In A. Thorn & M. Page (Eds.), Interactions Between Short-Term and Long-Term Memory in the Verbal Domain, pages 108-135. Hove, England, Psychology Press.

6 A computational model of nonword repetition, immediate serial recall, and nonword learning

Prahlad Gupta

Overview

A computational model is presented that tackles the issue of serial ordering both across word forms (as in list recall) and within word forms (as in nonword repetition). The model was set up to simulate performance in these two tasks. A number of patterns of relationship between performance in these tasks then fell out of the model, including the well-documented correlation between performance in the two tasks, and the typical profile of impairment in the "pure STM" neuropsychological syndrome. The model thus provides a concretely instantiated means of thinking about relationships between list recall and nonword repetition, and has the potential to play a useful role in investigation of these relationships.

Introduction

There is now abundant evidence to suggest that immediate list recall ability, nonword repetition ability, and the learning of new words are related in some way. There is also a measure of agreement that immediate serial recall (ISR) and nonword repetition are both tasks that draw on the mechanisms of phonological short-term memory fairly directly, and that the learning of new words is also in some way supported by phonological short-term memory (e.g., Baddeley, Gathercole, & Papagno, 1998; Brown & Hulme, 1996; Gathercole, Service, Hitch, Adams, & Martin, 1999; Gupta, 2003; Gupta & MacWhinney, 1997). Insofar as this is the case, these abilities constitute a fascinating domain of interaction between short-term and long-term learning and memory systems, and, moreover, one that is fundamental to some of the most centrally human cognitive abilities. However, there is no current widely-accepted mechanistic account of the observed patterns of relationship.

From the point of view of an individual learner, every new word is in effect a nonword when first encountered, and every known word was once a nonword to that learner. Greater facility in processing nonwords would therefore be expected to lead to greater facility in eventually learning them,

thus providing intuition for why there might be a relationship between nonword repetition and word learning. But why are immediate list recall and nonword repetition related? – in what sense might immediate repetition of a nonword be a phonological short-term memory task? In thinking about this, it is useful to consider that an auditorily presented nonword is a novel sequence of sounds, in much the same way that the typical list in an ISR task is a novel sequence of words or digits. This raises the possibility that a nonword may literally be processed like a list (i.e., like a novel sequence of speech units) when it is first encountered (Cumming, Page, & Norris, 2003; Gupta, 1996, 2002, 2005; Gupta & MacWhinney, 1997; Gupta, Lipinski, Abbs, & Lin, 2005; Hartley & Houghton, 1996). If this were the case, it would provide a simple explanation of the relationships observed between immediate serial recall and nonword repetition. Moreover, if mechanisms similar to those underlying immediate serial recall are operative in the repetition of nonwords, we would expect to observe serial position effects in repetition of the sequence of sounds comprising nonwords. Following this reasoning, Gupta (2002, 2005; Gupta et al., 2005) examined immediate repetition of individual auditorily presented polysyllabic nonwords, to determine whether repetition accuracy broken down by syllables within the nonwords would manifest primacy and recency; that is, whether the first and last syllables within the nonwords would be repeated more accurately than middle syllables. In a series of experiments, such primacy and recency effects were indeed obtained in repetition of individual nonwords of lengths four through seven syllables. These results are consistent with the idea that similar serial ordering mechanisms are operative in immediate serial recall of lists and in repetition of nonwords, especially when taken together with the considerable body of evidence indicating an association between nonword repetition and list recall.

These results serve, among other things, to emphasize the fact that some sequencing mechanism is needed to encode, maintain, and retrieve the serial order of these components of a nonword. The question then arises of exactly how this mechanism might be related to the mechanism underlying immediate serial recall of lists. Are these mechanisms identical? If so, how? If not, how do they differ? The present chapter describes a computational model that simulates performance in list recall and nonword repetition tasks, and uses these simulations to offer an account of the empirically observed relationships between performance in these tasks.

Functional requirements for a model

What would a model need to address, in order to provide an account of performance in immediate serial recall as well as nonword repetition? To simulate nonword repetition, a model clearly would need to address the serial ordering of sublexical constituents such as syllables/phonemes *within* word forms. That is, it would need, first, to *represent* such sublexical

constituents, and second, to provide a means for the encoding and retrieval of their *serial order*. Without this, it could not offer an account of how the novel sequence of sounds comprising a nonword such as *zitricaymus* can be immediately repeated, in sequence. To simulate immediate serial recall of lists of verbal items, a model would clearly need to address the serial ordering of the words constituting the list elements. That is, it would need, first, to *represent* these lexical elements, and second, to provide a means for the encoding and retrieval of their *serial order*. Without this, it could not offer an account of how the novel sequence of words comprising a list such as *cat dog ball chair* can be immediately repeated, in sequence.

To simulate performance in both tasks, therefore, a model would need to incorporate both lexical and sublexical representations, and to provide for serial ordering at both these levels of representation. However, these requirements have not so far been addressed by current models. Although there are now several important and insightful computational models of immediate serial recall (e.g., Botvinick & Plaut, 2006; Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1992, 1999; Hartley & Houghton, 1996; Houghton, 1990; Page & Norris, 1998; Vousden, Brown, & Harley, 2000), and indeed, the present work draws importantly on some of these models, most have been concerned primarily with addressing the numerous phenomena of immediate serial recall per se, rather than with the relationship between immediate serial recall and novel word processing. An exception to this is the work of Burgess and Hitch (1999) which did indeed seek to address this relationship (along with an impressive range of phenomena from ISR per se), but, like other models, it did not address the issue of serial ordering at multiple levels of representation. Addressing this issue is, however, a central motivation of the present work (see also Gupta, 1995, 1996; Gupta & MacWhinney, 1997), which is less concerned with achieving wide coverage of the phenomena of immediate serial recall. The aims of the present work can thus be seen as complementary to those of much existing computational modelling work in the domain of phonological short-term memory. Indeed, the central focus of the present work can be seen as lying precisely at the nexus of short-term memory and long-term memory and serial ordering in the verbal domain.

The current model: Overview

The core conceptualization underlying the present model essentially mirrors the preceding discussion of functional requirements: it incorporates lexical and sublexical levels of representation, and a serial ordering mechanism that provides for the encoding, maintenance and retrieval of sequences that are activated at these levels. Thus, as shown in Figure 6.1, the model incorporates word form and semantic levels of lexical representation, and syllabic and phonemic levels of sublexical representation. It may be worth noting that the incorporation of such levels of representation is completely

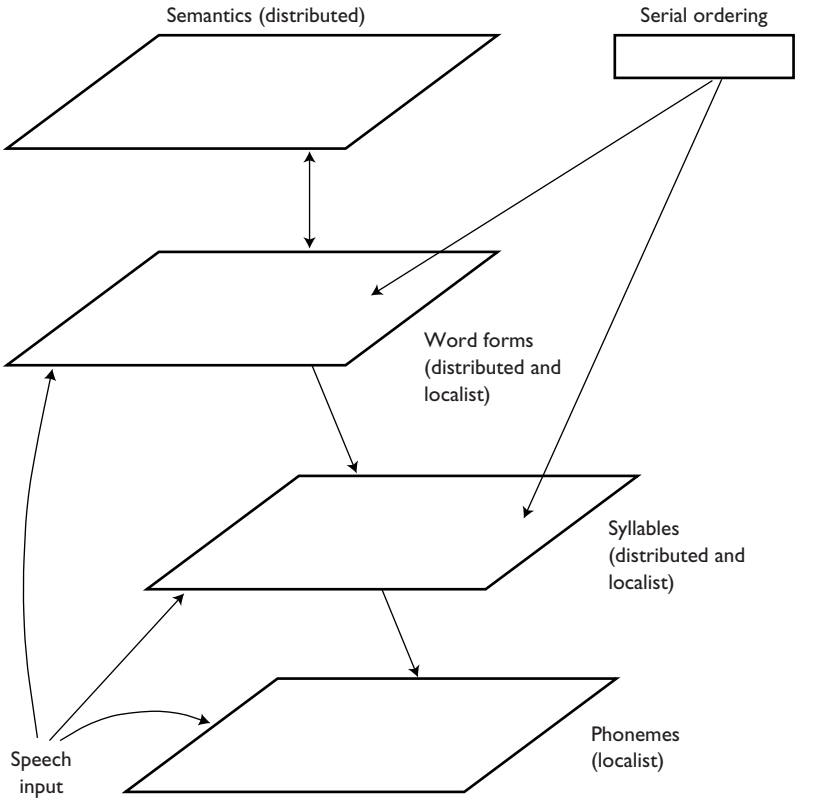

Figure 6.1 Overview of model.

consistent with the assumptions commonly made in models of lexical processing (e.g., Dell, 1986; Levelt, Roelofs, & Meyer, 1999).

Presentation of a word form to the model (depicted by the "speech input" arrows in Figure 6.1) results in sequences of representations being activated at the various levels of representation. For instance, presentation of the word form *zitricaymus* is manifested in the model as activation of the relevant sequence of phoneme representations at the phoneme level, activation of the relevant sequence of syllable representations at the syllable level, and activation of the relevant word form representation at the word form level. New representations are created on the fly as necessary at each of these levels, using mechanisms of the kind proposed by Grossberg (1987). Presentation of either a known word form or a novel word form thus gives rise to sequences of activations at the phoneme and syllable levels, and of a single activation at the word form level. A "list" is simply a sequence of these sequences. For instance, presentation of the list *cat dog ball chair* leads to activation of the sequences of phonemes and syllables constituting each

word in the list, and additionally, gives rise to a sequence of activations at the word form level (the representations of the word forms cat, dog, ball, and chair). Thus in a real sense, a word form is just a list of length one. For both a list or a single word form, the activated representation(s) at the word form level will lead to activations at the semantic level of representation. For known word forms, these will be the specific semantic representation that is associated with that word form, whereas for nonwords the evoked semantic representation will tend to be an indeterminate semantic representation that is a blend of those corresponding to known word forms that are similar to the nonword. Figure 6.1 also shows a "serial ordering mechanism" (to be discussed in greater detail below). This mechanism encodes (and maintains and retrieves) the serial order of a sequence of activations, at any level of representation it is connected to. As shown in the figure, the serial ordering mechanism has connections to both the word form and syllable levels of representation. Thus, the model provides for serial ordering at both lexical and sublexical levels of representation, thus completing the functional requirements outlined in the previous discussion.

The current model: Details

More details of the model's architecture are shown in Figure 6.2. The word form and syllable levels of representation are actually each composed of two sets of representations. One set of representations is *localist*, i.e., there is an individual unit representing the entire entity (a word form, at the word form level, or a syllable, at the syllable level). In the second set of representations at each level, the entity (word form or syllable) is represented as a pattern of activation that is *distributed* across a pool of units, with each unit in the pool representing a feature that comprises the entity; there is no individual unit that represents the whole entity, and this constitutes the critical difference between what are termed localist and distributed representations. The word form and syllable levels thus incorporate *both* localist and distributed representations (see Page, 2000, for a discussion of the merits of each type of representation).

As is typical in such artificial neural network or parallel distributed processing or connectionist models, the "units" are to be thought of as highly abstract neuron-like elements ("artificial neurons"). Such a unit has connections to and from other units. The outgoing connections from a unit are thought of loosely as axonal projections from a neuron. These projections make contact with other units, with such contacts being thought of loosely as synapses. The strength of any given "synapse" from unit *a* to unit *b* in such a model is instantiated as a *weight* on the connection from *a* to *b*. Thus a unit in such a model receives input from other units via the weighted incoming connections, summates its input, and, if the summed input crosses some threshold, transmits an output on its outgoing connections, thus providing input to other units to which it is connected. The input received

by unit *b* from unit *a* is the product of the output emitted by *a* and the weight on the connection from *a* to *b*. The connection weights in such a model at any given point are thought of as encoding the model's *long-term knowledge* at that point. They start at some initial random value, and are adjusted following each stimulus processing event. The gradual adjustment of connection weights in this way modifies the long-term knowledge, and is therefore thought of as *learning*. A variety of weight-adjustment or learning procedures exist for such models.

In the present model, the localist and distributed representations at each level (word form and syllable) are bidirectionally connected, with every unit in the localist pool having a connection to every unit in the distributed pool, and vice versa. The weights on all these connections are adjusted during training of the model, using a Hebbian learning algorithm, in which the adjustment of connection weights is a simple function of the activation of the sending and receiving units (and does not make use of the notion of divergence from some target activation for the receiving unit, which is termed *error*, and would additionally be used in an *error-driven* learning procedure; see McLeod, Plunkett, & Rolls, 1998, for an accessible treatment of such learning procedures). Therefore, after learning, at the word form level, activation of the localist unit representing a word form leads to activation of the corresponding distributed representation, and vice versa; and analogously at the syllable level.

As shown in Figure 6.2, the semantic level of representation is connected to the word form level (specifically, to the distributed representations at the word form level) via an intermediate set of "hidden units". All these connection weights are also learned during training of the model, using a Hebbian learning algorithm. Again, these constitute long-term memory in the model.

The distributed representations at the word form level are the phonologically structured representations of an entire word form. Specifically, the word form is represented as a string of syllables. The representation of a syllable is in terms of a CCVCC (i.e., Consonant-Consonant-Vowel-Consonant-Consonant) template. That is, the representation scheme for a syllable consists of a pool of units divided into five slots. Activation of units in the first slot denotes the first C (if any) of the syllable, activation of units in the second slot denotes the second C (if any) of the syllable, activation of units in the third slot denotes the V of the syllable, and so on. The entire word form is represented as a string of such syllable templates, stringing together as many syllables as needed to represent the whole word form. Thus the word form is represented in a CCVCC-CCVCC-CCVCC . . . format (note that the hyphens are shown here to clarify the fact that there are multiple syllables, but have no counterpart in what is presented to the model).

The long-term connections from the word form level to the syllable level are in fact from the distributed word form representations to the localist

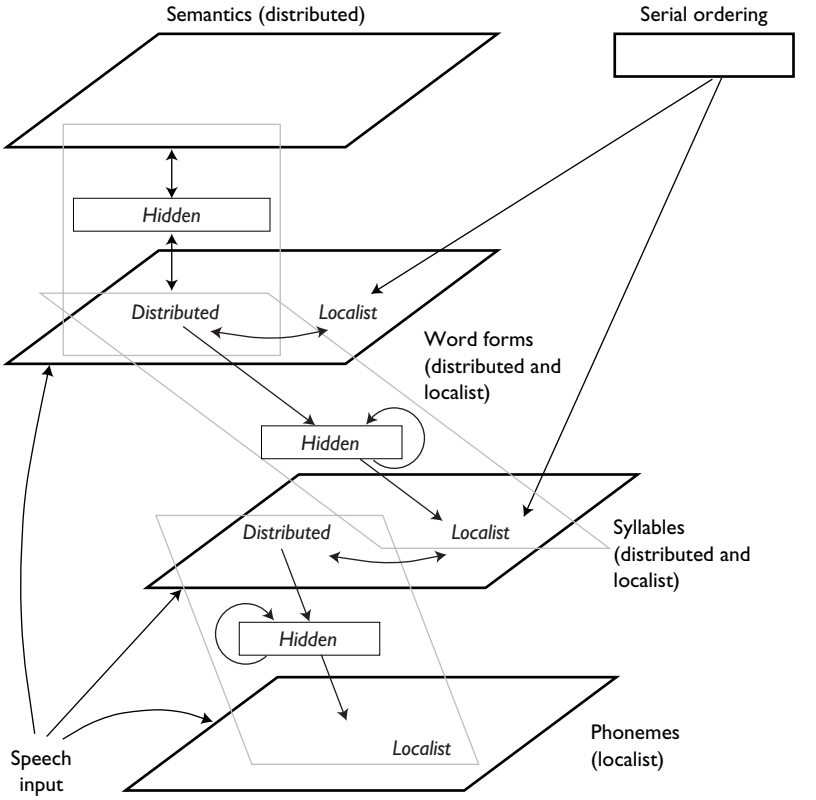

Figure 6.2 Detailed architecture of model.

syllable representations. The distributed representation at the word form level constitutes the input layer of a type of simple recurrent network (SRN; Elman, 1990a; Jordan, 1986) which is shown in the middle parallelogram in Figure 6.2. For present purposes, the crucial feature of such a network is that it can take as its input an unchanging (and hence *non-sequential*) representation of an entire sequence (the "plan" for the entire sequence), and combine this non-sequential input with information about its own sequence of internal states over time, so as to produce *sequential* output that "spells out" the constituent elements of the sequence represented by the plan, in serially ordered fashion, across successive time steps (for further detail on such networks, see Elman, 1990a; Gupta & Cohen, 2002; Jordan, 1986). In the present model, this SRN translates its input representation (the "plan" for the entire word form) into a sequence of localist outputs representing the sequence of syllables constituting that word form. This aspect of the model also constitutes long-term memory or knowledge. The SRN acquires its knowledge through adjustment of connection weights during training of the

model, which employs the back-propagation learning algorithm (an *error-driven* learning procedure; Rumelhart, Hinton, & Williams, 1986), as is typical for an SRN (e.g., Elman, 1990a, 1990b; Jordan, 1986).

The long-term connections from the syllable level to the phoneme level are in fact from the distributed syllable representations to the (localist) phoneme representations. The distributed representations at the syllable level are the phonologically structured representations of an entire syllable, arranged in a CCVCC format. (The representation scheme is identical to that used at the word form level for *each* of the syllables constituting the entire word form.) These constitute the input to a second SRN which is shown in the bottom parallelogram in Figure 6.2. This SRN translates its input representation, the "plan" for the entire syllable, into a sequence of localist outputs representing the sequence of phonemes constituting that syllable. This SRN is also trained via back-propagation, and it also constitutes long-term memory/knowledge in the system.

The serial ordering mechanism is a variant of the *avalanche* (Grossberg, 1978), and is a time-varying context signal of the general kind that has been employed in a number of computational models of verbal short-term memory (e.g., Brown et al., 2000; Burgess & Hitch, 1992, 1999; Hartley & Houghton, 1996; Houghton, 1990; Vousden et al., 2000). It encodes the serial order of linguistic representations as they are activated at the word form and syllable levels. The mechanism is an endogenous generator of a waveform. It automatically sweeps through the same waveform whenever it is reset. A useful metaphor (Brown et al., 2000; Vousden et al., 2000) is that the mechanism is like a "clock", which recreates the same waveform whenever it is reset to the 12:00 position. As the clock proceeds through its waveform, it takes "snapshots" of the activation of linguistic representations as they occur in sequence at the word form and syllable levels of representation as a result of presentation of speech inputs, and can later re-display these snapshots in sequence, thus recalling the previously presented list. The "snapshots" are implemented by decaying (and hence *short-term*) weights on connections to the localist representations at both the word form and syllable levels. Hebbian learning occurs in these short-term connections, when representations are activated in sequence by input at the word form and syllable levels. The mechanism can subsequently cause those sequences of activations to be replayed and thus recalled, as long as the connection weights have not decayed too much. The replaying consists literally of a recreation of the original sequence of activations in the linguistic system (for further detail, see Gupta, 1996; Gupta & MacWhinney, 1997).

The serial ordering device thus constitutes the model's short-term sequence memory. It is important to note, however, that no "copies" of the to-be-remembered item(s) are made or "stored". Rather, the short-term sequence memory is a serial ordering device that sets up associations to a sequence of activations in the linguistic system (which consists of long-term representations). A consequence of this is that the short-term and long-term

aspects of the model are inextricably linked in performance in a task such as immediate list recall or word form repetition. Performance depends critically on *both* the short-term and long-term memory systems in the model.

Development of the model

Development of the model to the present point has proceeded in three phases. The first step has been to construct a computational model (incorporating the architecture discussed in the previous section) that exhibits the basic ability to perform the two tasks of nonword repetition and immediate serial recall. The second step has been to calibrate the model so that its performance on the nonword repetition and immediate serial recall tasks is qualitatively similar to human behavioral performance. The third step has been to employ the model, as thus developed, to examine the relationship between immediate serial recall and nonword repetition.

As a precursor to simulations of nonword repetition and immediate serial recall, the model was "pretrained" to establish pre-existing linguistic knowledge and a known vocabulary. A set of approximately 3000 real words of English one through seven syllables in length was presented to the model. Each of the three components of the model representing long-term linguistic knowledge was trained on this set. The task for the system consisting of the semantic and distributed word form representations (see Figure 6.2) was to produce the correct semantic vector in response to a particular distributed word form representation and vice versa, for each of the 3000 words. Connection weights in this component were adjusted via a Hebbian learning algorithm, to and from the hidden layer localist units.[1] The task for the SRN system shown in the upper parallelogram in Figure 6.2 (consisting of the distributed word form representations and localist syllable representations) was to produce the correct sequence of localist syllable outputs in response to the distributed word form representation for each of the 3000 words. The task for the SRN system shown in the lower parallelogram in Figure 6.2 (consisting of the distributed syllable representations and localist phoneme representations) was to produce the correct sequence of phoneme representations in response to each distributed syllable representation contained in the 3000 words. Connection weights in these two SRNs were adjusted via back-propagation of error in each of their respective tasks. New units were created as necessary at the localist word form and localist syllable levels of representation, and connection weights between the localist and distributed representations at each level were also established during pretraining on the 3000 words. After each epoch of training, the model's performance on each word was tested by presenting the semantic representation for that word to the model's semantic level, allowing the activations to propagate from the semantic level to the word form, syllable, and phoneme levels, and examining the sequence of phonemes that was produced as a result. The word was considered to have

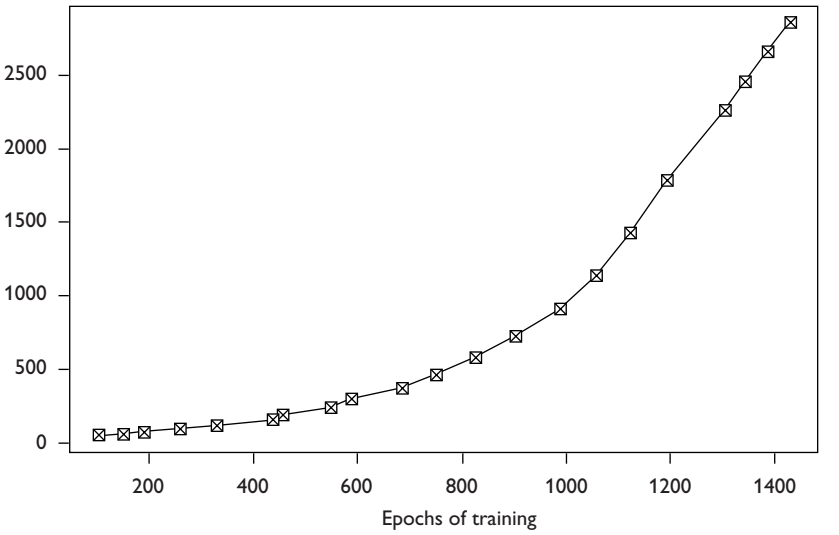

Figure 6.3 Growth of the model's vocabulary during pretraining.

been "learned" if all its phonemes were correctly produced, in correct order. Figure 6.3 shows the progress of vocabulary learning during pretraining. It is worth noting, however, that no claims are made about the psychological plausibility of the pretraining procedure as just described, nor about its correspondence with human vocabulary learning or development. The aim of pretraining was simply to establish a corpus of words that were known to the system (i.e., for which appropriate connection weights existed throughout the long-term components of the model). Achieving this aim provided the model with a vocabulary of known words that could be drawn on for simulated immediate serial recall, and provided a distinction in the model between known words and nonwords, for which there would be no appropriate pre-existing connection weights.

Simulation of nonword repetition and immediate serial recall

The goal of the simulations of nonword repetition (NWR) and immediate serial recall (ISR) was to achieve performance qualitatively similar to human empirical data. Moreover, the aim was to achieve this with a single set of parameters, so that the identical model would be performing each task and exhibiting humanlike behavior in both tasks (rather than have separate versions of the model for each task). The phenomenon chosen as the target for modelling of nonword repetition was the serial position functions that have recently been reported in human nonword repetition (Gupta, 2005; Gupta et al., 2005). For simulation of immediate serial recall, the phenomena chosen as modelling targets were the classic serial position effects in list

recall, list length effects, and the movement error gradients that have been documented for ISR (e.g., Henson, Norris, Page, & Baddeley, 1996). In each case, the aim was simply to achieve qualitatively similar behavior in the model (rather than precise quantitative fits).

Simulations of NWR were conducted by presenting nonwords of lengths one through seven syllables (25 of each length) to the model. The nonwords were drawn from the same corpus used in behavioral investigation of serial position effects in nonword repetition (Gupta et al., 2005). Each nonword was presented individually, and the model's repetition response recorded after each presentation. Each presentation consisted of activation of the appropriate sequence of representations at the syllable and phoneme levels in the model, and of a single representation at the word form level. The word form level representation consisted of a new localist node and the distributed representation of the entire word form. During this presentation, learning occurred in the connections from the word form level to the syllable level, the connections from the syllable level to the phoneme level, and bidirectionally between localist and distributed representations at each of the word form and syllable levels. All of these constituted learning in the long-term components of the model, and embodied the notion that learning occurs constantly in the system. Learning did not occur, however, between the semantic level and the word form level because for nonword representation, there is assumed to be no specific semantic representation that is activated. In addition to learning in these long-term components, learning occurred in the short-term connection weights from the serial ordering mechanism to the word form level (to the new word form that had been created and activated) and to the syllable level (to the sequence of syllable level representations that had been activated). Following this presentation, nonword repetition was simulated by resetting the serial ordering mechanism to its original state, followed by regeneration of its waveform. The recreated waveform, together with the connection weights created during presentation, led to reactivation of the previously presented patterns of activation at the word form and syllable levels.[2] The accuracy of this recall depends on a variety of factors including the extent to which adjacent portions of the waveform overlap in the "snapshots" they take of the word form and syllable levels, the rate of decay of the short-term connection weights, and the level of noise in the activations of word form level and syllable level units. Thus, accuracy of recall is not guaranteed to be perfect, and tends to yield serial position functions for recall at the target level. The model's performance in simulated NWR is shown in Figure 6.4. As can be seen, simulated NWR performance exhibits clear serial position effects, which are qualitatively similar to those observed behaviorally (Gupta et al., 2005).

Simulations of ISR were conducted by presenting lists of 1 to 12 known words (25 lists of each length) to the model, employing exactly the same model and parameters as in simulation of NWR. The lists were composed of one-syllable words drawn from the model's vocabulary. Each list was

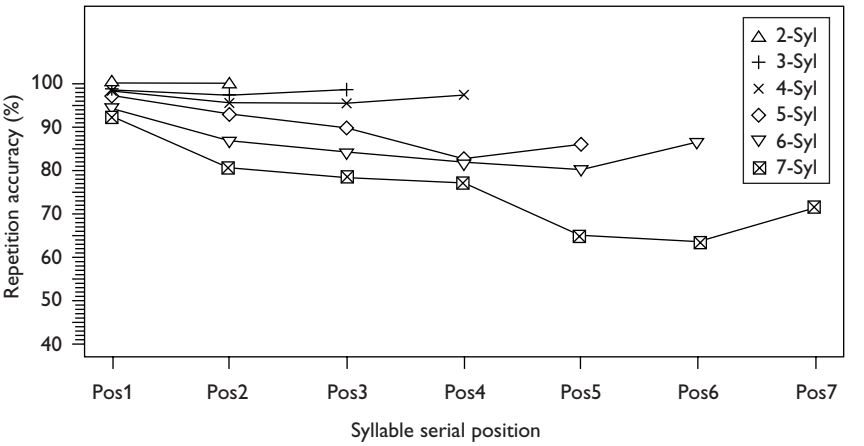

Figure 6.4 Serial position functions from simulated repetition of nonwords of 2 through 7 syllables in length.

presented one word at a time, and the model's recall of the list was recorded at the end of list presentation. Presentation of each word consisted of activation in the model of the appropriate sequence of representations at the phoneme level, and of a single representation at each of the word form and syllable levels. At every time step of this presentation, learning occurred in the short-term connection weights from the serial ordering mechanism to the word form level (to the currently activated word form representation) and to the syllable level (to the currently activated syllable representation).[3]

As with nonword repetition, recall of the list was simulated by resetting the serial ordering mechanism to its original state. Endogenous regeneration of the waveform, together with the connection weights created during presentation, led to reactivation of the previously presented patterns of activation at the word form and syllable levels, with accuracy of recall being dependent on the same factors as discussed above for nonword repetition. Figure 6.5 shows serial position functions for simulated ISR of lists of various lengths. As can be seen, the functions exhibit bowing and primacy and recency effects qualitatively similar to those characteristic of human list recall.

The errors observed in human immediate serial recall also follow characteristic patterns. In particular, when a list item that was presented in a particular serial position N is recalled in an incorrect position, the incorrect position is most likely to be very close to N, and relatively unlikely to be far from N. Thus the probability of an item presented in position N being recalled in an incorrect position M ($M = N \pm e$, $e > 0$) decreases with the distance of M from N, i.e., decreases as a function of e (e.g., Henson et al., 1996). Figure 6.6 shows that the model's ISR performance exhibits this property. The figure depicts movement error gradients for the model's ISR

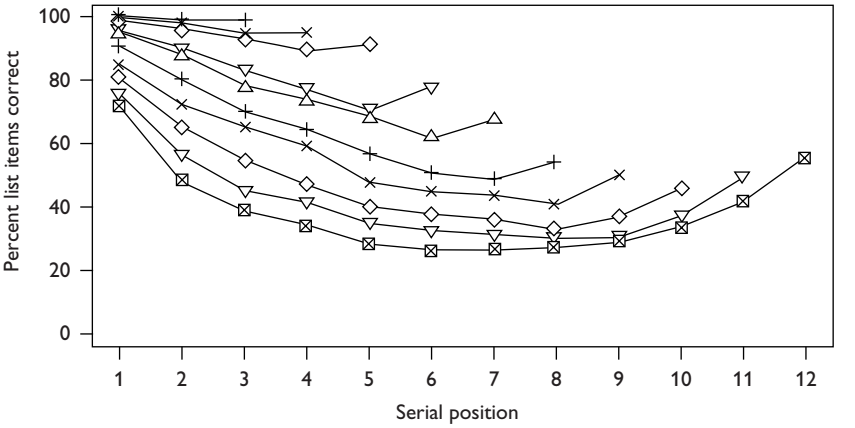

Figure 6.5 Serial position functions from simulated immediate serial recall of lists of 1 through 12 words.

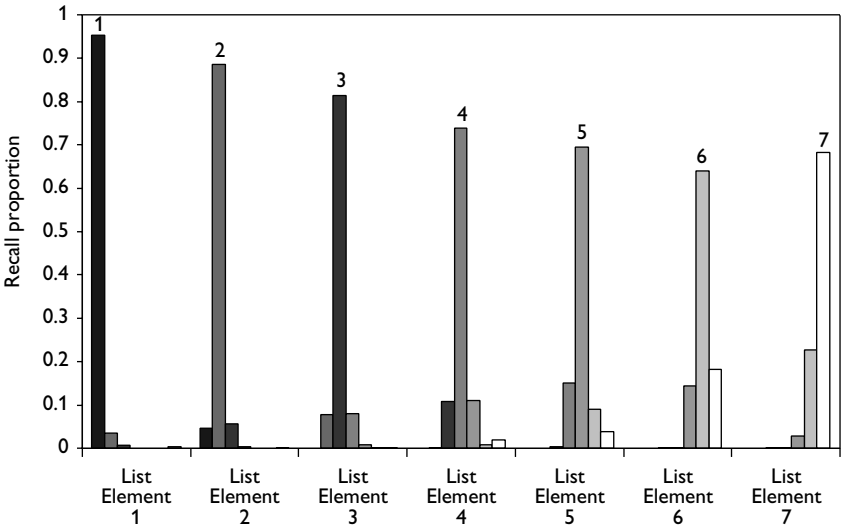

Figure 6.6 Movement error gradients for 7-item lists. Each set of bars arrayed along the x-axis corresponds to the item from a particular serial position in the stimulus list. From left to right within each set, the specific bars indicate the proportion of trials on which that stimulus item was recalled in each of the seven recall positions. (Some bars are not visible because some proportions are negligible.) Numerals at the top of each set of bars indicate, for that stimulus item, the output serial position at which it was most frequently recalled.

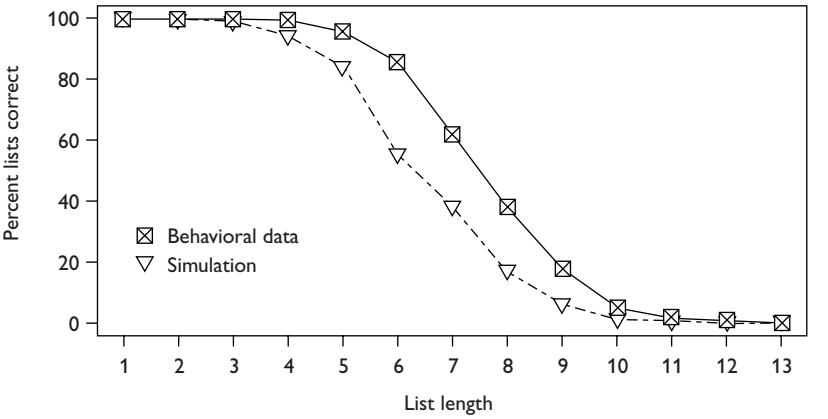

Figure 6.7 Effect of list length on ISR of lists of 1 through 13 words: human data and simulations.

of 7-item lists. Seven sets of bars are arrayed along the x-axis. Each set of bars represents a particular serial position in the target list (i.e., the presented list). The specific bars within each set indicate the frequencies of positions in which that target item was actually recalled. As can be seen, for each target position, the most probable recall position was the correct one. The adjacent positions were next highest in probability (denoting errors of $M = N \pm 1$), and the probability of the item being recalled further away than this was negligible. This profile of movement error gradients is qualitatively similar to that observed in analysis of human ISR performance (Henson et al., 1996).

Another classic characteristic of human ISR performance is the inverse-sigmoidal decrease in proportion of lists correctly recalled, as a function of list length. Figure 6.7 plots the proportion of lists correctly recalled as a function of list length, for both human performance (solid line) and simulated performance (dashed line). As can be seen, the list length effect observed in simulated ISR in the model is qualitatively similar to that observed empirically.

It is perhaps worth noting once again that the model's ability to perform both nonword repetition and immediate serial recall entails addressing the issue of serial ordering at multiple levels of representation. This difficult issue has not to our knowledge previously been addressed in an implemented computational model (but see Gupta, 1996). The two tasks are simulated, moreover, in a single model, with the identical architecture and long-term knowledge, and with identical parameter settings employed for both tasks. The ability to simulate both these tasks in a single model is therefore in itself of some interest as a computational demonstration. Going beyond this, however, the model achieves qualitatively humanlike performance in

simulated nonword repetition and immediate serial recall, by a number of measures, thus providing support for the general theoretical ideas incorporated in the model. This led to the next phase in the present work: employing the model, as described thus far, to investigate the empirically observed relationships between immediate serial recall and nonword repetition.

The correlation between nonword repetition and immediate serial recall

One of the earliest and most influential indications of a relationship between ISR and NWR ability came from the finding of correlations between performance in these tasks. These correlations have now been documented across a variety of ages and populations (e.g., Gathercole & Baddeley, 1989; Gathercole, Willis, Emslie, & Baddeley, 1992; Gathercole, Hitch, Service, & Martin, 1997; Gathercole et al., 1999; Gupta, 2003; Gupta, MacWhinney, Feldman, & Sacco, 2003). For instance, Gathercole et al. (1992) reported significant correlations between performance in NWR and ISR of .52, .67, and .45 respectively, for 4-, 5-, and 8-year-old children, Gathercole et al. (1999) reported a significant correlation of .32 in 13-year-old children, and in two experiments, Gupta (2003) reported significant correlations of .41 and .36 in adults.

Why do these correlations arise? Answers to this question have hitherto been offered only in rather general terms. The hypothesis incorporated in the present work is more specific: nonword repetition and immediate serial recall would be expected to be correlated, because the same serial ordering mechanism operates over both word form and syllable levels of representation. Given that the model incorporates this hypothesis, and yields credible simulations of both nonword repetition and immediate serial recall, we are in a position to test whether correlations between NWR and ISR do indeed fall out of a dependence on the same serial ordering mechanism.

Another question arises at this point, however: Why are the correlations in human performance not higher than have been reported? The correlations between NWR and ISR, as summarized above, are mostly in the range of .4 to .5 for older children and adults, indicating shared variance of .16 to .25. If the same mechanism does indeed provide for serial ordering in both NWR and ISR, why is the correlation between performance in these tasks not considerably higher? Perhaps the model can shed light on this question as well?

To examine whether NWR and ISR performance would, in fact, be correlated in the model, and to seek insight into the magnitudes of the correlations, several simulated "subjects" were created. The rationale here was to create a number of instantiations of the model incorporating the kind of variability that would certainly be present across human subjects – i.e., to simulate individual differences. Examining the correlation between

ISR and NWR performance across these instantiations of the model (the simulated subjects) would then be analogous to examining the correlation between ISR and NWR in the performance of a group of human subjects.

Each simulated subject was defined as a set of values of the parameters of the model. For all simulated subjects, this set of values was identical to that used in the simulations described above, except with regard to two parameters. One was a parameter governing the Hebbian learning rate in the connections from the serial ordering mechanism to the word form and syllable levels. This parameter can be thought of as a critical determinant of the efficacy of the serial ordering mechanism. The second parameter was one governing the learning rate between localist and distributed representations at the word form and syllable levels. This parameter can be thought of as affecting the efficiency of lexical and sublexical processing in the model. For each of these parameters, a range of parameter values was created, centered around the value that had enabled the model to yield the ISR and NWR performance described above. A value was picked randomly from each of these two parameter ranges, and this pair of parameter values was then taken to define one simulated subject. Seventy-five such pairs were chosen, thus creating 75 simulated subjects. Thus the simulated subjects incorporated individual differences in the efficiency of serial ordering and lexical/sublexical processing, but in all other respects were identical to each other and to the model that yielded the simulations described above. A simulation of ISR was then run using the parameter values of a particular "subject", and a simulation of NWR was also run using these same parameter values. This procedure was repeated for each of the 75 subjects. Thus each simulated subject was "tested" on both ISR and NWR.

For the purposes of assessing the correlation between ISR and NWR, each simulated subject's performance was gauged by a summary measure. Following the procedure adopted in empirical studies (e.g., Gupta, 2003; Gupta et al., 2003), ISR performance was measured in terms of list span, defined as the longest list length at which 50% or more of the presented lists were recalled correctly, and NWR performance in terms of the proportion of 2-, 4-, and 7-syllable nonwords correctly repeated. The correlation between these pairs of measures, derived from the 75 simulated subjects, was $r = .423$, $p < .01$.This is very close to the correlations observed in older children and adults.

The correlation arises in the model as a direct consequence of the engagement of the serial ordering mechanism at both the word form and syllable levels. To see why the correlation is not higher, it is useful to keep in mind what is involved in performance of ISR and NWR. The interaction of the long-term and short-term memory aspects of the system is particularly relevant here. As the model helps to make clear, although overall performance in ISR and NWR does depend on serial ordering mechanisms, it *also* depends on numerous other aspects of processing, including lexical and sublexical processing. This highlights the indubitable (but perhaps

sometimes overlooked) fact that measures of ISR and NWR performance in humans are obtained from participants who vary not just in their verbal STM mechanisms, but in multiple other performance characteristics that are separate from verbal STM. The correlation between performance in ISR and NWR in humans is thus computed across variance in these multiple different performance characteristics, and this is one reason why the correlation can be expected to be less than perfect. This is concretized in the model, in which the measures of ISR and NWR performance from the various simulated subjects are obtained across variation in the efficacy of serial ordering *and* of lexical/sublexical processing. As a result of the differing sources of variance, the correlation between NWR and ISR is not perfect, even though the model employs precisely the same serial ordering mechanism for both ISR and NWR. A cautionary note here is that there are other variables that in human performance would be expected to introduce uncorrelated variance in performance of ISR and NWR, and thus provide similar explanations for why the correlations are of only moderate magnitude. For instance, the serially ordered constituents of nonwords (i.e, syllables) carry coarticulatory information and a linguistic stress contour, whereas the serially ordered elements of lists (i.e., words) do not carry coarticulatory information, and typically have monotone stress. Moreover, lists contain pauses whereas nonwords do not. These diffferences are likely to affect performance in NWR and ISR in ways that reduce covariance in the tasks. In addition, human performance in NWR and ISR tasks is quite variable, with imperfect test–retest reliability, offering another possible reason for the observed magnitudes of correlations. For these reasons, the present simulations should be viewed more as a demonstration of the general point that variables that reduce covariance between NWR and ISR are certainly operative in performance of these tasks, and will lead to the correlation between these tasks being less than perfect even if their performance is dependent on the same serial ordering mechanism – rather than as a demonstration that lexical processing ability necessarily is the specific or sole variable that accounts for the modest magnitude of correlation.

While keeping these caveats in mind, it is nevertheless the case that the correlation between NWR and ISR performance in the model is qualitatively similar to that documented in human behavior. The emergence of this correlation in the model supports the present account of NWR and ISR performance, and in particular, the hypothesis that human performance in both tasks may depend on the same serial ordering mechanisms. The present account also offers an explanation of why these correlations are not higher. But are the correlations in the model really emergent, or has the model simply been "fitted" to yield them? In considering this question, it is worth emphasizing once again that the development and calibration of the model described in the preceding sections of this chapter attempted only to achieve NWR and ISR performance that was qualitatively humanlike. There was no attempt to tailor the model or adjust parameters to yield

correlation between performance in different tasks. In the present creation of simulated subjects, all except two parameter values were identical to those employed for calibrating the model in the first place. Moreover, the values of the remaining two parameters were highly constrained (being centered around their original values); and furthermore, the variation in these two parameter values across "subjects" was random. The procedure employed for *examining* correlations in the model was therefore not at all one of *adjusting* the model's parameters so as to *yield* correlations between NWR and ISR. The correlations observed between NWR and ISR performance in the model were thus not simply a consequence of parameter fitting. In a real sense, they fell out of the model, and thus offer some genuine insight into why these abilities may be related in humans.

The "pure STM" deficit and word and nonword processing

Another very influential indication of a relationship between ISR and the processing of nonwords has come from study of a neuropsychological syndrome termed the "pure STM" deficit. Patients with this deficit have dramatically impaired ISR span, but largely spared language comprehension and production, and other cognitive abilities (hence the name for the syndrome). As part of their generally spared language ability, such patients have little difficulty in repeating real words they know (Vallar & Baddeley, 1984). However, such patients present with great difficulty in repeating new words or nonwords (i.e., novel word forms), and also in learning them[4] (Baddeley, Papagno, & Vallar, 1988; Baddeley, 1993). This finding has been regarded as additional evidence for a relationship between ISR and NWR (e.g., Baddeley et al., 1988; Vallar & Baddeley, 1984).

This pattern (impaired ISR and NWR but preserved word repetition) has previously been addressed in computational work by Burgess and Hitch (1999), who described a simulation in which complete removal of a particular component of the model did lead to preserved word repetition coupled with a complete inability to perform ISR and NWR. As the authors noted, their model's *complete* inability to perform ISR and NWR was an exaggeration of the observed behavioral results. Although this was not explored further, it appears likely that partial (rather than complete) removal of the same component of the model would have achieved the desired pattern. Recall, however, that the Burgess and Hitch (1999) model (like all other computational models) did not address serial ordering at multiple levels of representation, and thus did not simulate serial ordering *within word forms*, which is a primary focus of the present work. The present simulations were thus aimed at investigating the pure STM pattern of impairment where serial ordering is computed both across word forms and within word forms.

To investigate the present model's ability to account for this pattern of results (viz., impaired ISR and NWR but preserved word repetition), its performance was examined under "impairment" of serial ordering. A

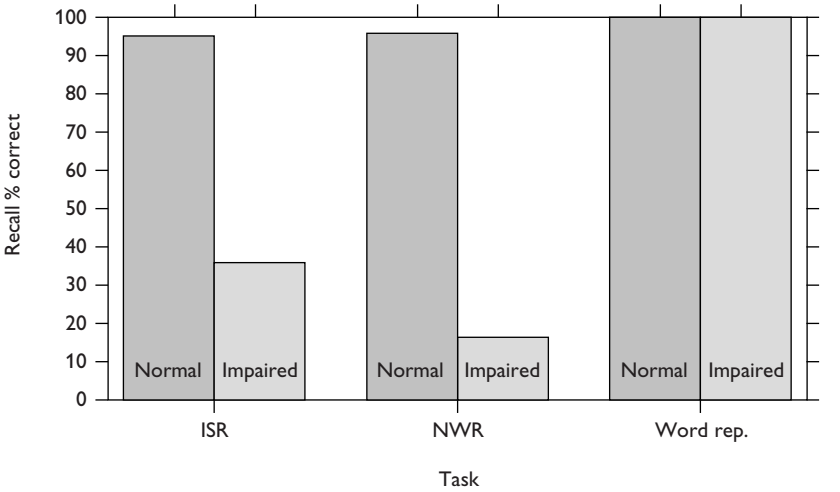

Figure 6.8 Simulation exhibiting a performance profile characteristic of a "pure STM" deficit.

"lesion" of the serial ordering mechanism was simulated using the parameter that governs the rate of Hebbian learning in connections from this mechanism to the word form and syllable levels. Recall that this parameter was also one of the two varied in the creation of the 75 simulated subjects described in the previous section. However, in the previous simulations, variation of this parameter was in a relatively narrow range around the mean value (i.e., within a "normal" range). In the present simulations, the parameter was set to a much lower value (i.e., to a "pathological" value, outside the "normal" range).

The model's performance was examined in three tasks. Each task was simulated both in an unimpaired state of the model, and under a lesion to the serial ordering mechanism (a "pure STM lesion"). The unimpaired state was identical to the set of parameter values used in establishing the basic simulations of humanlike NWR and ISR discussed earlier. The lesioned state differed from this only in the value of the one parameter discussed above. The three tasks examined were: (1) ISR of lists of 4 known monosyllabic words, (2) NWR of 4-syllable nonwords, and (3) Immediate repetition of known 4-syllable words. Figure 6.8 shows the model's accuracy in each of these tasks, both with and without the pure STM lesion. The y-axis shows the proportion of lists/nonwords/words correctly recalled. Even with the lesion, the model remained structurally able to perform tasks of ISR and NWR, but its performance on both was severely impaired by the lesion. Repetition of known words, however, was unaffected by the lesion.

To understand how these results arise, it may be helpful to recapitulate certain aspects of the functioning of the model. Recall that the model incorporates learning in both its long-term components and its short-term

components. Presentation of a polysyllabic novel word form (such as a new word or nonword – say, of three syllables) leads to short-term learning in the connections from the serial ordering mechanism to the word form and syllable levels, and thus to short-term encoding by the sequencing mechanism of the sequences of activations that occurred at the word form and syllable levels (see Figure 6.1). The sequence of activations that occurred at the word form level during presentation is simply a sequence of one activation – that of the newly created word form representation. The sequence of activations that occurred at the syllable level during presentation is the sequence of three syllables that comprise the word form. In addition to this short-term learning, exposure to the word form is also accompanied by long-term learning in the connections between and within the semantic, word form, syllable, and phoneme levels.

Immediate repetition of the just-presented novel word form requires recreation of the sequences of activation that occurred at the word form and syllable (and, ultimately, phoneme) levels. This is initiated by the serial ordering mechanism, which had encoded the sequence of one word form at the word form level, and the sequence of three syllables at the syllable level, and now reactivates these sequences. A crucial distinction between known words and nonwords arises here. For a known word, reactivation of the appropriate word form level representation is all that is needed to repeat the word, because there are strong long-term weights from the word form level representation to the appropriate sequence of syllables (these have come to be encoded, through many exposures, in the middle SRN – see Figure 6.2). Following a single exposure to a novel word form, however, the long-term weights from the new word form level representation to the appropriate syllable level sequence are not strong enough for activation of the word form representation to produce the correct syllable sequence. Successful repetition of a novel word form (but not a known word form) is thus crucially dependent on the short-term encoding of the sequence of syllables. This is why NWR and ISR were both impaired by the lesion of the serial ordering mechanism in the previous simulation: both are crucially dependent on the maintenance of a sequence of activations (at the syllable level for NWR and at the word form level for ISR) by the serial ordering mechanism. For repetition of a known word, by contrast, the serial ordering mechanism need only recreate a single activation at the word form level. Once this word form level representation is activated, the long-term weights can do the rest; the short-term serial ordering mechanism's encoding of the syllable level sequence is not needed. Repetition of a known word requires the serial ordering mechanism only to recreate a single word form level representation, and this is feasible even under considerable impairment.

In summary, the model exhibits a pattern of impairment qualitatively similar to that observed in pure STM patients. This result provides additional support for the present characterization of NWR and ISR and their relationship, and of how the pure STM deficit may arise in human patients.

Verbal short-term memory and novel word learning

The discussion thus far has described a variety of ways in which the present model offers an answer to the question of why performance in nonword repetition and immediate serial recall are related, and of how it accounts for a variety of empirical results that bear on this relationship. According to the theoretical view incorporated in this model, NWR and ISR draw on the same serial ordering mechanism, and the model provides a concrete demonstration of how this might work. But what about the relationship between nonword repetition and word learning? At the beginning of this chapter, it was suggested that the existence of a relationship between nonword repetition and word learning appears fairly intuitive, because greater facility in processing nonwords would be expected to lead to greater facility in eventually learning them. How can this intuition be concretized in terms of the present model?

As reiterated in the preceding section, the model incorporates learning both in its long-term components and in its short-term components. Exposure to a novel word form leads to both short-term and long-term learning, with the long-term learning being in the connections between and within the semantic, word form, syllable, and phoneme levels. Although, as discussed above, the long-term learning that occurs during a single exposure to a novel word form does not suffice to "learn" it, multiple exposures to nonwords will lead to multiple instances of long-term learning in these connections, and over multiple exposures, a nonword will turn into a known word. Indeed, this is precisely what happens during the "pre-training" of the model. Initially, repetition of the form is dependent on the short-term weights from the serial ordering mechanism to the word form level and (crucially) the syllable level. Eventually, however, once the long-term connection weights have reached sufficient strength, short-term maintenance of the syllable sequence becomes redundant (although it still occurs, if the serial ordering mechanism is intact). The learning of a new word form thus represents a transition from dependence on short-term learning to long-term knowledge.

However, nothing in this description actually provides an account of why the accuracy of nonword repetition might be related to the efficacy of word learning. This is because learning in the model (both short-term and long-term) occurs during *exposure* to word forms. The short-term learning-dependent accuracy of word form repetition at *recall* does not impact long-term learning in the model as it has thus far been developed. How then might repetition accuracy for a novel word affect its long-term learning?

The current model does, in fact, provide a simple way of thinking about this question. In particular, let us suppose that long-term learning occurred not only at *exposure* to a word form, but also during *repetition* of the word form. At repetition, what was learned would be *the repetition itself*, errors and all. The efficacy of long-term learning would then be dependent on the

accuracy of nonword (or novel word) repetition in a very direct manner. If repetition accuracy were high, the learning that occurred during repetition would supplement the learning that occurs during exposure. If repetition accuracy were low, however, the learning during repetition/recall would interfere with the learning at exposure. Although such learning-at-recall is not currently implemented in the model, the conceptualization offered by the model in this regard is fairly clear. The model thus helps concretize intuitions about *how* novel word learning may be related to nonword repetition (and thus to the serial ordering mechanism, or phonological short-term memory). It is also interesting that, in the domain of list recall, there is evidence for learning-at-recall: when specific to-be-recalled lists recur multiple times during a list recall session, facilitated performance on the repeating lists (the *Hebb effect*; e.g., Cumming et al., 2003; Hebb, 1961; Melton, 1963; Sechler & Watkins, 1991; Ward, 1937) is stronger if recall of the lists is overtly attempted than if there is no overt recall (e.g., Melton, 1963).

The question arises, of course, of whether the present hypothesis suggested by the model (i.e., that accuracy in learning-at-recall during novel word repetition will affect novel word learning) is a reasonable one for human performance. In fact, there is evidence to suggest that, at least in individuals with memory impairments, learning success may depend on the avoidance of performance errors – that is, on setting up an "errorless" training situation (e.g., Evans, Wilson, Schuri, Andrade, Baddeley, & Bruno, 2000; Glisky, 1995). The present hypothesis amounts to extension of these findings to normal word learning. The hypothesis also leads to testable predictions. According to the hypothesis, the greater the number of overt repetition errors made during the word learning process, the poorer the learning will be. What would affect the number of overt repetition errors made during word learning? Obviously, if overt repetition responses were not made at all during exposure events, this would limit the number of overt repetition errors made during learning. Therefore, according to the hypothesis under consideration, manipulating the presence or absence of overt repetition during word learning should affect the level of word learning obtained. However, such an effect will depend on the inherent level of accuracy of overt repetition. If this accuracy is high, then, even if overt repetitions are being made during exposure, there will be few inaccurate overt repetitions. Therefore, with high accuracy levels, the presence/absence of overt repetition during learning should make little difference to the ultimate word learning outcome. However, if the accuracy of overt repetition is low, then the presence of overt repetition during exposure will lead to many overt repetition errors, and the ultimate word learning outcomes should be lower than if there had been no overt repetition. Thus an interaction is predicted, such that the word learning outcome should be lower when overt repetitions are made during exposure than when they are not made – but only if the overt repetitions are inaccurate. What would affect the accuracy of overt repetitions? An obvious variable here is word

length: overt repetition errors would be greater, for instance, for five-syllable novel names than for one-syllable novel names. It can therefore be predicted that word learning should be lower with than without overt repetition during learning, but only at higher word lengths.

This prediction receives some preliminary empirical support. In two experiments (Abbs, Gupta, & Khetarpal, 2007), participants engaged in a word learning task in which they were visually presented with a drawing of a novel object, together with an auditorily presented novel "name" for the object. These presentations were exposure trials. The participant's task was to learn the pairing so as to be able to produce the name when cued with the visual image (drawing) of the object, in test trials. Following the paradigm of Gupta (2003), the drawings depicted "space aliens", and the names were possible nonwords of English. The learning session contained several blocks of exposure and testing, for each of four name–picture pairs.

In each of the two experiments, half the participants were instructed to overtly repeat the name on each exposure trial, and half were instructed simply to listen to the name at each exposure trial, and not to repeat it overtly. Thus each experiment incorporated a 2 (between: overt repetition during exposure trials/no overt repetition) × 6 (within: exposure block 1/2/3/4/5/6) mixed factorial design, with 36 participants in each experiment (18 in each between condition). However, in Experiment 1, the names were 3-syllable nonwords while in Experiment 2 the names were 4-syllable nonwords. What varied between the two experiments was therefore word length, which was expected to affect the number of overt repetition errors during exposure trials. Such an effect was in fact obtained, with repetition accuracy in the overt repetition condition during exposure trials being 92.2% in E1 (3-syllable names) and 82.3% in E2 (4-syllable names). The difference was significant: F (1, 34) = 5.56, MSE = 9.93, $p < .05$.

Figures 6.9(a) and (b) show the proportion of names correctly recalled on the six test trials ("recaps") for E1 and E2 respectively. In E1 (3-syllable names), there was no significant difference in recall for the "repeat" versus "silent" conditions ($p > .7$). However, in E2, recall was worse in the repeat than in the silent condition, and the difference was very nearly significant ($p = .054$). It should be noted that these results are preliminary, requiring further investigation of both robustness and alternative explanations. Nevertheless, they are consistent with the somewhat counter-intuitive predictions derived from the current hypothesis, and thus provide some initial support for the model. Perhaps more importantly, they demonstrate the testability of the model.

Conclusions

The present model is to our knowledge the only current computational model to tackle the issue of serial ordering both across word forms (as in list recall) and within word forms (as in nonword repetition), and hence the

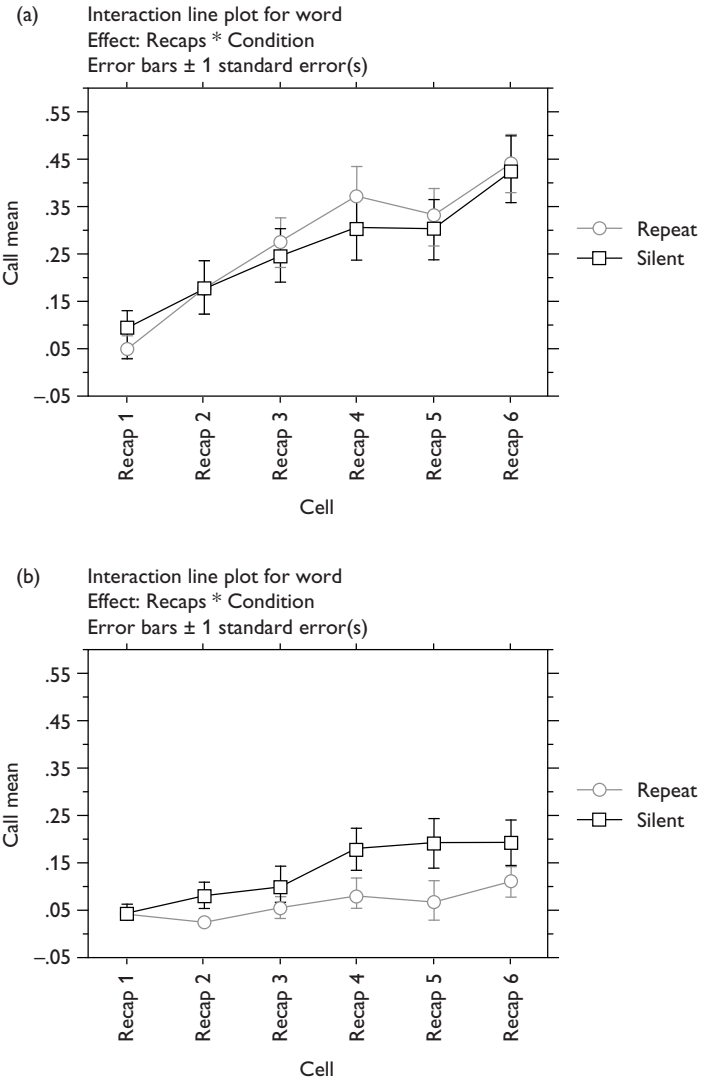

Figure 6.9 Effect of overt repetition of novel words during learning: (a) 3-syllable novel words; (b) 4-syllable novel words.

only current model that can offer an account of observed relationships between performance in these two tasks. Indeed, once the model had been set up to simulate performance in the two tasks, a number of patterns of relationship between performance in the tasks fell out of the model, including the well-documented correlation between performance in the tasks, and the profile of impairment in pure STM deficits. Additionally, the

model generates a testable hypothesis regarding the relationship between nonword repetition and the long-term learning of novel words.

Of course, the model has many limitations. It currently addresses only a very few of the many and varied phenomena of immediate serial list recall that have been accounted for by a number of other computational models (e.g., Brown et al., 2000; Burgess & Hitch, 1992, 1999; Page & Norris, 1998). As noted in the introduction, such coverage is not the goal of the present work, but nevertheless it constitutes a limitation if the model is viewed as a model of list recall. As an account of linguistic processing also, the present model has many limitations. It does not, for instance, incorporate information about variables such as linguistic stress or coarticulation, to name just two, which are likely to impact processing. Despite these limitations, the model does provide a concrete, computationally rigorous, and empirically testable means of thinking about *relationships* between the domains of serially ordered memory and language processing, and in particular, between list recall and nonword repetition; it thus has the potential to play a useful role in investigation of these relationships.

Acknowledgments

The author thanks Jamie Tisdale and Tobe Hagge, who conducted the bulk of the computer programming underlying this work, and Gary Dell, Nadine Martin, and Myrna Schwartz for helpful discussion during its development. This research was supported in part by grants NIH NIDCD R01 DC006499 to Prahlad Gupta, and NIH NIDCD RO1 DC01924 to Nadine Martin, Prahlad Gupta, and others.

Notes

1 These hidden layer units have a one-to-one correspondence with the distributed semantic and distributed word form representations. As the localist word form units also have a one-to-one correspondence with the word form level distributed representations, the question arises of whether the model could be simplified by merging these two sets of localist units – that is, by making the localist word form units also serve as the hidden units between the semantics and word-form levels. The reason for avoiding this merger was that sequencing at the word form level (via the avalanche) would then be equally influenced by semantic similarity and phonological similarity, which would be contrary to evidence suggesting that the effects of phonological similarity are much stronger than those of semantic similarity (e.g., Baddeley, 1968).

2 The patterns of activation at the syllable level in turn led to output of the appropriate phoneme sequences corresponding to the syllables.

3 Learning also occurred in all the long-term components of the model, but did not influence list recall, because the connection weights for the words had already been established during vocabulary pretraining.

4 It should be noted that they also do exhibit difficulty in certain other aspects of language processing such as repeating sentences longer than 8 syllables, and numeric sequences.

References

Abbs, B., Gupta, P., & Khetarpal, N. (2007). *Does overt repetition facilitate word learning?* Manuscript.

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology*, *20*, 249–264.

Baddeley, A. D. (1993). Short-term phonological memory and long-term learning: A single case study. *European Journal of Cognitive Psychology*, *5*, 129–148.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.

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.

Botvinick, M., & Plaut, D. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, *113*, 201–233.

Brown, G. D. A., & Hulme, C. (1996). Nonword repetition, STM, and word age-of-acquisition. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 129–148). Hove, UK: Psychology Press.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*, 127–181.

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

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551–581.

Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory*, *11*, 43–63.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

Elman, J. L. (1990a). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Elman, J. L. (1990b). Representation and structure in connectionist models. In G. P. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press.

Evans, J., Wilson, B., Schuri, U., Andrade, J., Baddeley, A., & Bruno, O. (2000). A comparison of errorless and trial-and-error learning methods for teaching individuals with acquired memory deficits. *Neuropsychological Rehabilitation*, *10*, 67–101.

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., Hitch, G. J., Service, E., & Martin, A. J. (1997). Short-term memory and new word learning in children. *Developmental Psychology*, *33*, 966–979.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, *13*, 65–77.

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.

Glisky, E. (1995). Acquisition and transfer of word processing skill by an amnesic patient. *Neuropsychological Rehabilitation*, *5*, 299–318.

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* (Vol. 5). New York: Academic Press.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, *11*, 23–63.

Gupta, P. (1995). *Word learning and immediate serial recall: Toward an integrated account*. PhD thesis, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA.

Gupta, P. (1996). Word learning and verbal short-term memory: A computational account. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Meeting of the Cognitive Science Society* (pp. 189–194). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gupta, P. (2002). Are nonwords lists? The effects of syllable serial position on repetition of individual nonwords. In *Abstracts of the Quebec'02 conference on short-term/working memory*. Quebec City, Quebec. Abstract.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology (A)*, *56*, 1213–1236.

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory*, *13*, 318–324.

Gupta, P., & Cohen, N. J. (2002). Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological Review*, *109*, 401–448.

Gupta, P., Lipinski, J., Abbs, B., & Lin, P.-H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language*, *53*, 141–162.

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

Gupta, P., MacWhinney, B., Feldman, H., & Sacco, K. (2003). Phonological memory and vocabulary learning in children with focal lesions. *Brain and Language*, *87*, 241–252.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, *35*, 1–31.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Oxford: Blackwell.

Henson, R., Norris, D., Page, M., & Baddeley, A. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology*, *49*, 80–115.

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.

Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach* (Report 8604). La Jolla, California: Institute for Cognitive Science, University of California, San Diego.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

McLeod, P., Plunkett, K., & Rolls, E. (1998). *Introduction to connectionist modeling of cognitive processes*. New York: Oxford University Press.

Melton, A. W. (1963). Implications of short-term mempry for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *2*, 1–21.

Page, M. (2000). Connectionist modeling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, *24*, 443–512.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. Rumelhart, & J. McClelland (Eds.), *Parallel distributed processing, Vol. 1: Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Sechler, E. S., & Watkins, M. J. (1991). Learning to reproduce a list and memory for the learning. *American Journal of Psychology*, *104*, 367–394.

Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, *23*, 151–161.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, *41*, 101–175.

Ward, L. B. (1937). Reminiscence and rote learning. *Psychological Monographs*, *49*, 1–64.