AISB Journal **1**(6), © SSAISB, 2005
http://www.aisb.org.uk

Modelling Vocabulary Acquisition: An Explanation of the Link between the Phonological Loop and Long-Term Memory

Gary Jones*, Fernand Gobet† and Julian M. Pine‡

* Centre for Psychological Research in Human Behaviour, University of Derby, Western Road, Mickleover, Derby DE3 9GX, *g.jones@derby.ac.uk*
† School of Social Sciences and Law, Brunel University, Uxbridge, Middlesex, UB8 3PH, *fernand.gobet@brunel.ac.uk*
‡ School of Psychology, University of Liverpool, Liverpool, L69 7ZA, *julian.pine@liverpool.ac.uk*

Abstract

The acquisition of vocabulary represents a key phenomenon in language acquisition, but is still poorly understood. Recently, the working memory model (Baddeley & Hitch, 1974) has been adapted to account for vocabulary acquisition (e.g., Gathercole & Baddeley, 1989). It is claimed that the phonological store, one of the components of working memory, offers a critical mechanism for learning new words. However, one of the theoretical weaknesses of this approach is that no account is given for the mechanisms and representations used in long-term memory learning. This paper presents a computer model combining the EPAM/chunking approach (Feigenbaum & Simon, 1984) with the working memory approach. Phonemic learning is simulated as the elaboration of a discrimination net. Naturalistic input, consisting of utterances from nine mothers interacting with their child, is used during the learning phase. Simulations show that the model can account reasonably well for the nonword repetition task described by Gathercole and Baddeley (1989), a task often presented as a powerful diagnostic of vocabulary learning.

1 Introduction

Children are remarkably adept at learning new verbal information. After an initial slow period from about 12 to 16 months when most children learn around 40 words, the learning rate increases such that in the next four months children will have learnt 130 more new words (Bates et al., 1994). By the beginning of their school years children learn up to 3,000 words per year (Nagy & Herman, 1987).

A major part of learning new words is learning the novel sequences of sounds that represent the word. However, it is difficult to directly examine the processes involved in learning the sound patterns of new words because it is impossible to be certain that the new sound pattern has never been encountered before. The use of nonwords which conform to the phonotactic rules of English provides a good test of vocabulary learning because it ensures that the (non)word to be learned is novel.

1.1 The Nonword Repetition Test

The nonword repetition (NWR) test (Gathercole, Willis, Baddeley & Emslie, 1994) was designed to investigate the role of phonological memory in word learning. The test involves the experimenter speaking a nonword and the child attempting to repeat it. Gathercole and Baddeley (1989) found that, compared to the other measures they used, the NWR test was the best predictor of vocabulary size, even after vocabulary scores (as calculated by the British Picture Vocabulary Scale, Dunn & Dunn, 1982) were partialled out of correlations.

Gathercole and Baddeley (1990) used the NWR test to categorise children into two groups, those with low NWR scores, and those with high NWR scores. Children in the high NWR group were better at learning nonword labels than children in the low NWR group. Gathercole, Willis, Emslie, and Baddeley (1991) found better NWR performance on nonwords that were rated high in wordlikeness than nonwords rated low in wordlikeness. These NWR studies show the influence that vocabulary knowledge has upon the learning of new words.

The nonword repetition test involves two sets of nonwords, one set having single consonants (e.g. **rubid**) and one set having clustered consonants (e.g. **glist**ow). There are twenty nonwords in each set, divided into four groups of five based on the number of syllables in the nonword (one to four). Several studies using these types of nonwords have consistently found that repetition accuracy decreases as the number of syllables in the nonword increases, excepting one-syllable nonwords (e.g., Gathercole & Adams, 1993; Gathercole, Willis, Emslie & Baddeley, 1991), and that accuracy is worse for clustered consonant nonwords.

Based on these findings, NWR ability would seem to provide a good test of phonological memory and is a good indicator of vocabulary size. The NWR findings can be explained within the theoretical framework of the working memory model.

1.2 The Phonological Loop Explanation of NWR Findings

The working memory model (Baddeley & Hitch, 1974) has recently been adapted to account for vocabulary acquisition (e.g., Gathercole & Baddeley, 1993). It is claimed that the phonological loop part of the model is a critical mechanism for learning new words. The phonological loop has two linked components: the phonological short-term store, and the sub-vocal rehearsal mechanism. Items in the store decay over time (around 2,000 ms, Baddeley, Thomson & Buchanan, 1975). The sub-vocal rehearsal mechanism (involving sub-vocal articulation in real-time) can refresh items in the store in a serial, time-based manner (Gathercole & Martin, 1996). The store is linked to the central executive part of the model, which provides a link to long-term memory (LTM) (Gathercole & Baddeley, 1993). The influence of LTM is acknowledged (e.g., Gathercole, Willis, Baddeley & Emslie, 1994), but the nature and definition of the link is yet to be defined.

The phonological loop hypothesis is able to explain the basic NWR findings involving nonword length because of the decay that takes place in the phonological store (items remain in the store for 2,000 ms unless refreshed). Longer nonwords take longer to rehearse and so their representations in the phonological store are not refreshed as often as shorter nonwords. Repetition ability will therefore decrease for longer nonwords. The poorer repetition performance for clustered consonant nonwords can be explained in a similar way (because although the clustered consonant nonwords contain the same number of syllables as the single consonant nonwords, they contain more phonemes).

Differences in NWR ability between children of the same age were originally attributed to differences in rates of subvocal rehearsal (Gathercole & Baddeley, 1993). However, recent findings show that children do not use subvocal rehearsal until around seven years of age (see Cowan & Kail, 1996, for a review). This has led to the phonological store being assumed to be the primary language learning device, with differences in language learning across children of the same age being explained by the quality of the phonological representation of just-spoken items (Baddeley, Gathercole & Papagno, 1998). The lack of a rehearsal process for children below seven years of age does not appear to hinder the model. Brown and Hulme (1996) show that NWR phenomena can be explained solely by a decay based model (which the phonological store now becomes for children under seven years of age). This model will be discussed later.

The phonological loop is able to explain a lot of the vocabulary acquisition phenomena using a very simple mechanism. However, it fails in two critical areas: there is no explanation of how words are learned, and there is no explanation of how the loop interacts with LTM. Speculative explanations have been given as to how the loop may interact with LTM (e.g., Baddeley, Gathercole & Papagno, 1998; Gathercole & Baddeley, 1989). Gupta and MacWhinney (1997) have proposed a formal specification but their model has not yet been implemented computationally.

The phonological loop hypothesis also lacks precision because it is part of a verbal theory. For example, there is no definition of how rehearsal rate changes based on the length of the sound pattern being rehearsed (except to say that long strings are rehearsed slower than short strings). Implementing the loop within a computational architecture forces precision because the theory is implemented as a running computer program. Several computational implementations of the phonological loop exist.

1.3 Existing Computational Implementations of the Phonological Loop

Burgess and Hitch (1992) detail a connectionist network which was primarily intended to model serial order effects (the recall of a set of items in the correct presentation order). The decay in the phonological store is represented by decay on the weights between layer nodes (decay is proportional to the number of phonemes to be output). Rehearsal is synonymous with articulation: the most active word in the network is selected for output once the phonemic input has been processed. The model can explain word length and articulatory suppression effects, but does not explain any of the NWR findings. In addition, no phoneme or word learning takes place; the model provides no theory as to how phonemes and words are created in LTM.

Brown and Hulme (1996) give an account of a trace-based decay model which bears resemblance to the phonological store (the model intentionally has no rehearsal process). Time is represented in 0.1-sec slices; input items are split into segments such that each segment corresponds to a time slice. Longer words therefore take up more segments and so occupy more time slices. As each segment of an input item enters the store, it is given a fixed initial strength, which decays over time. This means that the early *segments* of an input item suffer more decay than the later segments, as well as earlier input *items* suffering more decay than later input items.

The probability of recalling an item is the product of the current strength of each of the segments of the item. The probability is increased to reflect the influence of LTM during recall; wordlike nonwords would therefore have their probability increased more than non-wordlike nonwords. Using these mechanisms, the model can account for recall effects for different lengths of both words and nonwords (Brown & Hulme, 1996). However, the model does not account for any learning processes.

There are problems with each of the existing phonological loop models. For the phonological loop to successfully provide an account of vocabulary learning, a precise specification of its interaction with LTM is required. EPAM is a computational modelling architecture which is able to provide such a specification.

2 Implementing the Phonological Loop within the EPAM Architecture

EPAM is a computational modelling approach whereby a discrimination network is built based on the input that the model receives. In terms of Artificial Intelligence (AI), the approach can be seen as being very similar to tries (Fredkin, 1960), and particularly the ”suffix” trie whereby input is presented to the trie as a whole and then as every part of the suffix (e.g., W H A T, then H A T, then A T, then T). In much the same way as tries, a hierarchy of the input is built, as will be shown later. Tries have commonly been used in AI to represent dictionaries (e.g., Arslan & Egecioglu, 2004), but have also been used in other domains such as matching for similarity in video databases (e.g., Park & Hyun, 2004). Discrimination networks have also been used in AI (e.g., in expert systems, Gerevini et al., 1992), although the main area for the EPAM form of discrimination networks has been in simulating various areas of human cognition, such as learning, memory, and perception in chess (De Groot & Gobet, 1996; Gobet, 1993; Gobet & Simon, 2000; Simon & Gilmartin, 1973), verbal learning behaviour (Feigenbaum & Simon, 1984), the digit-span task (Richman, Staszewski & Simon, 1995), the context effect in letter perception (Richman & Simon, 1989), and the acquisition of syntactic categories (Freudenthal, Pine & Gobet, 2005; Gobet, Freudenthal & Pine, 2004; Jones, Gobet & Pine, 2000) (see Gobet et al., 2001, for an overview). EPAM provides a modelling environment which is well suited for describing how sound patterns can be learnt. When a sentence is heard, it is heard as a sound pattern in the form of a sequence of phonemes. This sequence of phonemes needs to be processed and stored in a hierarchical fashion (to illustrate the order of the sound patterns).

EPAM provides a simple mechanism by which this goal can be accomplished. Furthermore, there would seem to be an easy method by which the sound patterns in LTM (i.e., the resulting discrimination network) can be linked to the phonological store - when sound patterns come in, they can be matched to those that exist in LTM and thus any sequence of sound patterns that match do not have to be stored individually in the phonological store - a link can be placed there to the relevant item in the discrimination network.

Precisely how EPAM will be linked to the phonological store will be explained later.

2.1 The EPAM Architecture

EPAM learns by building a discrimination network. The discrimination network is a hierarchical representation of the input and consists of nodes connected to one another by links. Nodes contain information and links between nodes contain tests which must be fulfilled before they can be traversed. For the purposes of modelling the learning of sound patterns, EPAM has been simplified and is henceforth referred to as EPAM-VOC.

When an input (e.g., a sequence of phonemes) is given to the network, EPAM-VOC traverses down the hierarchy as far as possible. This is done by starting at the top node (the root node) and selecting the first link whose test is fulfilled by the first part of the input. The node at the end of the link now becomes the top node and the rest of the input is applied to all the links below this node. When a node is reached where no further traversing can be done (e.g., the input fulfils none of the tests of the nodes links, or the node is a leaf node), EPAM-VOC compares the information at the node with the input information. Learning now occurs in two ways.

1. *Discrimination*. When the input information mismatches the information given at the node, a new link (i.e., test) and node are added to the tree below the node that has just been reached. The new test will relate to the mismatched part of the input.

2. *Familiarisation*. When the input information is under-represented by the information at the node (e.g., features from the input are not present in the information at the node), new features (from the input) are added to the information in the node. In EPAM-VOC, the *image* of a node will always consist of all the information in the links that lead to the node.

Discrimination therefore creates nodes and links, and familiarisation creates or modifies the information contained in nodes. Examples of the discrimination and familiarisation learning mechanisms will be given later.

2.2 Learning Sound Patterns in EPAM-VOC

EPAM-VOC provides an efficient method for representing items in LTM. The basic idea is to give as input to the model the utterances from mothers speech so that EPAM-VOC can learn phonemes and combinations of phonemes. Mothers' utterances will be converted into a sequence of phonemes before being used as input. This will be done using the CMU Lexicon database (available at http://www.speech.cs.cmu.edu/cgi-bin/cmudict) which cross-references words with their phonemic representations. The use of phonemic input assumes that some form of phonemic feature primitives already exist to distinguish one phoneme from another.

EPAM-VOC will begin with a null root node. When it receives an input (a sequence of phonemes), new nodes and links will be created. At first, most of the new nodes and links will just be for single phonemes, as EPAM-VOC learns to master individual phonemes. As learning progresses, the information at nodes will become sequences of phonemes and therefore segments of speech (e.g., specific words) rather than just individual sounds (i.e., phonemes). To accomplish this, the EPAM learning mechanism is altered in two ways. First, before a sequence of phonemes can be learnt, the individual phonemes in the sequence must have been learnt. Second, when individual phonemes are learnt, they are linked to the root node (in this way all sequences of phonemes are below the node which represents the initial phoneme in the sequence).

Let us consider an example of the network learning the utterance "What?". Using the CMU Lexicon database, this utterance is converted to the phonemic representation "W AH1 T" (all of the phonemes used in the database map onto the standard phoneme set for American English). Note that the phonemic input to the model does not specify gaps between words, but does specify the stress of particular phonemes (0=unstressed; 1=primary stress; 2=secondary stress).

The first part of the input ("W") is applied to all of the root nodes' links in the network. If the network is empty, there will be no links. At this point EPAM-VOC must discriminate because the information "W" mismatches the information at the root node (the root node information is null). The discrimination process creates a new node, and a link from the root node to the new node with the test "W". EPAM-VOC must then familiarise itself with the input, in order to create the "W" information in the image of the node. EPAM-VOC then moves on to the remainder of the input (i.e., "AH1 T") much like a suffix trie. In a similar way as for "W", the phoneme "AH1" will be learnt. EPAM-VOC then moves on to the remainder of the input (i.e., "T"), and in a similar fashion, learns the phoneme "T". Thus when the input is received the first time, the individual phonemes "W", "AH1" and "T" are learnt.

When encountering the input for the second time, the link "W" can be taken, and the input can move to the next phoneme, "AH1". As node "W" does not have any links, discrimination occurs below the "W" node, creating a new node below the "W" with a link of "AH1". Familiarisation then fills this node with the contents "W AH1". The remainder of the input (i.e., "T") is then examined, but as this has already been learnt, the processing of the input terminates.

The third time the input is received, the "W" link can be taken, with the input moving on to "AH1". As there is an "AH1" link below the "W" node, this link can be taken, and the input can move on to the "T". As there is no "T" link below the "W AH1" node, discrimination occurs. A new node is created below the "W AH1" node with the link "T". Familiarisation will fill in the contents of the new node with "W AH1 T". Thus after three presentations of the input, the network is as shown in Figure 1. The simple example serves to illustrate how EPAM-VOC works; in the actual learning phase each utterance line is only used once, encouraging a diverse network of nodes to be built. Note that EPAM-VOC needs to know individual phonemes before they can be learnt as part of a sequence of phonemes. For example, should the network in Figure 1 see the utterance "Which?" ("W IH1 CH"), it will traverse down the "W" link, and move on to the next part of the input (i.e., "IH1 CH"). However, the network does not know the phoneme "IH1", and so it needs to discriminate at the root node, learning the individual phoneme "IH1" before moving on to the remainder of the input "CH" (and learning this as an individual phoneme also).

2.3 Implementing the Phonological Loop and Linking it to Long-Term Memory

The model now requires a specification of the phonological loop and a method by which the loop interacts with LTM. The findings relating to the NWR test (the standard test of the phonological loop) were all carried out on children below the age of six. As children below the age of seven are believed not to rehearse, the rehearsal part of the loop should not be used to simulate the NWR findings reported above.

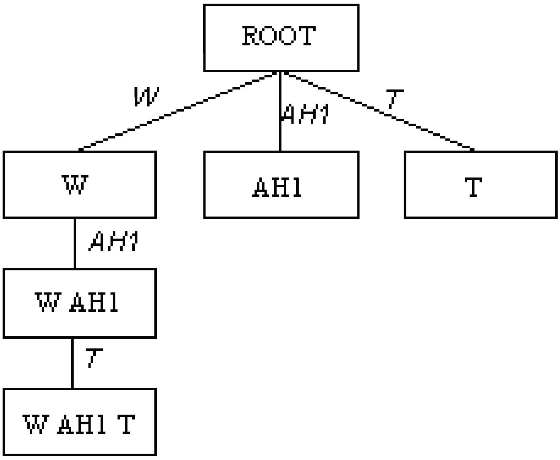

Figure 1: Structure of an EPAM-VOC net after receiving the input "W AH1 T" three times.

The storage part of the phonological loop is a decay based store which allows items to remain in the store for 2,000 ms. The model will have a time-limited store which will allow 2,000 ms of input (i.e., consistent with the phonological loop estimates). The input will be cut-off as soon as the time limit is reached, because there is no rehearsal to refresh the input representations.

The cumulative time required by the input provides a theory of how the amount of information in the phonological store is mediated by LTM. When an input is heard, LTM (the EPAM-VOC network) is accessed and the input is recoded using the minimum number of nodes possible. Rather than the actual input being placed in the phonological store, the nodes which capture the input are used. The length of time taken to represent the input is therefore calculated on the number of nodes that are required to represent the input. The time allocations are based on values from Zhang and Simon (1985), who estimate 400 ms to match each node, and 84 ms to match each syllable in a node except the first (which takes 0 ms). As the input will be in terms of phonemes, with approximately 2.8 phonemes per syllable (based on estimates from the nonwords in the NWR test), the time to match each phoneme in a node is 30 ms.

Using the example input "What about that?" ("W AH1 T AH0 B AW1 T DH AE1 T") and the network as given in Figure 1, the actual input to the model will be "W AH1 T AH0 B AW1" because this is all that can be represented in the phonological store within the 2,000 ms timescale. The "W AH1 T" part of the input is represented by a single node, and is allocated a time of 460 ms. Most of the other phonemes are not known to the model and are therefore assumed to take the same time as a full node (400 ms) (the time allocated to each phoneme is assumed to be constant). This means that only three more phonemes can be represented within the phonological store (the actual input to the model having a time allocation of 1,660 ms). When the EPAM-VOC network is small, only a small amount of the input information can be represented in the store, and so new nodes will not contain much information. When the EPAM-VOC network is large, a lot of the input information can be represented in the store and so the model can create new nodes which contain large amounts of information.

3 Simulating the Nonword Repetition Results

There are two main sets of results for the NWR test. One set was tested on children of four and five years of age (Gathercole & Baddeley, 1989; these results were reported in the introduction). The problem with this study is that the children are of an age where they already have a reasonably large vocabulary size. A second set of NWR results is reported by Gathercole and Adams (1993), who used a simpler version of the test on children of two and three years of age. They found the modified test still allowed phonological memory skills to be reliably tested.

The EPAM-VOC model is able to learn sequences of sounds from inputs that are strings of phonemes (converted from speech utterances). It should therefore be capable of simulating both sets of NWR data (2-3 year olds and 4-5 year olds) by modifying the input that is given to the model to reflect the type of input that will be received by these age groups.

In the simulations that will be presented, the model normally learns something for every input it receives. The EPAM parameter of 8 s to learn a single node was dispensed with because of the long time scale involved in the simulations. Given the same input to the model, there should be no significant difference to the results whatever the time taken to learn a node.

The NWR test for the model will be performed by presenting each nonword to the model (as a string of phonemes), and seeing if the components of the nonword can be accessed within the same time limitations that were used for the input (see above). By definition, the information at one node will not be able to represent all of a nonword (because the nonword will never have been received as input, and the presentation time is assumed to be too short to build a new LTM chunk). The information from several different nodes will be required to represent the nonword. If the number of nodes and the phonemes in each node can fit into the time limit, the nonword is repeated accurately, otherwise the nonword is repeated incorrectly. The models' NWR test does not involve articulation because the current EPAM-VOC model does not include a theory of articulation.

3.1 Simulation of Two to Three Year Old Children

For the simulation of the NWR test for children of 2-3 years of age, naturalistic input was used for the model. The input consisted of the mother utterances from nine mothers interacting with their 2-3 year old children, taken from Theakston, Lieven, Pine and Rowland (2000). The average number of utterances for each mother was 25,711 (range 17,474-33,452). The duration of the phonological store was changed from 2,000 ms to 1,750 ms, because there is a high probability that the phonological store of very young children has less duration than adults (existing timing estimates for the phonological store have been based on studies involving adults).

The model was run once for each of the mother's input, resulting in nine different simulations. The NWR test for the model consisted of presenting each nonword as input to the model and seeing if it could represent the nonword within the 1,750 ms time capacity. Note that Gathercole and Adams used a simplified version of the NWR test (using 1-3 syllable nonwords and not distinguishing between single and clustered consonants). They also performed a word repetition test. The results of the children and the model (after seeing 30 percent of the mother's utterances as input) for both the nonword and word repetition tests are shown in Figures 2 and 3.

The model performs at ceiling for the one and two-syllable words and nonwords. A minimum of four phonemes can fit into the 1,750 ms time limit in the phonological store (one phoneme uses up 400 ms at most). In the words and nonwords used, the average number of phonemes for one-syllable items is 3.2 and for two-syllable items is 5.0. The model has therefore chunked at least one pair of phonemes contained in each of the words and nonwords used in the tests. The children do not perform at ceiling for any of the conditions. Nevertheless, the model still provides a significant correlation with the child data ($r(4)=0.859$, $p<.05$).

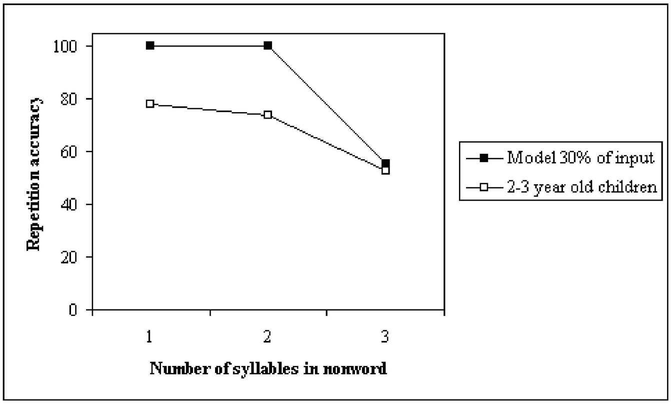

Figure 2: NWR accuracy for 2-3 year old children and the model.

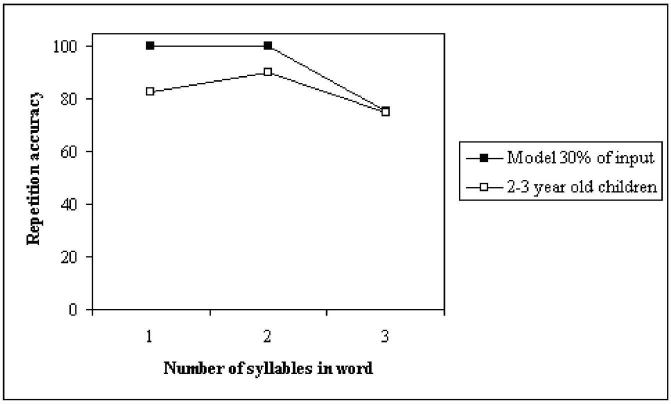

Figure 3: Word repetition accuracy for 2-3 year old children and the model.

The children may not perform at ceiling for one and two-syllable items because of noise during either recognition or articulation of the item. For example, simple nonwords such as *nate* may be perceived by the child to be a real word (e.g., *mate*) and therefore mis-articulated (similarly real words such as *hate* could also be mis-articulated). The studies by Gathercole and colleagues do not perform any analysis of error types so it is difficult to ascertain why children perform poorly for short words and nonwords. However, the model supports the hypothesis of mis-articulation on two counts. First, in performing at ceiling for one and two-syllable items, the model suggests that all 2-3 year old children should be capable of repeating back one and two-syllable words and nonwords. Second, the model provides a very good match for the three-syllable words and nonwords, and items of this length are very unlikely to be mis-attributed to being other words (because they will share relatively few characteristics with other words of the same length, unlike shorter items).

The results show that the EPAM-VOC model can produce repetition results which are comparable to young children using a simple learning mechanism and naturalistic input. In particular, repetition performance for three syllable items is closely matched by the model. The simulation also raises questions about children's performance on repetition tests for one and two-syllable items.

3.2 Simulation of Four to Five Year Old Children

Carrying out a simulation for each set of mother's utterances is not expected to provide a representative sample of the input that 4 and 5 year olds receive, because by this age the children are beginning school, and beginning to read. Each of the nine mother's utterances were therefore matched to a random selection of words from the CMU Lexicon database on a 50/50 basis for use as input. However, in order to simulate the difference between 4 and 5 year old children in terms of the amount of language they will have heard, 60 percent of the input and 80 percent of the input was seen by the model respectively. For example, for Anne, there were 31,393 mother's utterances. A random sample of half of these (15,696 utterances) were taken together with 15,696 random lexicon words. The simulation of 4 year olds used 60 percent of this resulting file and the simulation of 5 year olds used 80 percent of it. Nine such files were created (one for each mother), resulting in nine simulations. Each simulation presented the model with an equal amount of mother's utterances and lexicon words. The phonological store capacity was reverted back to 2,000 ms based on the assumption that the phonological store reaches full capacity by 4 years of age. Figures 4 and 5 show the comparisons of the results of the simulations and the children, for single and clustered nonwords.

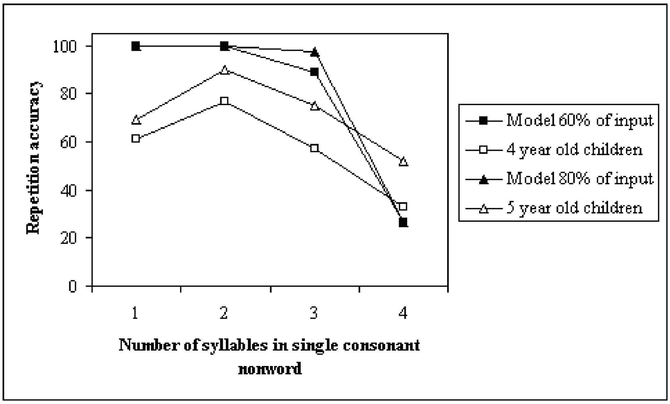

Figure 4: Single consonant NWR accuracy for 4 and 5 year old children, and the model.

The children's performance for one-syllable nonwords is actually worse than for two-syllable nonwords (which should be more difficult). The poor performance for children's repetition of one-syllable nonwords may be due to the acoustic characteristics of their monosyllabic stimuli (Gathercole & Baddeley, 1989), which is consistent with the mis-articulation hypothesis suggested earlier.

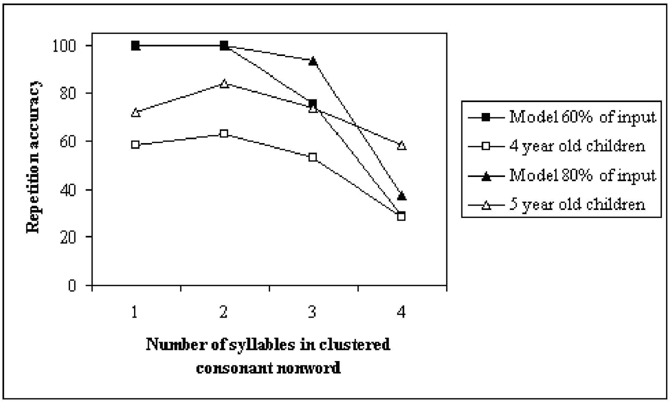

Figure 5: Clustered consonant NWR accuracy for 4 and 5 year old children, and the model.

Although at first glance the fit between the model and the children may not look substantial, there are in fact significant correlations between the 60 percent model and 4 year olds ($r(6)=0.932$, $p<.01$) and between the 80 percent model and 5 year olds ($r(6)=0.847$, $p<.01$).

The model again performs at ceiling for one and two-syllable nonwords. However, the three and four-syllable performance by the model is interesting. The model over-performs for three-syllable nonwords and has a tendency to under-perform for four-syllable nonwords. The three-syllable nonwords average 7.4 phonemes whereas the four-syllable nonwords average 10.1 phonemes. In order to repeat four-syllable nonwords correctly, the model has to chunk up large groups of phonemes that make up the nonword, whereas relatively few have to be chunked to correctly repeat three-syllable nonwords. Building up large chunks in the model is very dependent on the variety of the input that the model sees - the more varied the input, the more rich the chunks in the model. Under-performing on four-syllable nonwords therefore suggests a lack of variation in the input. *This highlights the role of the input as a mediating factor in repetition performance.* The problem for the model, given that the variation of the input is critical, is in determining the type of input that a 4 or 5 year old child will have heard. Clearly, this is an impossible task and any attempt to replicate the input will be a crude approximation. For example, even though the model receives half of the mother's utterances as input, this only constitutes 3,046 different words on average. The lexicon words are used as an attempt to bolster this amount, but clearly they fail to replicate the diversity of input that 4 and 5 year old children receive. The model thus provides a reasonable approximation of repetition performance based on what would seem to be a reasonable, but not perfect, approximation of the input. The results suggest that a more realistic input would produce a good match to the data.

4 Discussion

The simulations have shown that the EPAM-VOC model is able to approximate the NWR performance of both 2-3 and 4-5 year old children. The model accomplishes this by using a combination of a simple learning mechanism, naturalistic input, and a simple implementation of the phonological loop. This represents a parsimonious approach to learning novel sound patterns. In particular, the model is able to give a specific account of how existing vocabulary knowledge influences the learning of new sound sequences.

The EPAM learning mechanisms are very sensitive to the input that the model receives. This allows the model to make very precise predictions. For example, the sound patterns of the most frequent words in the input will be learnt first. New words which consist of frequent sound patterns will be learned quicker than new words which consist of infrequent sound patterns. The model also allows comparisons of how different approaches to memory can affect learning. A time based store can be compared to a chunk based store (e.g., Miller, 1956) and the effects on learning can be examined.

One aspect of the model that does not correspond to children's vocabulary learning is that the model merely learns sequences of sound patterns - it does not learn words per se. Although the resulting discrimination network (after training) will include a lot of vocabulary, there will also be sound patterns that do not correspond to actual words (for example "W AH1 T AH0" from the beginning of "What about that"). The child must therefore process the input in a more discerning way than the model does, in order to determine word boundaries. This process of "segmentation" is very important and has attracted a great deal of interest in its own right (Brent & Cartwright, 1996; Kazakov & Manandhar, 2001; Perruchet & Vintner, 1998) (see Jusczyk, 1999, for a review). Clearly the model presented here represents first steps in the computational modelling of vocabulary learning, with the next steps involving how to incorporate the processes of segmentation.

This work represents a new modelling research program which aims to examine the extent to which the linguistic input a child receives can account for the child's vocabulary development. While the detail of the simulations could be improved, an important contribution of this paper is to provide mechanisms showing how the phonological store links to LTM. The phonological store was shown to mediate LTM learning by limiting the amount of phonemes that could be learnt in LTM. In turn, LTM mediated how much information could be represented in the phonological store by chunking phonemes such that more information could be stored over time. In addition, the model, which learns from naturalistic input, has been used to explain a variety of other phenomena using very similar mechanisms to those employed here. The use of the same computational approach in various domains such as vocabulary learning, the acquisition of expertise, verbal learning, and the acquisition of syntactic categories, ensure a model that has few degrees of freedom.

5 Acknowledgements

The authors would like to thank two anonymous reviewers for their helpful comments during the preparation of this paper.

6 References

Arslan, A. N., & Egecioglu, O. (2004). Dictionary look-up within small edit distance. *International Journal of Foundations of Computer Science, 15,* 57-71.

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

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation*, (Vol. 8), 47-90. New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21,* 85-123.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93-125.

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

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

Cowan, N., & Kail, R. (1996). Covert processes and their development in short-term memory. In S. E. Gathercole (Ed.), *Models of short-term memory,* 29-50. Hove, UK: Psychology Press.

De Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess: Studies in the heuristics of the professional eye.* Assen: Van Gorcum.

Dunn, L. M., & Dunn, L. M. (1982). *British Picture Vocabulary Scale.* Windsor: NFER-Nelson.

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8,* 305-336.

Fredkin, E. (1960). Trie memory. *Communications of the ACM, 3,* 490-499.

Freudenthal, D., Pine, J. M., & Gobet, F. (2005). Resolving ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research, 6,* 17-25.

Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. *Developmental Psychology, 29,* 770-778.

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. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81,* 439-454.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Hove, UK: Lawrence Erlbaum Associates.

Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.), *Models of short-term memory,* 73-100. Hove, UK: Psychology Press.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12,* 349-367.

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.

Gereveni, A., Perini, A., Ricci, F., Forti, D., Ioriatti, C., Mattedi, L., & Monetti, A. (1992). POMI - An expert system for integrated pest-management of apple orchards. *AI Applications, 6,* 51-62.

Gobet, F. (1993). A computer model of chess memory. In *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gobet, F., Freudenthal, D., & Pine, J. M. (2004). Modelling syntactic development in a cross-linguistic context. *Proceedings of the COLING 2004 Workshop "Psychocomputational Models of Human Language Acquisition"* (pp. 53-60). Geneva: COLING.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5,* 236-243.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24,* 651-682.

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

Jones, G., Gobet, F. & Pine, J. M. (2000). A process model of children's early verb use. *Proceedings of the Twenty Second Annual Meeting of the Cognitive Science Society* (pp. 723-728). Mahwah, NJ: Lawrence Erlbaum Associates.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3,* 323-328.

Kazakov, D., & Manandhar, S. (2001). Unsupervised learning of word segmentation rules with genetic alogorithms and inductive logic programming. *Machine Learning, 43,* 121-162.

Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-93.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition*, 19-35. Hillsdale, NJ: Lawrence Erlbaum Associates.

Park, S., & Hyun, K. H. (2004). Trie for similarity matching in large video databases. *Information Systems, 29,* 641-652.

Perruchet, P., & Vintner, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*, 246-263.

Richman, H. B., & Simon, H. A. (1989). Context effects in letter perception: Comparison of two theories. *Psychological Review, 96,* 417-432.

Richman, H. B., Staszewski, J., & Simon, H. A. (1995). Simulation of expert memory with EPAM IV. *Psychological Review, 102,* 305-330.

Simon, H. A., & Gilmartin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5,* 29-46.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2000). The role of performance limitations in the acquisition of 'mixed' verb-argument structure at stage 1. In M. Perkins & S. Howard (Eds.), *New directions in language development and disorders.* Plenum.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory and Cognition, 13,* 193-201.