A SOM Model of First Language Lexical Attrition

Benjamin D. Zinszer (bdz107@psu.edu)
Ping Li (pul8@psu.edu)
Department of Psychology, Pennsylvania State University
University Park, PA 16802 USA

Abstract

First language lexical attrition remains a difficult phenomenon to study empirically, due to its long-term and dynamic effects. Based on observations from existing case studies, we propose a connectionist model to simulate the effects of first language lexical attrition. The model exhibits a plausible time-course for first language lexical comprehension, highlights the independence of productive and receptive attrition trajectories, and predicts an age of onset effect for early cases of L1 lexical attrition.

Keywords: language attrition, lexicon, modeling, self-organizing map, connectionism

Introduction

Many people learn and forget a second or third language during the course of their lifetimes. Less often, a major migration may cause someone to forget all or part of his native language. While a great deal of research has been dedicated to the first and second language acquisition, relatively little is known about language loss (hereafter, attrition) in the individual speaker.

Lexical Attrition

In the last decade, there has been an increasing amount of work devoted to the study of language attrition, specifically in L1 or first language attrition. Apparent age-related effects have been observed in the attrition of L1 phonology (Hytenstam et al, 2009; Pallier et al, 2003). However, long-term lexical attrition has remained largely undocumented, partly due to the lack of rigorous experimental methodologies for the study of lexical attrition.

Nonetheless, one could reasonably expect the long-term course of lexical attrition to differ from that of phonology. Previous research examining the interplay between language learning and cognitive functions has identified differing memory stores for lexical and phonological acquisition (Hernandez & Li, 2007; Ullman, 2001). To the extent that continued performance in the L1 depends on different memory representations, the effects of attrition on phonology and the lexicon may be independent.

The current body of L1 lexical attrition research provides some general observations about the relationship between age of onset (AoO)[1], length of residence (LoR) and the degree of attrition. A case study of letters written by an L1-German immigrant to the United States revealed an ongoing process of lexical attrition even fifty years after AoO (Hutz, 2004). In another case study, an L1-German speaker with a similarly long LoR of 47 years in the United States demonstrated substantial lexical relearning in a natural conversational setting (Stolberg & Münch, 2010). That relearning is possible after such a long time raises the question of whether lexical attrition is truly a case of forgetting, in which L1 knowledge is destroyed in memory, or whether it is the access to L1 knowledge that is primarily affected by attrition.

The most evident problem in current L1 attrition research is the difficulty of reliably measuring change across time. As demonstrated in the case studies, loss of L1 abilities may be a slow and gradual process spanning years or decades. As a result, even longitudinal studies over a few years capture only a snapshot of a highly dynamic language system. The limited span of longitudinal data provided by any single study makes it extremely difficult or statistically impossible to identify the time course of development. While large samples with cross-section age variables (such as age of acquisition in the L2 literature) can mitigate these problems, advanced language users who experience L1 attrition are relatively scarce, making a cross-sectional sample nearly impossible.

One small-scale quantitative study has tested L2 lexical attrition through the relearning paradigm. De Bot, Martens, & Stossel (2004) found a relearning advantage in foreign language study for forgotten words over new words, revealing that the forgotten words, though inaccessible, persisted in memory. While this study found a general adherence to an exponential forgetting curve in which relearning savings are possible below the productive threshold, the findings are difficult to generalize due to the limited size of the vocabulary and the limited scope of the study.

Given the difficulties in systematic control of important learning variables (such as age, language proficiency, and L2 exposure), language attrition research has remained mostly a descriptive enterprise. Computational modeling may serve to turn language attrition research to an experimental science, due to its flexibility in parametric manipulation of the relevant variables and in testing relevant theoretical hypotheses. To date, very little work has been done in the computational modeling of language attrition. The goal of this study is to make a first attempt in providing a detailed computational account of the developmental time course of language attrition in the lexical domain.

---

[1] "Age of onset" here refers specifically to the beginning of attrition. Due to the difficulty of identifying this event, AoO is typically marked by the change of language environment (e.g., geographic migration), prior L2 exposure notwithstanding.

Computational Models

To our knowledge there has been only one computational model specifically designed to address lexical attrition. Meara's (2004) Boolean model of lexical attrition used a simple connectionist paradigm to simulate the effect of intra-lexical relationships on the time course of attrition. Meara's model exhibited self-organized criticality, that is, the wide-spread and sudden deactivation of lexical nodes at unpredictable intervals. This effect may be interpreted as largely a product of the inter-node dependencies inherent to Boolean models, but more importantly, Meara found that when the mean activation was taken across ten models, the resulting curve showed a gradual decline. This study highlights the troubling possibility that empirical research of lexical attrition in human subjects is hiding potential criticality effects. Increasingly sophisticated computational models may yet fill this gap.

Self-organizing feature maps (SOM) are a promising option in modeling lexical attrition. SOM is a connectionist modeling paradigm which represents data in a network of clustered nodes. Previous research has established the utility of SOM in producing cognitively plausible models of language development (see Li, 2009, for a review; see also Richardson & Thomas, 2008 and Mayor & Plunkett, 2010).

The potential for extending SOM to lexical attrition is suggested by its flexibility in simulating the effects of competing input sets. Age-related dynamic cross-linguistic competition in L2 learning has been demonstrated with other SOM-based models (Li & Farkas, 2002; Zhao & Li, 2007). Furthermore, effects of sensitive period or catastrophic interference have also been shown with the manipulation of learning parameters in SOM (Richardson & Thomas, 2008).

Computational modeling offers the possibility of a unified account of language learning, attrition, and relearning phenomena, integrating empirical research in these fields under more durable hypotheses. The present study aims to produce a SOM model: (1) to replicate the sustained gradual erosion of L1 lexical knowledge in both production and comprehension, (2) to compare the respective rates of attrition for comprehension and production, (3) to produce a plausible time course for long-term L1 lexical attrition, and (4) to reveal age of onset effects in L1 lexical attrition.

Method

In this study, a dual self-organizing feature map (SOM) model is trained in a first language (L1) and at varying ages of onset (AoO) in a second language (L2) while L1 training decreases or stops. Performances of the model in comprehension and production are tracked throughout training.

The Model

The self-organizing feature map (SOM) is a connectionist modeling paradigm wherein each node contains a vector of weights corresponding to each member of the input vector (see Kohonen, 2001 for a detailed explanation of SOM).

Node weights falling within a defined neighborhood around the input vector are adjusted towards the input based on their distance from it. Over many epochs of training, this adjustment results in topography-preserving orders, such that similar inputs are represented by nearby clusters of nodes in the map while dissimilar inputs by distinct and distant clusters. The typography-preserving characteristics of SOM are particularly well suited for examining the effects of cross-language lexical competition in a dynamically evolving system as in lexical attrition.

Architecture The model designed for this study employs two such SOMs (see Figure 1). The first SOM was trained on the phonological representations of words. This phonological map self-organizes according to the basic phonemic elements in a word, clustering words of a similar sound together. The phonological map was composed of 1600 nodes on a 40 by 40 rectangular grid. The second SOM was trained using the semantic representation of words. The semantic map clustered words of similar meaning, category, and part of speech. The semantic map was composed of 900 nodes arranged on a 30 by 30 rectangular grid. The semantic map was designed to be smaller than the phonological map because it received half as many unique input representations (see *Stimuli and Training*). The two maps were joined by Hebbian connections (see Hebb, 1949 for model and biological basis). A single Hebbian connection is represented by a weight that multiplies activation between the two nodes it connects. Every node on one map was connected to every node on the opposite map, for a total of 1.44 million (1600 x 900) Hebbian connections.

Functions and Parameters After the presentation of each input stimulus, the maps and Hebbian connections were updated according to a set of learning functions. These functions defined which sets of nodes and weights are adjusted and how much they are adjusted.
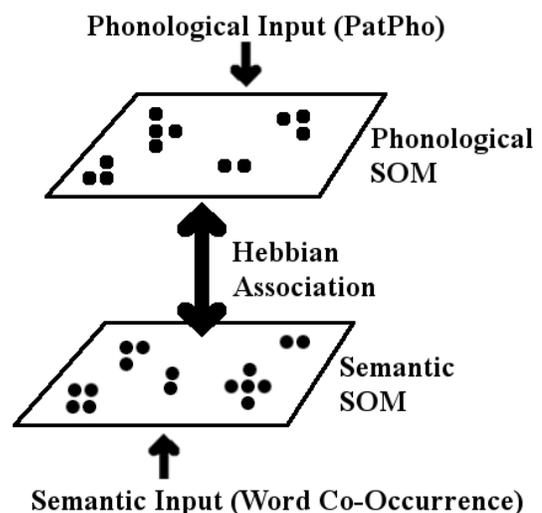
Figure 1: The model is composed of two self-organizing feature maps. Activation in each map is propagated to the other by means of Hebbian Connections.

On the phonological and semantic maps, the node whose weights most closely match those of the input set (measured as minimal Euclidean distance between input and each node) is designated as the Best Matching Unit or BMU. The nodes around the BMU are updated according to a neighborhood function approximating a Gaussian curve with a maximum value of one at the BMU.

The radius of the neighborhood is variable between trials and measured in terms of the Cartesian distances between nodes on the rectangular grid. In this study, the radius was initially set at one half the size of the smaller map (15) to allow maximum adjustment in early trials. With each epoch the radius was allowed to decrease by one if the quantization error was less than in the preceding trial. With this approach, performance of the model was not directly tied to a manipulation of the radius size, but rather the radius size and model performance were allowed to covary through early training stages.

Updates to SOM weights were proportional to a node's value on the Gaussian neighborhood curve, resulting in a smaller change for more distant nodes, and no change for nodes outside the neighborhood. All updates were also multiplied by the SOM's learning rate, a value between zero and one which limits the amount of change that can occur in a single trial. A learning rate of 0.2 was set for both maps.

Hebbian connection adjustment was determined by co-activation in both maps. Activation for each node within the BMU's neighborhood was inversely proportional to Euclidean distance between the node's weights and the input vector. Each Hebbian connection was then adjusted by multiplying the activation of the nodes on each map and the Hebbian learning rate. The Hebbian learning rate was set to 0.1 in this model. Following each trial, Hebbian weights were normalized to values between zero and one.

Stimuli and Training

Two types of stimuli were provided to the model for training. Vectors containing phonological representations of words were presented one at a time to the phonological map. Simultaneously, vectors containing semantic representations of the same words were presented to the semantic map. This paired presentation allowed each map to organize around its respective input and then form connections between the phonological and semantic representations on their respective maps.

Phonological input vectors were generated using the PatPho system for English (Li & MacWhinney, 2002) and Mandarin Chinese (Zhao & Li, 2009). The dimension of each phonological vector was 63 units. Semantic vectors were obtained from the English stimulus set used to train the DevLex-II model. Each semantic input vector was 200 units long, derived from word co-occurrence patterns (see Li, Zhao, & MacWhinney, 2007 for details). In order to help the model discriminate between highly similar words (such as *red* and *blue* or *grandma* and *grandpa*) a nominal amount of noise was randomly added to the semantic data before training began for each model.

Most importantly, the English semantic representations were paired with both Chinese and English phonological representations during training. While emergentist models of bilingualism such as the Unified Competition Model have accounted for semantic and lexical transfer in second language acquisition (MacWhinney, 2005), prior computational models of language acquisition have failed to account for the largely shared conceptual space between two languages. Due to the importance of L2 negative transfer in L1 lexical attrition (Hutz, 2004; Schmitt, 2010) a computational account would be incomplete without a common semantic representation.

Words for the training set were selected from the MacArthur-Bates Communicative Developmental Inventories (English: Dale & Fenson, 1996; Chinese: Hao et al, 2008). Originally, 140 rough translation equivalents were obtained by comparing the English index with the English glosses in the Chinese index. Because intonation was not coded in the phonological representation, several words were eliminated as homophones. A few other words were removed because they could not fit the PatPho template for phonological encoding or did not have readily available co-occurrence data for semantic input. In total 116 English and 116 Chinese words were phonologically and semantically encoded for input to the model.

All instances of the model were trained for 500 epochs. The L1 (Chinese) was trained first, and at varying numbers of epochs (AoO) L1 input ceased and L2 (English) input began. AoO was varied in intervals of 50 epochs from 50 to 400. Ten models were trained for each of the eight AoOs.

Performance Tests

Following each training epoch, production and comprehension of the L1 was tested throughout the entire lifespan of the model.

For modeling purposes, comprehension was defined as the activation of the correct BMU on the semantic map when a phonological stimulus was presented to the phonological map. This activation was achieved by means of the Hebbian connections. After presentation of the stimulus, activation on the phonological map was calculated by the same method described in Functions and Parameters (above). Activation levels in the phonological map were then multiplied through their Hebbian connections. The incoming activation on the semantic map was summed for each node, and the most activated node on the semantic map was found. This most activated node was then compared to a list of semantic BMUs. If the most activated node was also the correct BMU, comprehension had occurred. If no BMU occupied the most activated node, the most activated node was compared to the closest BMU (by Cartesian distance) on the map. In the event that two or more BMUs on either map occupied the same node, all of these BMUs were disqualified from the comparison, preventing their corresponding words from passing the comprehension measure.
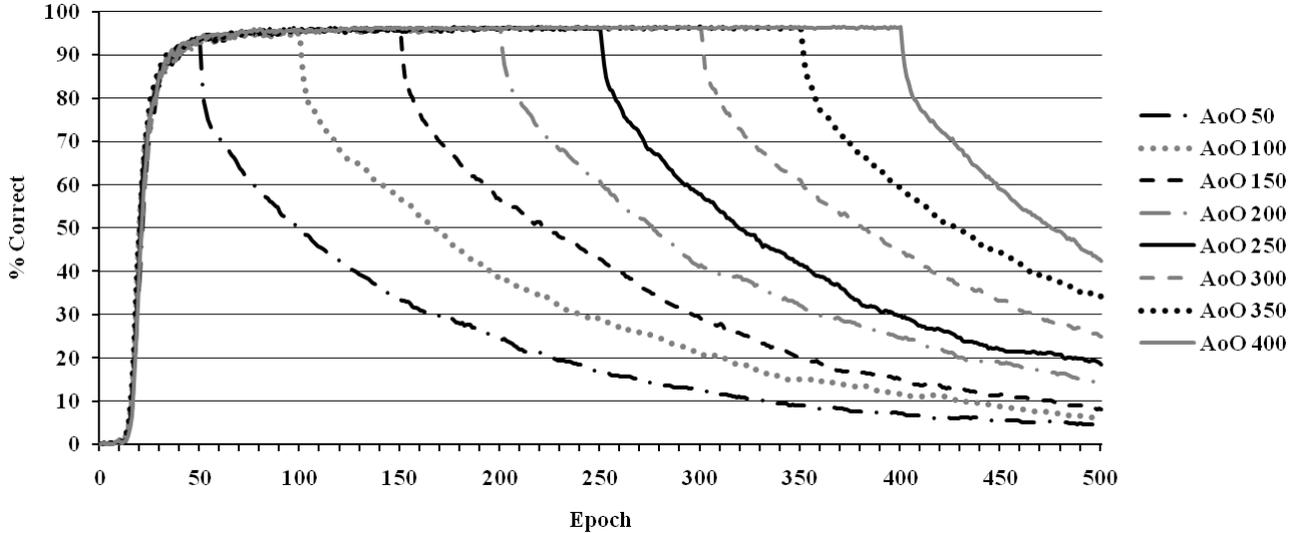
Figure 2: Mean comprehension scores in each AoO condition are graphed across the duration of the models (measured in training epochs). AoO values are also measured in training epochs, as depicted along horizontal axis.

Production was defined by the reverse process of comprehension. A stimulus input was provided to the semantic map, and activation was propagated by the same Hebbian connections to the phonological map, and the most activated unit was compared by the same criteria to the phonological BMUs. One important distinction in the case of production is that the most activated unit was only compared to the L1 BMUs to avoid inter-language confusion.

Results

L1 Comprehension

Mean comprehension curves were calculated across ten models for each AoO condition. Performance for each condition exceeded 92% by 50 epochs (the earliest AoO). AoO conditions later than 50 epochs exceeded 95% comprehension by 100 epochs. Maximum L1 comprehension after 100 epochs was 96.6% (112 out of 116 L1 words) for all models (un-averaged) with an AoO greater than 100. After AoO began, L1 comprehension decreased monotonically. Figure 2 shows the L1 comprehension curves for all eight AoO conditions. At the onset of L2 training, L1 comprehension seems to approximate an exponential decay for each AoO condition. Differences between L1 curves are described below (see section *Age of Onset Effects*).

L1 Production

L1 production declined severely and immediately for all AoO conditions. All models across all conditions performed below 5% correct productions within four epochs of the AoO and remained low throughout L2 training. Due to the low performance, no further analysis was applied to these data. See the *Discussion* section for a further treatment of this topic.

Age of Onset Effects

By visual inspection, AoO was inversely related to rate of attrition for the earlier AoO conditions. To quantify this relationship, the number of epochs required for each AoO condition to drop below 75% comprehension was calculated. Many models in the AoO 50 and 100 conditions did not reach maximum L1 comprehension performance by L2 onset. Therefore the performance calculation compensated by adding to performance measures the difference between each model's maximum L1 comprehension and the overall maximum (112) before calculating the number of epochs necessary to reach the threshold.

Figure 3 approximates the rate of attrition for each AoO by showing the number of epochs elapsed after L2 onset before the 75% L1 comprehension threshold was reached. Error bars indicate the two standard errors of the mean for
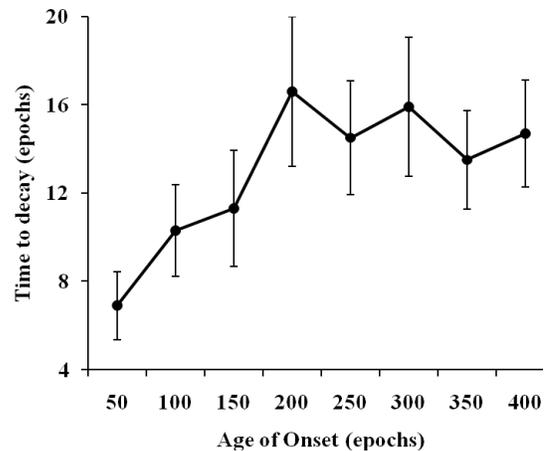
Figure 3: Mean number of epochs to reach 75% L1 comprehension or less, adjusted for incomplete learning. Error bars are two standard errors of the mean.

each AoO condition. ANOVA revealed a highly significant difference ($p<0.001$) in mean decay rates between conditions. Post-hoc tests (Tukey, with a family alpha of 0.05) showed that the AoO 50 condition was significantly different than AoO 200-400 (but not 100 and 150), while AoO 100 was also significantly different from 200.

Discussion

While an examination of learning in connectionist models may be interesting in its own right, the results of this study are most informative with regard to the dynamic trajectories of human first language attrition. Prior studies in L1 attrition have found age effects in phonological attrition, but no such effect has been demonstrated for lexical attrition. Nonetheless, a review of the current L1 lexical attrition literature reveals that lexical attrition is a long-term and dynamic process.

Performance measures for L1 comprehension and production after AoO indicated great instability in the production while comprehension declined more gradually. A potential source of declining performance in both measures was the changing Hebbian weights. Because the weights were normalized with each trial, the magnitude of change to the Hebbian connections due to a stimulus is not strictly dependent on activation levels. This effect is analogous to a decay (or forgetting) rate, as all connection weights were reduced relative to the learning rate.

A major source of instability, and a probable driving factor behind the rapid decline in production, was the reorganization of the phonological SOM. The operational definition of comprehension assumed activation of the correct phonological representations (if present) and tested the consequent activation on the semantic map, rendering comprehension relatively resistant to changes in the phonological map. By contrast, production required that the static semantic representations correctly activate the highly plastic phonological representations. Faced with moving targets, productive performance was at a distinct disadvantage, even when activation was artificially restricted to L1 candidates and criteria were loosened to allow for "close enough" matches.

Although the degree and rate of decline for production may be exaggerated by the model, this finding does reinforce the dissociation of receptive versus productive abilities. Due to this dissociation, studies which primarily measure productive errors in speakers undergoing L1 attrition may overestimate the degree of loss. Stolberg and Münch (2010) found that lexical/semantic production errors decreased by approximately half over the course of 15 conversations in the subject's L1. In light of the dissociation between comprehension and production, the degree to which these errors represent receptive L1 lexical attrition remains in question.

The relearning demonstrated in Stolberg and Münch's study points to the possibility of persistent, though temporarily inaccessible L1 representations. Results from the described model suggest that these representations do persist in memory, reactivated with relatively little practice long after becoming unavailable for production. De Bot et al (2004) confirmed the presence of latent lexical representations in the L2 through a short term relearning task. Foreign language students showed a relearning advantage for words to which they had been previously exposed but forgotten over learning new words. Our model stands to bridge these studies by demonstrating that these latent representations may also explain the observed L1 lexical attrition phenomena, further guiding L1 attrition studies toward seeking L1 representations that may have fallen below the threshold of retrieval for production.

The model also exhibited a highly plausible decay function for first language lexical comprehension. Previously only retrospective analyses, such as that by Hutz (2004), have been available for lexical attrition across a lifetime. Semantic transfer errors identified in Hutz's case study (e.g. *"Das ist feine mit mir"* which is a literal translation of the English idiom "That's fine with me", rather than the equivalent German idiom "*damit bin ich einverstanden*") grew at a diminishing rate over 55 years. The decay of comprehension in this model is highly compatible with Hutz's findings in semantic transfer, indicating that the model's performance curves may represent a component of the generalized time course for L1 lexical attrition.

Moreover, variation of age of onset revealed a possible inverse relationship with the rate at which the comprehension decay occurred. Particularly in the 50 AoO condition, we observed attrition occurring at a higher rate than for later AoOs. This rate, coupled with the slightly lower L1 pre-attrition performance (92% versus 97%), points to the effects of incomplete learning for early onset attrition. Empirical studies have shown that early rather than late exposure to L2 may lead to stronger influence from L2 to L1, causing certain elements of L1 to give way to L2 patterns more easily (e.g., in object naming patterns and categorization; see Pavlenko & Malt, in press). On the other hand, the stronger AoO effects at early stages may be accounted for by the substantial brain plasticity for new languages within the critical period (Pallier et al., 2003).

In the model, it is apparent that the importance of AoO is diminished in cases of later onset. The ostensible leveling-off may be attributable to the limitations placed on Hebbian entrenchment by the normalization. The strength of early AoO effects and high variability in later AoOs reflects Johnson and Newport's (1989) observation of age-related effects in second language acquisition. Like Johnson and Newport's data, our findings are at best ambiguous about the role of age in late second language onset. To what degree the performance of our model was due to incomplete L1 learning versus age-related acceleration of decay requires further investigation.

Conclusion

In empirical literature the study of language attrition has remained a qualitative and descriptive enterprise, due to the

lack of rigorous experimental methodologies for reliably measuring change across time. Coupled with the difficulty of finding a sufficient number of language users who experience L1 language attrition, the extant research makes it difficult to identify any time course of development. In this study, we provided a SOM-based computational model of lexical attrition as a first attempt to systematically investigate mechanisms of language attrition. Specifically, our model is able to produce a gradual decline in L1 lexical performance, suggesting a plausible course of decay in first language comprehension that is compatible with the observations of existing case studies. Furthermore, our model highlights the potential for independent effects on comprehension and production within a single language user. Finally, our model shows age of onset effects in relation to the rate of attrition and points to the possible role of incomplete L1 learning. Such effects are important for understanding the dynamic changes in the competition of two languages during learning.

References

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

de Bot, K., Martens, V., & Stoessel, S. (2004). Finding residual lexical knowledge: The Savings approach to testing vocabulary. *International Journal of Bilingualism, 8* (3), 373-382.

Hao, M., Shu, H., Xing, A., & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods, 40* (3), 728-733.

Hebb, D. (1949). *The organization of behavior: A neuropsychological theory*. New York, NY: Wiley.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133* (4), 638-650.

Hutz, M. (2004). Is there a natural process of decay? A longitudinal study of language attrition. In M. S. Schmid, B. Köpke, M. Keijzer, & L. Weilemar, *First language attrition: Interdisciplinary perspectives on methodological issue*. Amsterdam: John Benjamins Publishing Company.

Hyltenstam, K., Bylund, E., Abrahamsson, N., & Park, H. (2009). Dominant-language replacement: The case of international adoptees. *Bilingualism: Language and Cognition, 12*(2), 121-140.

Johnson, J., & Newport, E. (1989). Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. *Cognitive Psychology, 21*, 60-99.

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

Li, P. (2009). Lexical Organization and Competition in First and Second Languages: Computational and Neural Mechanisms. *Cognitive Science, 33*, 629-664.

Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In R. Heredia & J. Altarriba, *Bilingual sentence processing, 17,* 59–85. North-Holland.

Li, P., & MacWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods Instruments and Computers, 34* (3), 408–415.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic Self-Organization and Early Lexical Development in Children. *Cognitive Science, 31* (4), 581-612.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. De Groot, *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press.

Mayor, J., & Plunkett, K. (2010). A neuro-computational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*. (in press).

Meara, P. (2004). Modelling vocabulary loss. *Applied linguistics, 25* (2), 137–155.

Pallier, C., Dehaene, S., Poline, J.-B., LeBihan, D., Argenti, A.-M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cerebral cortex, 13* (2), 155-61.

Pavlenko, A. & Malt, B. (2010). Kitchen Russian: Cross-linguistic differences and first language object naming by Russian-English bilinguals. *Bilingualism: Language and Cognition*, (in press).

Richardson, F. M., & Thomas, M. S. (2008). Critical periods and catastrophic interference effects in the development of self-organizing feature maps. *Developmental science, 11* (3), 371-89.

Schmitt, E. (2010). When boundaries are crossed: Evaluating language attrition data from two perspectives. *Bilingualism: Language and Cognition*, 13 (1), 63-72.

Stolberg, D., & Münch, A. (2010). "Die Muttersprache vergisst man nicht" –or do you? A case study in L1 attrition and its (partial) reversal,. *Bilingualism: Language and Cognition, 13* (1), 19-31.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition, 4* (2).

Zhao, X., & Li, P. (2007). Bilingual lexical representation in a self-organizing neural network. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Zhao, X., & Li, P. (2009). An online database of phonological representations for Mandarin Chinese. *Behavior Research Methods, 41* (2), 575-83.