinearity may reflect the complex dynamical competition of different factors in our neural network. In particular, when a large

lesion on semantic layer occurs, the structure on the layer is widely destroyed, but it is possible that at a later stage, some strong associations between semantic and phonemic layers are resilient to noise, leading to the pattern that a late-occurring lesion (e.g., epoch 80) gives better recovery than lesions occurring midway (e.g., epoch 40). When lesion occurred mid-course, the recovery cannot take advantage of the cross layer associations because no strong associations have been formed, while at the same time the network's sensitivity to large changes has dropped significantly. Thus, noise at epoch 40 has a more devastating effect as the network transitions from an unstable state to a stable state in the dynamical space.

Conclusion

There are three conclusions that we can draw from the simulated results of our self-organizing neural network.

First, our model captures important empirical phenomena in children's early word production. This ability is due to the simple computational principles of self-organization and associative learning built into DevLex-II. The simulations further attest to the utility of self-organizing neural nets as models of language acquisition.

Second, our models shows that lexical acquisition depends on the interaction of many factors, including the self-organization of relevant linguistic information (phonological structure, phonemic sequence, and semantic organization), and the development of associations across the domains (form-to-meaning, meaning-to-form links). Word production errors may be due to poor structure in the representation, incomplete associations, or both.

Third, our model shows that individual differences in lexical acquisition may be attributable to (a) input characteristics (such as frequency and length of words in the input), (b) associative capacity (see also Li, Zhao, & MacWhinney, 2005), and (c) delayed or damaged learning, which by itself is a joint function of the nonlinear dynamic interaction among timing, severity, and recoverability of lesion.

Acknowledgments

References

Bates, E. (1999). Plasticity, localization and language development. In S.H. Broman and J.M. Fletcher (Eds.), *The changing nervous system: Neurobehavioral consequences of early brain disorders*. New York: Oxford University Press.

Bates, E., & Roe, K. (2001). Language development in children with unilateral brain injury. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience*. Cambridge, MA: MIT Press.

Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28*, 125-127.

Elman, J., Bates, A., Johnson, A., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Gupta P, & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language. 59*, 267-333.

James, D., & Miikkulainen, R. (1995). SARDNET: A self-organizing feature map for sequences. In G. Tesauro, D. S. Touretzky, and T. K. Leen (Eds.), *Advances in Neural Information Processing Systems 7* (NIPS'94; pp.577-584). Cambridge, MA: MIT Press.

Jescheniak, J.D., & Levelt, W.J.M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 824-843.

Kohonen, T. (2001). *The self-organizing maps* (3rd ed.). Berlin: Springer.

Li, P. (2003). Language acquisition in a self-organising neural network model. In P. Quinlan (ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks*. Hove & Briton: Psychology Press.

Li, P., Farkas, I., & MacWhinney (2004). Early lexical development in a self-organizing neural network. *Neural Networks. 17*, 1345-1362

Li, P., & MacWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments, and Computers, 34,* 408-415.

Li, P., Zhao, X., & MacWhinney, B. (2005) From avalanche to vocabulary spurt: Dynamics in self-organization and children's word learning, under review.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum.

Marchman, V. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 5(2),* 215-234.

Menn, L., & Stoel-Gammon, C. (1993). Phonological development: learning sounds and sound patterns. In Gleason, J.B., (Ed.), *The development of Language* (3rd ed.). New York : Macmillan.

Miikkulainen, R. (1997). Dyslexic and category-specific aphasic impairments in a selforganizing feature map model of the lexicon. *Brain and Language, 59*, 334-366.

Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics, 25,* 201–221

Thomas, M. & Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behavioral and Brain Sciences, 25(6)*, 727-750.

Thomas, M. (2003). Limits on plasticity. *Journal Of Cognition And Development, 4(1)*, 95–121

Vargha-Khadem, F., Isaacs, E., & Muter, V. (1994). A review of cognitive outcome after unilateral lesions sustained during childhood. *Journal of Child Neurology, 9(Suppl.),* 2S67-2S73.