Word Learning and Verbal Short-Term Memory: A Computational Account

Prahlad Gupta
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign
Urbana, IL 61801
prahlad@uiuc.edu

Abstract

Recent behavioral evidence suggests that human vocabulary acquisition processes and verbal short-term memory abilities may be related (Gathercole & Baddeley, 1993). Investigation of this relationship has considerable significance for understanding of human language, of working memory, and of the relationship between short- and long-term memory systems. This paper presents a computational model of word learning, nonword repetition, and immediate serial recall. By providing an integrated account of these three abilities, the model provides a specification of how the mechanisms of immediate serial recall may be related to mechanisms of language processing more generally. Furthermore, the model provides fresh insight into the observed behavioral correlations between word learning and immediate serial recall. According to the model, these correlations can arise because of the common dependence of these two abilities on core phonological and semantic processing mechanisms. This contrasts with the explanation proposed in the working memory literature, viz., that word learning is dependent on verbal short-term memory (Gathercole et al., 1992). It is discussed how both explanations can be reconciled in terms of the present model.

Introduction

A variety of recent evidence suggests that human vocabulary acquisition processes and aspects of human verbal short-term memory may be related. In children, reliable correlations have been obtained between digit span, nonword repetition ability, and vocabulary achievement, even when other possible factors such as age and general intelligence have been factored out (e.g., Gathercole & Baddeley, 1989; Gathercole et al., 1992). Studies of normal adults suggest that factors known to affect verbal short-term memory also interfere with word learning ability (e.g., Papagno et al., 1991). It also appears that there is a population of neuropsychologically impaired patients in whom language function is largely preserved, but who exhibit selective deficits in verbal short-term memory and in word learning ability (Baddeley et al., 1988). It is not possible to describe these studies in detail here (see Gathercole & Baddeley, 1993, for a review). The point is that there is now a considerable body of evidence to suggest that word learning, verbal short-term memory, and nonword repetition are a related triad of abilities.

The studies mentioned above have been conducted within the framework of the *working memory model* (Baddeley, 1986). In that model, one subsystem of working memory is verbal short-term memory. This subsystem has been termed the "articulatory loop", and its study has relied on immediate serial recall (ISR) tasks, in which a subject is presented with sequences of unrelated verbal items (such as digits or words), and is required to recall the sequence in correct order, immediately following its presentation. The articulatory loop consists of two parts. One part consists of a *phonological store* for verbal material, within which memory traces decay within 1-2 seconds. The second part consists of mechanisms that enable *rehearsal*, a process that can "refresh" decaying traces in the phonological store (Baddeley, 1986).

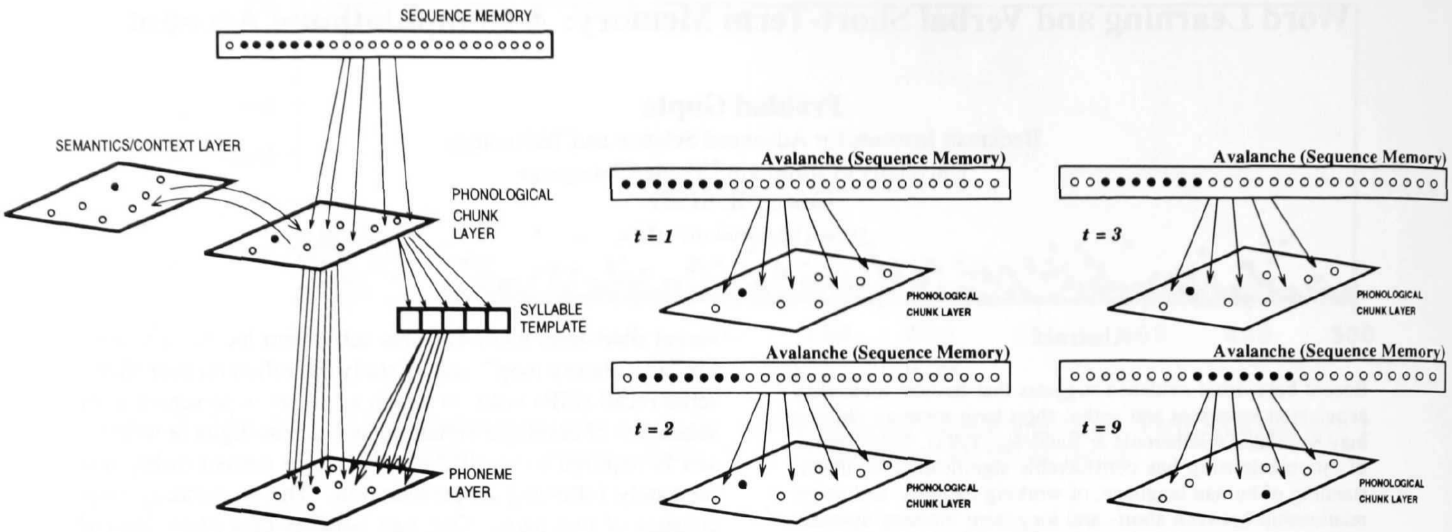
Within this paradigm, the relationship between word learning and verbal short-term memory has been interpreted as indicating that the articulatory loop, and in particular, the phonological store, underlies vocabulary learning (e.g., Gathercole et al., 1992). However, this conjecture has not been elaborated in processing terms, and it is unclear what the nature of such shared processing might be. This paper presents a computational model that attempts to specify in detail what the relationship might be between word learning, nonword repetition, and immediate serial recall.

Such investigation of shared mechanisms underlying verbal short-term memory and vocabulary acquisition is important for at least two reasons. First, it offers a new processing-oriented approach to examining vocabulary acquisition. Second, exploration of this connection can illuminate the relations between short- and long-term memory systems.

A Computational Model

Architecture and computational mechanisms. The goal of the model described in this paper is to examine the computational basis of the relationship between word learning, nonword repetition, and immediate serial recall. The tasks for the model, therefore, are to simulate (1) immediate repetition of novel word forms, (2) the learning of novel word forms, and (3) immediate serial recall of lists of known word forms. In attempting this, the model builds on many previous ideas (Burgess & Hitch, 1992; Grossberg, 1978; Hartley & Houghton, in press; Houghton, 1990), each of which addresses certain, but not all, aspects of the target phenomena.

The model is depicted in Figure 1(a). There are three crucial levels of representation. The Phoneme Layer is a level of output phonology at which phonemes are represented. At this



(a) (b)

Figure 1: A model of word learning and immediate serial recall. (a) Architecture of the model. (b) Sequencing mechanisms.

level, for example, there are representations of the phonemes /a/, /t/, and so on. Second, there is a level at which word forms are represented (the Phonological Chunk Layer). At this level, there are representations for word forms; these representations are shared by input and output phonology. A third level represents semantic and/or contextual information about word forms, and is designated the Semantics/Context Layer. Information about the meaning of the word form *dog* is represented at this level, as also information about contexts of usage in which the word form *dog* has been encountered. In fact, semantics is viewed loosely as a special case of context.

These levels of representation are related via connection weights. The Semantics/Context Layer is bidirectionally connected to the Phonological Chunk Layer, so that representations at these two levels influence each other interactively. The Phonological Chunk Layer has connection weights to the Phoneme Layer. Production of a word form is a serially ordered process, therefore the representation of a word form at the Phonological Chunk Layer has to be able to produce a specific sequence of phonemes at the Phoneme Layer. Each of these levels of representation is comprised of a pool of units, each of which receives input from other units, summates its input, and produces an output which is a sigmoidal function of its summed input. The model also incorporates a *syllable template* between the Phoneme Layer and Chunk Layer, which functions as a parser, assigning syllable structure to the incoming stream of phonemes, and also imposing syllable structure on the output of word forms from the system (Hartley & Houghton, in press).

There is a general sequencing mechanism (designated as the Sequence Memory) that provides immediate memory for sequences of word forms. The Sequence Memory has connections to the Phonological Chunk Layer, which enable it to replay a sequence of activations that have occurred recently at the Phonological Chunk Layer. The connection weights from the Sequence Memory are subject to decay, and therefore the memory for specific sequences is short-lived. The Sequence Memory is the present model's version of the *phonological store* postulated by the working memory model (Baddeley, 1986). The basic mechanism I adopt for such sequence memory is the *avalanche* (Grossberg, 1978). A variant of this mechanism has been incorporated in Houghton's (1990) Competitive Queueing (CQ) architecture, and used in its CQ form by Burgess & Hitch (1992), Hartley & Houghton (in press), and others.

An avalanche is composed of an array of units, as shown for the Sequence Memory in Figure 1(a). A crucial property of an avalanche is that a wave of activation propagates along this pool of units, activating them in a specific and replicable sequence. Presentation of a sequence of word forms to the system is modeled as a sequence of activations of the appropriate word form units at the Phonological Chunk Layer. Concurrently with the sequence of activations at the Chunk Layer, the wave of activation propagates along the Sequence Memory avalanche, and at each time step, connection weights from the avalanche units to the Chunk Layer are adjusted by a Hebbian process. This means that each Sequence Memory avalanche unit encodes whatever pattern of activation was present over the Chunk Layer at the time step(s) when that avalanche unit was active. The process of recall requires that the wave of activation must travel along the avalanche once again. When it does, each avalanche unit will recreate its encoded pattern of activations over the Chunk Layer, provided the connections weights from the avalanche to the Chunk Layer have not decayed too much.

Figure 1(b) depicts several time steps of processing during presentation of a sequence of inputs to the Chunk Layer. The figure shows the wave of activation propagating along the Sequence Memory avalanche units over several time steps of presentation. At each time step, Hebbian adjustment occurs on the connection weights from the avalanche to the Chunk Layer. At recall, reinstantiation of the wave of activation over the avalanche units will lead to production of the sequence of Chunk Layer activations.

Thus the Sequence Memory (SM) shown at the top of Figure 1(a) is implemented as an avalanche that encodes sequences of word forms. In addition, each node at the Phonological Chunk Layer also represents an avalanche. Thus phonemes are bound to a Chunk Layer avalanche in much the same way as Chunk Layer nodes are bound to the SM avalanche. This constitutes a second level of sequencing, whereby the Chunk Layer representation of a particular word form can encode and reproduce the serial ordering of its constituent phonemes.

The model assumes the existence of *word recognition* processes, input from which causes activation of one Phoneme Layer node at a time, and one Chunk Layer node at a time, at each time point in processing, during presentation of word forms.

Simulations

Simulation of word learning. As an example of word learning in the model, let us consider how the novel word form /zæt/ is learned (it may be helpful to refer to Figure 1(a) as needed). Learning occurs during presentation of the word form, as follows. Propagation of activation along the Sequence Memory (SM) avalanche units is initiated. At time step 1, the phoneme unit for /z/ is activated (at the Phoneme Layer). The appropriate Syllable Template node is activated. A new node is allocated at the Phonological Chunk Layer. A new node is also allocated at the Semantics/Context Layer. The following automatic processes occur: (1) Hebbian adjustment of Chunk → Phoneme weights, and Chunk → Template weights. (2) Hebbian adjustment of SM → Chunk weights. (3) Hebbian adjustment of Chunk ↔ Semantics/Context weights. (4) Decay of Chunk → Phoneme weights and Chunk → Template weights. (5) Decay of SM → Chunk weights.

At the next time step, the phoneme unit for /z/ is inactivated, and the phoneme unit for /æ/ is activated. The same automatic processes take place. This procedure is repeated at presentation of /t/. These processes provide the basis for word learning. Note that there are two aspects of learning. First, learning increases the Chunk ↔ Semantics/Context weights. For novel words, these weights are developed on the fly during presentation, and will be lower than those that have developed (over multiple exposures) for known words. Second, learning results in the development of Chunk → Phoneme weights and Chunk → Template weights. For novel word forms, these weights are also developed online during presentation. There is no difference in the magnitude of these weights for known and novel word forms. However, for known words, these weights are assumed to have saturated, and neither increase nor decay during presentation. For novel word forms, these weights are subject to decay. Thus known words have higher Chunk ↔ Semantics/Context than do nonwords; and the Chunk → Phoneme weights and Chunk → Template weights do not change, for known words, whereas they decay for nonwords. These two effects characterize the difference between words and nonwords in the model.

Note also that the automatic processes described above occur irrespective of whether /zæt/ is a known or novel word form. That is, processing during presentation of a *known* word form is identical to that during presentation of a nonword – except that, as noted above, for the already known word, the Chunk → Phoneme weights and Chunk → Template weights will neither increase nor decay.

Simulations of word learning were set up to model a cued-recall task. Ten two-syllable, three-syllable, and four-syllable word forms were used, taken from the Children's Test of Nonword Repetition (Gathercole et al., 1994). 30 simulations were run at each word length. During presentation of each set of word forms (i.e., all the word forms at a particular word length), word learning occurred, as described above. As part of this process, a semantics node was created for each word form. After all the word forms had been presented once, cued recall was tested. The semantics nodes associated with each word in the set were activated, one at a time, representing the cueing of a word with its semantics. Interactive activation then resulted in the activation of chunk nodes. At each time step during recall, activations of the semantics/context units, the chunk units, and the phoneme units were updated. Correct recall of a word would require that the correct sequence of phonemes be produced.

All the words were learned on the first presentation. At all three word lengths, performance was perfect on the very first cued recall trial. Thus the model exhibits word learning performance corresponding to human subjects' abilities to learn words within a very few presentations (Carey, 1978; Dollaghan, 1985).

Simulation of nonword repetition. The automatic processes that occur when a novel word form is presented to the system have already been described as part of the foregoing discussion of word learning. Testing whether that novel word has been *learned* involves testing whether the Semantics/Context → Chunk weights are sufficiently strong to produce the correct Chunk Layer pattern of activation, in the absence of any support from the Sequence Memory, whose weights are assumed to have decayed beyond the point where they can contribute to the chunk retrieval process; and whether the retrieved Chunk Layer pattern can spell out its constituent sequence of phonemes at the Phoneme Layer. Testing *immediate repetition* of the nonword, by contrast, involves testing whether the SM → Chunk weights are still strong enough to allow retrieval of the correct Chunk Layer pattern of activation, and hence activation of the correct sequence of phonemes at the Phoneme Layer.

Simulations were run of immediate repetition of the same 30 nonwords used in the word learning simulations, with 30

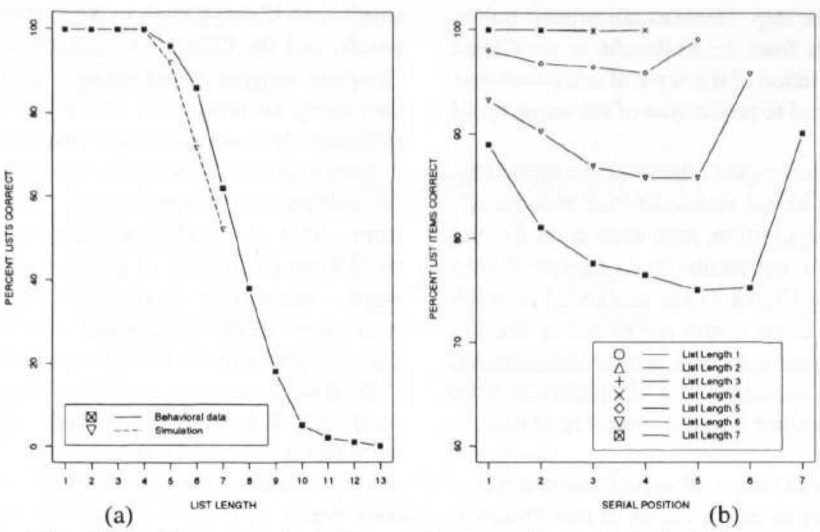


Figure 2: Performance on ISR of lists of known words. (a) Percent lists correctly recalled: simulation, and behavioral data (Guildford & Dallenbach, 1925). (b) Serial position curves (simulation).

attempts to repeat each set of 10 nonwords. Repetition performance was 100% correct at all three syllable lengths. This corresponds to essentially error-free performance in normal adult human subjects. The errors that adults do make in nonword repetition are due to uncertainties in the perception of nonwords, rather than than to any difficulty in being able to repeat them. The model has no analogue of imperfect perception, and so does not make any errors in repetition, even of unfamiliar word forms.

Simulation of immediate serial recall. Immediate serial recall was simulated by presenting sequences of word forms to the model, and then examining recall, when the Sequence Memory avalanche was re-activated immediately following presentation. Figure 2(a) shows the simulated proportion of lists correctly recalled at each list length, as well as the human data (Guildford & Dallenbach, 1925). Figure 2(b) shows simulated serial position curves for ISR of lists of known words, for list lengths 1 through 7. For list lengths 5 and above, the curves show the characteristic primacy and recency effects. Note that the model's performance is perfect for list lengths 1 through 4, as shown in Figure 2(a); therefore in Figure 2(b), the serial position curves for these four list lengths show 100% correct performance at all serial positions, and are superimposed.

It should be noted that what the model simulates is serial recall abilities that are automatic and non-strategic. These are the abilities involved when subjects report "reading out" the recall sequence from memory, as they commonly do for lists of up to 5 digits. Such memory involves no rehearsal. It corresponds to the initial part of the serial recall curve, where performance is almost perfect for up to approximately 5 digits (Guildford & Dallenbach, 1925). As can be seen from Figure 2(a), the model's performance drops below that of human subjects for lists beyond length 4. This reflects the fact that human performance beyond list length 4 or 5 is increasingly dependent on strategies that are not incorporated in the current model.

Word Learning and Serial Recall: The Relationship

We have seen that the model's performance of immediate serial recall (ISR), nonword repetition (NWR), and word learning (WL) is in agreement with human behavioral data. Let us now examine the relationship between these abilities in greater detail. There are four structural components of the model: (1) the Sequence Memory (SM); (2) the Chunk Layer; (3) the Phoneme Layer and Syllable Template (which I classify as one component); and (4) the Semantics/Context Layer. The model's performance in ISR, NWR, and WL depends on the effectiveness of the mappings between these components.

In the model, the effectiveness of these mappings depends on the strengths of connection weights between these components, and the magnitude of the weights in turn depends on the rate of weight change (the learning rate) for each of these sets of connections. These learning rates are parameters of the model that have been set so that the model approximates normal adult human performance, as described in the various simulations so far. The learning rates will be denoted by α for the weights from the Chunk Layer to the Phoneme Layer and Syllable Template, together designated as the Chunk $\rightarrow$ Phoneme weights; θ for the SM $\rightarrow$ Chunk weights; and γ for the Chunk $\leftrightarrow$ Semantics/Context weights.

To understand how NWR, WL, and ISR might be differentially dependent on the various components of the model, the effectiveness of each component was varied, by varying the associated learning rate. In all simulations described earlier, the values of learning rates were $\alpha = 0.43, \theta = 1.06, \gamma = 0.3$. The value of each of these learning rate parameters was now varied, one at a time. For example, Figure 3(a) shows the effect of varying α; the seven combinations of learning rate shown were obtained by combining seven different levels of α with the fixed values $\theta = 1.06, \gamma = 0.3$. The seven learning

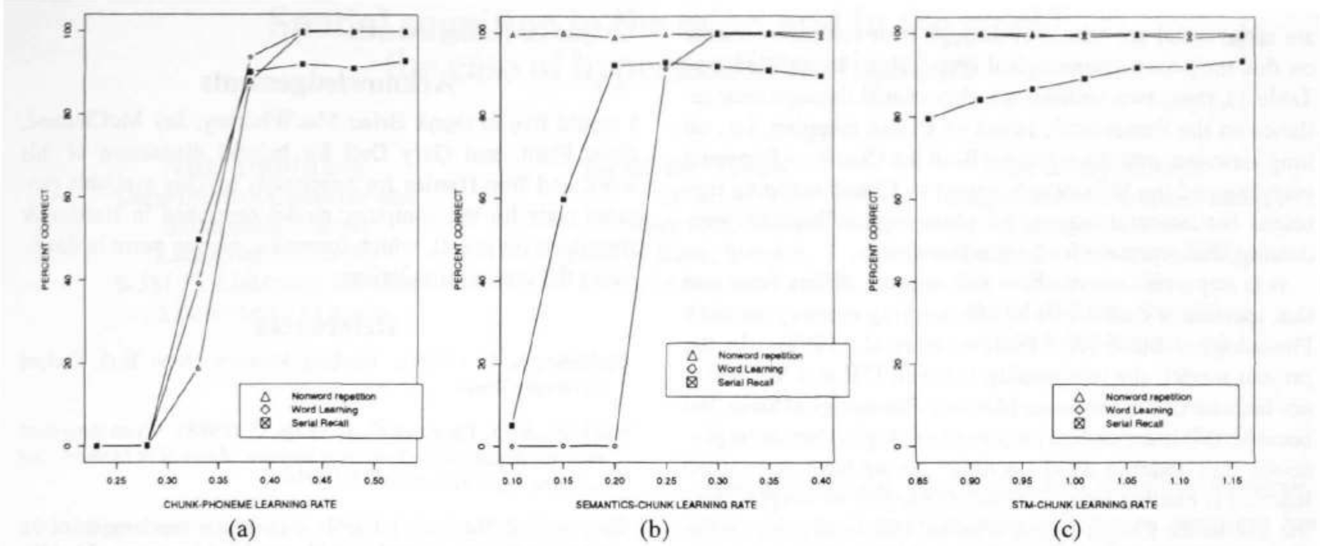


Figure 3: Effect of various learning rates on Nonword Repetition, Word Learning, and Immediate Serial Recall. (a) Effect of learning rate α of weights from Chunk Layer to Phoneme Layer and Syllable Template. (b) Effect of learning rate γ of bidirectional weights between Chunk Layer and Semantics/Context Layer. (c) Effect of learning rate θ of weights from SM to Chunk Layer.

rate combinations in Figures 3(b) and (c) were constructed in analogous fashion, varying either γ or θ respectively, in combination with fixed values of the other two parameters.

For NWR, each data point represents the proportion of 4-syllable nonwords repeated correctly, averaged over 30 simulations of NWR at the particular combination of learning rates. For WL, each data point represents the proportion of novel 4-syllable word forms produced correctly in cued recall, averaged over 30 simulations at the particular combination of learning rates. For ISR, each data point represents the proportion of lists of 5 known words correctly recalled, averaged over 30 simulations.

As can be seen from Figure 3, nonword repetition is dependent on the rate of change (α) of weights from the Chunk Layer to the Phoneme Layer and Syllable Template. However, it is unaffected by the other learning rates. Word learning is sensitive to α, and also to the rate of change (γ) of bidirectional weights between the Chunk Layer and the Semantics/Context Layer. However, it is unaffected by the rate of change (θ) of weights from the SM to the Chunk Layer. Immediate Serial Recall is dependent on all three rates of weight change. These dependencies are summarized in Table 1.

To consider what these dependencies indicate, it is worth reiterating what the variations in "learning rates" represent. The Chunk $\rightarrow$ Phoneme mapping represents long-term phonological knowledge (knowledge about the serial order of phonemes within words). The effectiveness of this knowledge is represented by the strength of connection weights in this component of the model. In the model, the strength of these connection weights depends on the learning rate α. Variation in the learning rate parameter α is thus a shorthand for variation in the strength or effectiveness of long-term phonological knowledge in the system. Similarly, variation of the learning rate parameter γ represents variation in the strength of long-term knowledge about the semantics of words in the system. Variation in the learning rate parameter θ represents variation in the inherent capacity of the *short-term* Sequence Memory.

Ability	Ability is affected by:		
	Chunk $\rightarrow$ Phoneme learning rate (α)	Context $\leftrightarrow$ Chunk learning rate (γ)	SM $\rightarrow$ Chunk learning rate (θ)
Nonword repetition	$\surd$	X	X
Word Learning	$\surd$	$\surd$	X
Immediate Serial recall	$\surd$	$\surd$	$\surd$

Table 1: Analysis of dependencies in the model: Factors influencing Nonword Repetition (NWR), Word Learning (WL), and Immediate Serial Recall (ISR).

The dependencies in Table 1 thus indicate that ISR ability is affected by the effectiveness of long-term phonological knowledge, long-term semantic knowledge, and short-term phonological store capacity. Word learning ability is affected by the strength of long-term phonological knowledge and long-term semantic knowledge. Nonword repetition ability is affected by the strength of long-term phonological knowledge.

Conclusions

These results suggest a more precise specification of the relationship between NWR, WL, and ISR. In the model, these abilities are related because they are all dependent on the effectiveness of the Chunk $\rightarrow$ Phoneme mapping. That is, the model ascribes the relationship between these three abilities to their common reliance on long-term phonological knowledge about phoneme serial ordering.

Thus one of the ways in which word learning and immediate serial recall are related is through their common reliance on this long-term phonological knowledge. In addition (see Table 1), these two abilities are also related through their reliance on the Semantics/Context $\leftrightarrow$ Chunk mapping, i.e., on long-term semantic knowledge. Both the Chunk $\rightarrow$ Phoneme mapping and the Semantics/Context $\leftrightarrow$ Chunk mapping represent fundamental aspects of phonological/linguistic processing, and represent long-term knowledge.

It is important to note how this account differs from one that ascribes a causal role to the working memory model's Phonological Store (cf. Gathercole et al., 1992). In the present model, the relationship between ISR and WL arises *not* because of the Sequence Memory/Phonological Store, but because ISR is dependent on core phonological/semantic processes that underlie word learning. As we have seen, word learning is unaffected by the rate of change θ of weights from the SM to the Chunk Layer, whereas ISR is affected by the effectiveness of the Chunk $\rightarrow$ Phoneme mapping, and by the effectiveness of the Semantics/Context $\leftrightarrow$ Chunk mapping.

If the present model were to incorporate connectivity from the Sequence Memory directly to the Phoneme Layer, it is likely that this would aid the effectiveness of NWR, WL, and ISR. All three abilities would then be dependent on a mechanism (the Sequence Memory/Phonological Store) that primarily subserves ISR. Thus the present model does not rule out the possibility of a causal role for the phonological store in word learning. Rather, it suggests that NWR, ISR and WL may be related even in the absence of such a connection.

It is perhaps most likely that these abilities are related in both ways. This would help to explain the findings that, whereas nonword repetition ability is more predictive of vocabulary ability (than vice versa) at earlier ages, vocabulary ability becomes more predictive of nonword repetition ability and ISR ability (than vice versa) by about 8 years of age (Gathercole et al., 1992). This would be explicable in the present model in that word learning might benefit from the support of Sequence Memory $\rightarrow$ Phoneme Layer connections more when the effectiveness of the Chunk $\rightarrow$ Phoneme mapping is low than when it is high. A similar explanation has been proposed by Gathercole et al. (1992). The present model provides a computational basis for seeing why this might be so.

Of course, word learning and ISR may *also* be related through their common reliance on rehearsal mechanisms. According to this view, in immediate serial recall, rehearsal can help maintain phonological representations of the recall stimuli in an active state. In vocabulary acquisition, likewise, rehearsal may aid the formation of new phonological representations, by allowing the learner repeated access to the word form so as to consolidate the new memory.

In conclusion, in its present form, the model makes the important demonstration that word learning and immediate serial recall may be related at the level of core phonological processing. Using the model to investigate how these abilities might be related via the phonological store and via rehearsal is a goal for further research.

Acknowledgements

I would like to thank Brian MacWhinney, Jay McClelland, Dave Plaut, and Gary Dell for helpful discussion of this work, and Tom Hartley for generously making available program code for the computer model described in Hartley & Houghton (in press), which formed a starting point in developing the present simulations.

References

Baddeley, A. D. (1986). *Working Memory.* New York, Oxford University Press.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27, 586–595.

Burgess, N. & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, 31, 429–460.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic Theory and Psychological Reality.* Cambridge, MA, MIT Press.

Dollaghan, C. (1985). Child meets word: "fast mapping" in preschool children. *Journal of Speech and Hearing Research*, 28, 449–454.

Gathercole, S. E. & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28, 200–213.

Gathercole, S. E. & Baddeley, A. D. (1993). *Working Memory and Language*. Hillsdale, NJ, Lawrence Erlbaum.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28, 887–898.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, 2, 103–127.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in Theoretical Biology*, Volume 5. New York, Academic Press.

Guildford, J. P. & Dallenbach, K. M. (1925). The determination of memory span by the method of constant stimuli. *American Journal of Psychology*, 36, 621–628.

Hartley, T. & Houghton, G. (in press). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current Research in Natural Language Generation*. New York, Academic Press.

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language learning. *Journal of Memory and Language*, 30, 331–347.