JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 27, 1613-1639 (2011)

Towards Understanding Child Language Acquisition: An Unsupervised Multimodal Neural Network Approach*

ABEL NYAMAPFENE
College of Engineering, Mathematics and Physical Sciences
University of Exeter
Exeter, EX4 4QF, UK

This paper presents an unsupervised, multimodal, neural network model of early child language acquisition that takes into account the child's communicative intentions as well as the multimodal nature of language. The model exhibits aspects of one-word child language such as generalisation to new and unforeseen utterances, a U-shaped learning trajectory and a vocabulary spurt. A probabilistic gating mechanism that predisposes the model to utter single words at the onset of training and two-words as training progresses enables the model to exhibit the gradual and continuous transition between the one-word and two-word stages as observed in children.

Keywords: child language acquisition, one-word child language stage, two-word child language stage, unsupervised multimodal neural network, neural multinet, Hypermap

1. INTRODUCTION

How children learn to speak is still one of the most important problems in cognitive modelling, and, in the absence of a generally accepted theory of language acquisition, computer-based models (or computational models) capable of acquiring languages on the basis of exposure to linguistic data may help to find out how the language learning process could work [1].

Computational models enable proposed theories to be evaluated and improved through computer simulation [2]. Firstly computational models improve the clarity with which theories are proposed. This is because a theory has to be stated clearly enough for it to be implemented on a computer. Secondly, when the computational model is run, comparisons can be made of the model's output with empirical data. Any discrepancies between the two can lead to a reformulation of the proposed theory. Moreover, computational models are flexible, and this enables new hypotheses to be made and tested out.

Neural network models are based on our current knowledge of computation within the brain, and their objective is to generate the observed behaviour using "brain-like processing". The idea behind neural network modelling of cognitive processes is that the functional processes and representations found in a cognitive system are likely to be constrained by the sorts of computations that the neural substrate can readily achieve [3]. Consequently, it is generally believed that the behaviour of neural network models helps to shed light on how the brain may implement cognitive tasks.

This paper presents an unsupervised multimodal neural network model of child language acquisition that simulates the transition of child language from the one-word utterance stage to the two-word utterance stage. The term *multimodal* is defined by the

Received December 23, 2009; revised March 15, 2010; accepted May 6, 2010.
Communicated by Chin-Teng Lin.

Oxford English Dictionary as "characterised by several different modes of occurrence or activity; incorporating or utilising several different methods or systems". Multimodal processing enables associations to be made between the various information modes, making it possible to translate from one mode to the other. For instance, a person listening to someone's voice can be triggered to remember the visual features of the speaker. It is now widely accepted that cognitive processes are generally multimodal in nature as evidenced by the ability of the brain to integrate information from entirely different input modalities [4-6]. For instance, child language acquisition is now viewed as a multimodal task in which different modalities such as perceptual entities, communicative intentions, and speech, are inextricably linked [7-9]. During the process of language acquisition, the child learns to decipher this inextricability and eventually emerges from preverbal communication, to single word utterances, then to two-word utterances and finally to full-blown native language.

2. RELATED COMPUTATIONAL WORK

Our model is not the first neural network model for child language acquisition to take multimodality into consideration. Plunkett, Sinha, Moller and Strandsby [10] proposed a supervised neural network model of early lexical development that learnt to associate linguistic labels with visual patterns of dots. From a language acquisition perspective, children do not receive constant feedback about what is correct or incorrect in their speech, or the kind of error corrections they must make on a word-by-word basis as suggested by supervised training algorithms. Rather, language acquisition in the natural setting is regarded as being essentially a self-organising process that proceeds without explicit teaching [9] and self-organising maps (SOMs) [11] in which learning proceeds in an "unsupervised manner" without the use of explicit teaching signals offer better biological plausibility than supervised neural networks [12]. Consequently, in contrast to Plunkett, Sinha, Moller and Strandsby, and in line with more recent models of child language acquisition, this paper proposes a self-organising approach to modelling child language acquisition.

The Hebbian-linked SOMs architecture by Miikkulainen [13, 14] has recently been used as the basis for a number of multimodal self-organising neural network models of child language acquisition [12, 15, 16]. This approach to modelling child language acquisition is inspired by the widely held opinion that multimodal tasks are processed by the brain by means of separate modality-specific modules that are linked to each other [17-22]. In the Hebbian-linked SOMs architecture, the SOMs are trained to encode different data domains, and their outputs work as inputs to the links associating them. The connection weights of these links are updated through Hebbian learning [23] in such a manner that a mapping is established between the data domains separately encoded by each of the linked SOMs.

However, the notion that the brain represents and processes information as separate, modality-specific concepts is increasingly being challenged by the alternative notion that uni-modal processing routes within the brain ultimately converge onto a single amodal set of conceptual representations [24, 25]. Recent advances in neuroimaging techniques to detect and measure brain activity appear to support the idea that inputs from the dif-

ferent modalities converge onto the same set of modal-nonspecific (amodal) representations. For instance, in a positron emission tomography (PET) neuroimaging analysis of word and picture processing, Vandenberghe, Price, Wise, Josephs, and Frackowiack [26] concluded that the conceptual representation system was undifferentiated by modality. More recently, Bright, Moss, and Tyler [27] have carried out a meta-analysis of four PET studies, two using pictures as stimuli and two using words, and they have concluded that the representation of information at the semantic and/or conceptual level is not differentially affected by the mode of input (words or pictures).

With regard to child language acquisition, Bloom [28] has suggested that when an infant hears a word (and perhaps a larger speech unit like a phrase or sentence), the word is entered in memory along with other perceptual and personal data that include the persons, objects, actions and relationships encountered during the speech episode. The model presented in this paper is based on the counterpropagation network and it explicitly encodes early child language utterances as composite multimodal elements each comprising the phonological utterance, communicative intention and the referent perceptual entity. We believe that this approach is more in conformity with Bloom's ideas than Hebbian-linked self-organising map models where the different aspects of a child's utterance are separately encoded in the different self-organising maps.

By and large, current models of child language generally try to find a mapping between the child's utterances and the perceptual entities and events the child wishes to speak about without taking into account the child's perceived communicative intentions into account. For instance, the Plunkett, Sinha, Moller and Strandsby model [10] tries to find a mapping between pictorial labels and the linguistic representation of these labels. Similarly, Hebbian-linked self-organising map models of child language acquisition simulate child language at the one-word stage simply as a mapping between linguistic word-form labels (*i.e.* names) and associated perceptual entities and events [12, 16] and ignore the child's perceived communicative intentions. However, in reality, child language acquisition involves more than acquiring the ability to use linguistic word-forms to name perceptual entities and events. For instance, in addition to naming, children at the one-stage have other communicative intentions such as the desire to comment on the events and perceptual entities within their environment, making requests to their caregivers, registering their objection about something and so on [7-9]. As Elman [29] notes, children have drives, desires, as well as things that draw their attention and things they ignore. In short, children are actors in the process of cognitive development, and not just passive absorbers of external stimuli who are completely at the mercy of their environment. Hence, by ignoring a child's communicative intention, current computational models of child development have ignored the actor role of children in the process of cognitive development.

A neural network model of child language that takes communicative intentions into consideration was built by Abidi and Ahmad [15] in the early 1990's. This system, known as ACCLAIM (A Connectionist Child Language development and Imitation Model), is a multi-net neural network comprising both supervised and unsupervised neural networks specifically built to simulate child language development at the one-word and two-word stages. Within this multi-net, two neural networks are used to simulate one-word child speech. The first neural network consists of a pair of self organising maps that are Hebbian-linked to each other. One of the self-organising maps encodes percep-

tual entities whilst the other map encodes linguistic word forms. The second network is a supervised network, trained using the backpropagation algorithm to map non-naming communicative intentions to their corresponding one-word utterances. However, current opinion in child language suggests that naming is just one of the many communicative tasks children get to acquire on their way to full mastery of their first language [7, 9, 30]. Consequently, it would be logical to assume that the same set of brain networks that are involved in mapping names to their corresponding entities will also be involved in mapping the other communicative intentions to their corresponding single word utterances. In contrast to ACCLAIM, the model of child language acquisition presented in this paper consists of a common unsupervised neural network that learns to simulate all the different types of utterances made at the one-word stage.

With regard to two-word child language, Gleitman and Newport [31] have noted that children at this stage appear to determine the most important words in a sentence and, almost all of the time, use them in the same order as an adult would. Brown [32] has also identified a small set of eight basic semantic relations in children's two word utterances that appear to determine their word order. To generate two-word utterances in ACCLAIM, two semantically related concepts are applied to the Hebbian-linked SOM architecture responsible for mapping concepts to linguistic word forms. The two words corresponding to the applied concepts are then fed into a supervised neural network trained to output word pairs in their correct order. However, as we have noted earlier, child language acquisition is regarded as an essentially unsupervised process. Consequently, in the model presented in this paper we simulate the two-word stage with an unsupervised temporal neural network instead.

ACCLAIM has modelled the one-word and two-word child language stages as two separate processes that are not linked to each other in time. However, early child language acquisition occurs in stages, and the transition from one stage to the next is generally gradual and continuous [33]. In the model presented in this paper, a time-dependent probabilistic approach is used to switch processing from the one-word stage network to the two-word stage network as training progresses.

3. THE PROPOSED COMPUTATIONAL MODEL

The model for one-word to two-word child language acquisition reported in this paper consists of a modified counterpropagation network that simulates child language at the one-word stage and a temporal Hypermap that simulates child language at the two-word stage. These two networks are linked together by a time-dependent probabilistic gating network that is initially biased towards the counterpropagation network outputs. During simulation, the two networks are trained simultaneously, and as training progresses the gating network shifts output preference from the counterpropagation network to the temporal Hypermap.

3.1 The Modified Counterpropagation Network

As shown in Fig. 1, the original full counterpropagation network [34] provides bidirectional mapping between two sets of input patterns. It consists of two layers, namely

the hidden layer, trained using Kohonen's self-organising learning rule [11], and the output layer which is based on Grossberg's outstar rule [35]. The Kohonen layer encodes the mapping between the two sets of patterns whilst the Grossberg layer associates each of the Kohonen layer neurons to a set of target output values. Each Kohonen neuron has two sets of weights, one for each of the two patterns being mapped to each other.

In the model proposed in this paper, the Kohonen layer of the counterpropagation network is used to associate the corresponding input modal vectors. For a multimodal input comprising m modes, the Kohonen layer neurons will each have m modal weight vectors, with each vector corresponding to a modal input (Fig. 2). After training, when a modal input is applied to the network, the modal weights of the winning neuron will contain information on all the other modal inputs of the particular modal input. By reading off these weights, we can get the corresponding modal inputs to a particular modal input.

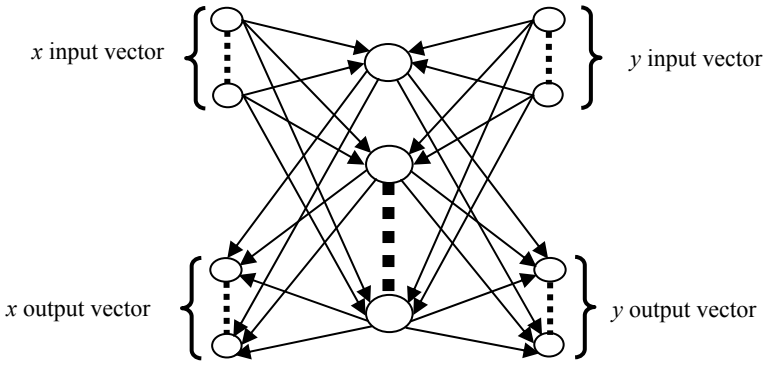

Fig. 1. Model of the full counterpropagation network showing the mapping between two different data types.

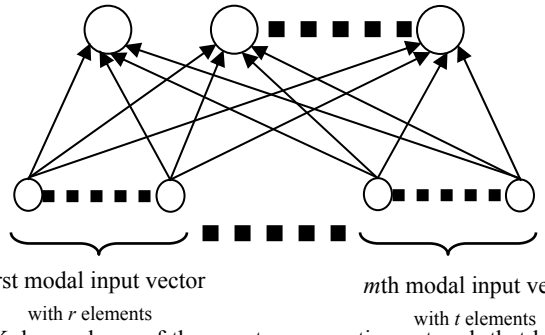

Fig. 2. Model of the Kohonen layer of the counterpropagation network that has been modified to simultaneously accept multiple input data types.

For each multimodal input vector, the winning neuron is the one with the least overall Euclidean distance between its individual modal weight vectors and the corresponding modal component vectors of the multimodal input. To compute the overall Euclidean distance for each neuron we first determine the normalized squared Euclidean distance

for each modal input:

$$d_j^2 = \frac{1}{n}\left\|\bar{x}_j - \bar{w}_j\right\|^2 = \frac{1}{n}\sum_{k=1}^{n}(x_{jk} - w_{jk})^2$$

where $\bar{x}_j$ and $\bar{w}_j$ are the modal input and weight vector respectively with n elements each. The overall Euclidean distance for the neuron is then obtained as follows:

$$D = \sqrt{\sum_{j=1}^{m} d_j^2}.$$

After selecting the winning neuron, the weights of the modified counter propagation network are updated in the same way as for the basic SOM.

Table 1. Audio-visual data percentage correct classification after 200 cycles training.

Network		Monomodal Classification		Crossmodal Classification	
Map size	Type	Audio to Audio	Visual to Visual	Audio to Visual	Visual to Audio
8 × 8	Hebbian-linked	90	84	34	28
	Counterpropagation	86	72	82	76
10 × 10	Hebbian-linked	98	74	26	38
	Counterpropagation	60	76	64	72
12 × 12	Hebbian-linked	96	78	30	16
	Counterpropagation	92	88	92	88

Table 2. Audio-visual and phonological-semantic crossmodal MSE after 200 cycles.

Network		Data			
		Audio-Visual		Phonological-Semantic	
Map size	Type	Audio to Visual	Visual to Audio	Phonology to Semantics	Semantics to Phonology
8 × 8	Hebbian-Linked	4.07	14.33	0.64	9.86
	Counterpropagation	3.45	9.63	0.80	9.08
10 × 10	Hebbian-Linked	3.88	12.89	0.61	9.34
	Counterpropagation	3.46	9.55	0.77	8.23
12 × 12	Hebbian-Linked	2.98	14.54	0.55	8.05
	Counterpropagation	4.68	8.68	0.68	9.51

3.1.1 Evaluation of the modified counterpropagation network

Two bimodal datasets were used to assess the crossmodal properties of the modified counterpropagation network using the Hebbian-linked SOMs architecture by Miikkulainen as a benchmark [36]. The crossmodal properties of interest were the crossmodal mean squared error (MSE) and crossmodal data classification performance. To assess these two properties an input of a given modality is applied to the trained bimodal network and the corresponding output in the other modality is retrieved. The crossmodal mean squared error (MSE) is the MSE between the retrieved vector and target vector of

the output modality. The classification performance refers to the percentage of input-output pattern pairs in which both the input and output are categorised as belonging to the same class despite their different modalities.

The first dataset, due to de Sa and Ballard [37], is composed of auditory and visual representations of consonant – vowel utterances. This dataset is made up of 96 repetitions of /ba/, /va/, /da/, /ga/ and /wa/, making up a total of 480 audio-visual data items. The auditory representation of each utterance is a 216 dimension feature vector, whilst the corresponding visual representation is a 125 dimension feature vector. A training set of 400 patterns was created by selecting at random 80 patterns from each consonant-vowel set. The remaining patterns, which totalled 80 patterns, constituted the test set which was used for crossmodal performance evaluation.

The second dataset, due to Li and MacWhinney [16], consists of the phonological and semantic representations of 500 words extracted from the Toddler's List found in the MacArthur-Bates Communicative Development Inventories (CDI) [38]. Each phonological representation comprises 54 phonetic features, whilst each semantic representation comprises 200 features which are a concatenation of features derived from word co-occurrence probabilities and those derived from WordNet, a computational thesaurus that provides semantic classification of the English lexicon [39]. A training set comprising the phonological and semantic representations of 400 words was randomly selected from the data set. The test set used for crossmodal performance evaluation consisted of the phonological and semantic representations of the remaining 100 words.

Crossmodal classification performance was assessed using the de Sa and Ballard data whilst crossmodal mean squared error (MSE) was assessed using both data sets. For each of the two datasets, modified counterpropagation networks and Hebbian-linked networks of map size 8×8, 10×10, and 12×12 were trained over 200 cycles. A learning rate with an initial value of 0.9 that decayed linearly over the training period to a constant value of 0.01 was used in each instance.

As shown in Table 1, compared to the modified counterpropagation network, the Hebbian linked SOM architecture gave higher percentages of correct classifications for both the audio-to-audio and visual-to-visual monomodal classification tasks. However, the modified counterpropagation network had higher significantly percentages of correct crossmodal classifications for both audio-to-visual and visual-to-audio mapping. These results suggest that using a single map to encode multiple data modes, as is the case with the modified counterpropagation network, has an adverse effect (possibly due to modal interference) on monomodal classification, but significantly improves crossmodal classification.

With regard to output mean squared error (MSE), Table 2 shows that crossmodal performance with respect to MSE is asymmetrical for both architectures. This may be due to the extent to which a particular data modality is able to map uniquely to the other modality. For instance, the same data item in a given modality may map to more than one data element in the other modality, which means that the expected output, though it may be valid, may not be the actual one expected. Both networks had broadly the same performance on the Li and MacWhinney phonological-semantic dataset, with the modified counterpropagation network exhibiting a marginally better performance on the de Sa and Ballard audio-visual dataset.

On the basis of this assessment, it can be concluded that the modified counter-

propagation network is a suitable candidate for unsupervised multimodal neural network processing tasks such as child language acquisition.

3.2 The Temporal Hypermap

Kohonen [40] proposed the Hypermap, a self-organising map that could be trained to recognise a pattern that occurs in the context of other patterns. In the Hypermap approach, contextual information is used to select a subset of nodes from the network. Pattern information is then used to determine the best-matching node from amongst the subset of nodes selected using contextual information. Sequence processing is achieved by using, for each pattern in the sequence to be encoded, the most recent patterns prior to it, or some processed form of them, as context. However, it is difficult to recall an entire sequence in its correct temporal order using partial sequence data, or to store multiple sequences and recall them individually using contextual information.

Araújo and Barreto [41] extended the Hypermap by incorporating lateral weights to encode the temporal order of items in a sequence, and by splitting the context vector into a *time-varying context* vector, which could be used to identify a particular sequence item, and a *fixed context* vector, which could be used to identify a particular sequence. A particular sequence item can be recalled by presenting to the network its fixed context, time varying context and pattern vector. Following the identification and recall of this item, all the other sequence items coming after it in the sequence are recalled in their correct temporal order by means of lateral Hebbian connections. However, the absence of a short-term memory mechanism in Araújo and Barreto's network means that the network can not be used to anticipate and recall an entire network on the basis of partial sequence data. In addition, the network can not recall a stored sequence in its correct temporal order on the basis of fixed context information alone unless the fixed context vector is used as the pattern vector of the first item in the sequence.

This paper proposes a temporal Hypermap that learns to store and recall without catastrophic interference multiple sequences in which patterns may recur within and across several sequences. The temporal Hypermap achieves this by incorporating short term memory and inhibitory links. The short term memory dynamically encodes the time-varying context of each sequence item, making it possible to recall a stored sequence from its constituent subsequences. Wang and Yuwono [42, 43] have previously used short term memory to encode sequences, but unlike us, they have used a distributed representation, which, owing to its susceptibility to symbol interference, places limitations on the sequence length that can be stored and recalled in intact form. The inhibitory links incorporated in the temporal Hypermap establish the temporal order of the items in a stored sequence, thereby making it possible for a sequence to be retrieved in its correct temporal order when its corresponding fixed context vector is applied to the temporal Hypermap.

3.2.1 The temporal hypermap architecture

In order to analyse and describe the model, we have adopted the sequence processing terminology by D. Wang and M. A. Arbib [44, 45] whereby a sequence is defined as a finite set of pattern items:

S: $s_1 - s_2 - \ldots - s_n$,

where $s_j, j = 1, \ldots, n$ is a component of the sequence S and the length of the sequence S is n. If a sequence includes repetitions of the same subsequence in different contexts, it is called a complex sequence; otherwise it is a simple sequence.

When retrieving complex sequences, the correct successor can be determined only by knowing components prior to the current one. The shortest prior subsequence that unambiguously determines a sequence component s_j in a sequence S is defined as the *time varying context* of the component, and the length of this time-varying context is referred to as the degree of the component. The time varying contexts of the individual components in a sequence may have different degrees and the maximum component degree specifies the degree of the whole sequence. Where multiple sequences are being simultaneously processed, the maximum sequence degree specifies the degree of the whole network.

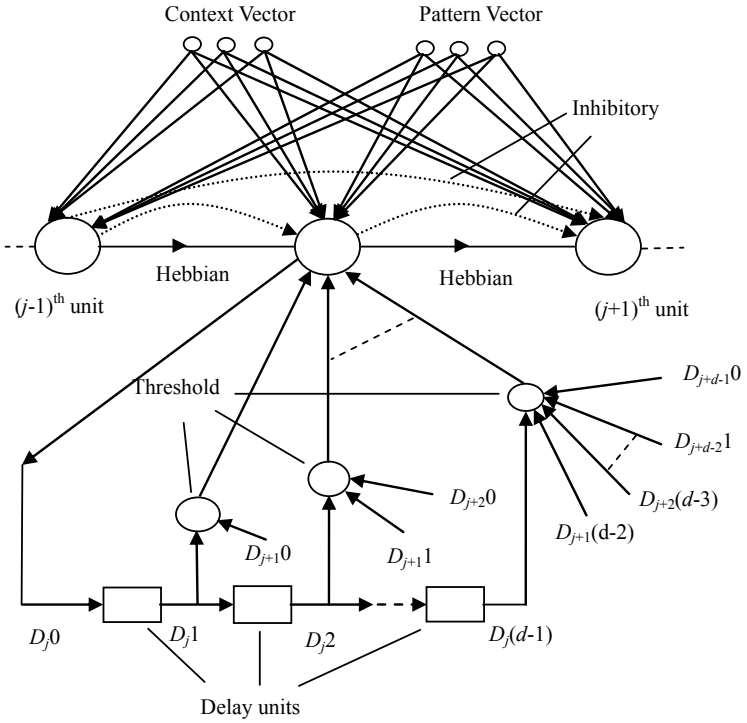

Fig. 3. Temporal Hypermap neuron and links to adjacent neurons. The current neuron is referred to by the subscript j, whilst the next neuron in the sequence is referred to by the subscript $j + 1$, and the immediately prior neuron is referred to by the subscript $j - 1$.

As shown in Fig. 3, the temporal Hypermap consists of a map whose neurons each have two sets of weights – context weights and pattern weights. Pattern weights encode sequence components whilst context weights encode the sequence identity. Each neuron has a short term memory mechanism comprising a tapped delay line [46] and threshold

logic units [47] whose purpose is to encode the time varying context of the sequence component encoded by the neuron. Consecutive neurons in a sequence are linked to each other through Hebbian weights [23], and inhibitory links extend from each neuron to all the other neurons coming after it in the same sequence. The Hebbian links preserve temporal order when spreading activation is used to recall a sequence [48], whilst the inhibitory links preserve temporal order when fixed context information is used to recall an entire sequence [49, 50].

3.2.2 Encoding sequence components with the temporal Hypermap

A sequence is stored by applying its components to the temporal Hypermap in their correct temporal order. Each temporal Hypermap neuron encodes exactly one sequence item. This prevents ambiguity the retrieval of sequences in cases where identical pattern vectors are repeated in the same sequence or occur in other sequences. Each neuron has a responsibility variable that is set to one if the neuron has been assigned to a sequence component; otherwise it remains reset at zero [41]. When an input is applied, only those neurons with responsibility variables equal to zero are allowed to compete.

A winner-take-all mechanism is used to assign each sequence component to a free neuron as follows: On applying the context and weight vectors of a sequence component to the temporal Hypermap, the network computes the context distances and pattern distances of all the unassigned neurons, *i.e.* those neurons in which the responsibility factor is set to zero. The context distance for a neuron is the Euclidean distance between the neuron's context weight vector and the input component context vector, whilst the pattern distance is the Euclidean distance between the neuron's pattern weight vector and the input component pattern vector. The product of these two distances gives the contextually weighted pattern distance for the neuron. The neuron with the smallest contextually weighted pattern distance is selected as the winner, and its context and the pattern weights are set equal to the context vector and pattern vector of the input sequence item respectively.

3.2.3 Encoding sequence temporal order with inhibitory links

The inhibitory scheme for preserving temporal order during sequence recall through fixed context information operates as follows: When a context vector identifying a stored sequence is applied to the network, all the neurons belonging to the sequence are selected as winners. However, because of the inhibitory links, only the first neuron in the sequence is activated since it is the only one receiving no inhibitory signals from other neurons. In the next time-step when this neuron is deactivated, its inhibitory effect on all the other neurons is removed. Consequently, the next neuron in the sequence is activated since there is no longer any signal inhibiting it. This process of neuron activation and deactivation is repeated across the network, and in this way the sequence is reproduced in its correct temporal order. This scheme was first proposed by Estes [49], and used by Rumelhart and Norman [50] in their model of how skilled typists generate transposition errors.

The inhibitory scheme is implemented as follows: During sequence storage, all the neurons that have been used to store the components of the current sequence are identified through their responsibility and context activation functions which are both set to

one. Inhibitory links are extended from these neurons to the current winning neuron to ensure that the neuron will only be selected only after they have all been deactivated.

The temporal Hypermap uses a one-shot encoding scheme to implement the lateral Hebbian weights between consecutive neurons in a sequence. During sequence encoding, the weights on the unidirectional links between consecutive neurons in a sequence are set to one. Lateral weights between non-consecutive neurons are maintained at a value of zero.

Assuming that a sequence neuron is initially activated, when that neuron is deactivated, a pulse propagates along the unidirectional Hebbian link connecting the neuron to the next neuron in the sequence. This process continues until the final neuron in the sequence has been activated.

3.2.4 Encoding time-varying context for each sequence component

Each neuron has a tapped delay line whose length is set to one less the degree of the network. For instance, if the degree of the network is d, then each neuron will have a delay line with $d - 1$ delay units. Each time an input pattern matches the pattern vector of a neuron, the neuron in question generates a pulse which is applied to the tapped delay line. This pulse propagates along the delay line for $d - 1$ time-steps, after which it is cleared out of the delay line.

At the output of each delay unit is a tap that feeds into a threshold logic unit. There is one threshold logic unit for each delay tap position. As seen in Fig. 4, the inputs to each threshold logic unit are the output of the tap position to which it is connected on its neuron's delay line as well as all the simultaneously active tap positions on later neurons in the sequence. For instance, the threshold logic unit connected to the output of the first delay unit in a neuron will also receive as input the non-delayed output of the next neuron in the sequence; the threshold unit connected to the output of the second delay unit on a neuron will have three inputs, namely the output of the second delay unit on the neuron, the output of the first delay output of the next neuron in the sequence, and finally the output of the second delay unit of the third consecutive neuron in the sequence.

The threshold level of each threshold logic unit is set equal to the number of inputs connected to the logic unit, which is one more than the delay tap position to which it is connected. In addition, the output of each threshold logic unit is scaled to give an output activation value equal to its threshold level. Consequently, when two or more threshold logic units are simultaneously activated in the temporal Hypermap, the threshold logic unit with the highest threshold level wins the competition. With regard to the use of partial subsequences to prompt the recall of a stored sequence, the winning threshold logic unit is the one encoding the shortest subsequence long enough to distinguish between stored sequence items responding to the inputted subsequence. When this threshold logic unit fires, its associated neuron also fires and generates an output and the rest of the sequence is then generated through spreading activation by means of the lateral Hebbian links.

3.2.5 Evaluating the temporal Hypermap

The sequence processing abilities of the temporal Hypermap were assessed using the session titles for the 1994 IEEE International Conference on Neural Networks. This

dataset contains a high degree of overlap among the sequences, and is therefore ideal for assessing the extent to which the network can handle multiple sequences with repeating items. Wang and Yuwono [42] used this dataset to evaluate their own neural network model of sequence processing, known as the anticipation model, and consequently, this data enables the model's performance to be benchmarked against the anticipation model.

The temporal Hypermap stores and correctly recalls each of the eleven sequences following the input of sequence identity vectors. This is in contrast to the anticipation model which only manages to recall without error five of the eleven stored sequences. Like the anticipation model, applying to the temporal Hypermap a subsequence sufficiently long enough to be uniquely identified enables the network to correctly recall the stored sequence starting from the first item of the inputted subsequence. This is because sequence items are stored as chains of associations, each of which is triggered by a unique subsequence. This ability to recall a sequence from an applied subsequence gives the temporal Hypermap the flexibility to generate part of a stored sequence from any position of the sequence. Such flexibility is consistent with human ability to identify and recall common sequences such as familiar songs, melodies, or action sequences when exposed to their segments.

3.3 Realising the Multi-Net Architecture for One-Word to Two-Word Transition

To simulate one-word child language acquisition, Kohonen training [19] is used to train the modified counterpropagation network using multimodal child language data comprising the perceptual information, such as events and entities, that the child wishes to speak about in its environment, the actual single word utterances made by the child, and lastly, the child's perceived communicative intention.

To simulate two-word child language acquisition, the temporal Hypermap is trained using one-short encoding on the perceptual information that the child wishes to speak about in its environment, the actual two-word utterances made by the child, and lastly, the child's perceived communicative intention.

To simulate the transition from the one-word stage to the two-word stage, inhibitory links are extended from the temporal Hypermap to the counterpropagation network. These inhibitory links suppress the counterpropagation network output whenever the temporal Hypermap has an active output. This ensures that only one network is able to generate an output at any given instance. These inhibitory links are controlled by a time-dependent probabilistic gating mechanism that predisposes the counterpropagation network to generate an output during the early stages of simulation, and the temporal Hypermap to generate an output during the latter stages of simulation. In this way, the model output changes gradually from one-word utterance simulation to two-word utterance simulation.

The gating mechanism's likelihood of activating the temporal Hypermap in preference to the counterpropagation network for a given input is governed by the equation:

$$p_u(n) = (1 - p_u(0))\frac{n}{n_T} + p_u(0) \tag{1}$$

where n is the current Kohonen training cycle for the counterpropagation network, n_T is

the total number of counterpropagation training cycles, $p_u(0)$ is the initial probability for making two-word utterances prior to training determined through an analysis of the child language corpus being used for simulation, and $p_u(n)$ is the network's probability for making two-word utterances in the current cycle n.

4. CHILD LANGUAGE SIMULATION DATA

The data used for running the child language simulations is taken from the Bloom 1973 corpus [7], which is part of the Child Language Data Exchange System (CHILDES) corpora [51].

This corpus consists of utterances recorded from a child, Alison, who was born on July 12, 1968. This corpus consists of six samples taken at ages: 1 year 4 months and 21 days, 1 year 7 months and 14 days, 1 year 8 months and 21 days, 1 year and 10 months, 2 years 4 months and 7 days, and 2 years and 10 months. A common setting was used in all the recording sessions. This setting consisted of three pieces of furniture and a rug in front of a blank wall. There was a big wooden Windsor-type double chair that could seat two people comfortably. This is referred to as the "big chair" in the transcription and it was centre stage. To the right of it was a child-size moulded plastic chair, and between the two chairs was a triangular low table.

Each session included a snack with cookies, a container of apple juice, and several paper cups. A group of toys was brought to all of the sessions. These toys consisted of Alison's doll, a metallic dump truck about 30 cm long, and a set of rubber farm animals (bull, cow, calf, horse, colt, lamb and pig). Other toys were used in one or another of the sessions, but not in all of them. These included a jar of bubble liquid, a group of hand and finger puppets, a 12 cm plastic doll wrapped in a blanket and a photograph of a girl in a plastic frame. The snack was carried in a canvas tote bag ("the bag") which was Alison's own and which contained an extra diaper and napkins.

4.1 Dataset for Simulating One-Word Child Language Acquisition

In an investigation of the relation between children's single-word utterances and maternal single-word utterances, Ninio [52] observed that over 90 percent of the children's single word utterances were similar to what their mothers said in similar communicative circumstances. This large overlap of child speech with maternal speech suggests that little error will be incurred if the child's utterances are used as input to the model in lieu of the mother's utterances, which are significantly more complex. On this basis, the single word utterances made by the child Alison in the Bloom 1973 corpus have been selected for use as training data for the model proposed in this paper. For each of these utterances the actual utterance, the perceptual entity being referred to, and the conceptual relation that have been inferred from the discourse between the child and its mother are recorded, as illustrated in Table 3.

4.2 Coding Scheme for Children's Perceptual Entities

Bloom [7] suggests that children employ category level abstraction as well as salient

Table 3. Our interpretation of some of the one-word utterances by the child Alison.

Actions and Mother's words	Alison's Utterance	Perceptual Entity	Conceptual Relation
(A) touching Mother's hip	there	mum	locate object
(M) helps Alison down. (A) turns, trying to get up again	more	chair	request event recurrence
(M) what is this? (A) pointing to chair	chair	chair	name object
(A) takes cookie; reaching with other hand towards others in bag	more	cookie	request object recurrence
(M) where's your juice?	gone	juice	Comment object disappearance

Table 4. Coding scheme for perceptual entities.

Code Block Name	Digit Name	Digit Number	Digit Meaning
Category defining features	Is a person?	D_0	1: Positive confirmation 0: Negative confirmation
	Is an animal?	D_1	
	Is a Vehicle?	D_2	
	Is furniture?	D_3	
	Is clothing?	D_4	
	Is Food?	D_5	
	Is Household Item?	D_6	
	Is Place?	D_7	
	Is doll?	D_8	
Person features	Is Self?	D_9, D_{10}, D_{11}	1st digit set to '1' if feature is present, otherwise it is set to '0'. 2nd digit set to '1' if feature is relevant but absent, otherwise it is set to '0'. 3rd digit set to '1' if feature is irrelevant, otherwise it is set to '0'.
	Does Person Care?	D_{12}, D_{13}, D_{14}	
Animal features	Has furry coat?	D_{15}, D_{16}, D_{17}	
	Is a cat (else dog)?	D_{18}, D_{19}, D_{20}	
	Has horns?	D_{21}, D_{22}, D_{23}	
Food	Is cookie? (else juice)	D_{24}, D_{25}, D_{26}	

features to distinguish between perceptual entities. For instance, children are believed to distinguish various objects by observing aspects such as 'size', 'shape', and 'colour' and even, at times, their 'function'.

A perceptual entity coding scheme based on the suggestion by Bloom is presented in this paper (Table 4). This scheme has eight categories of perceptual entities. These are *person*, *animal*, *vehicle*, *furniture*, *clothing*, *food*, *bodypart* and *place*. Each category is represented by a single binary digit. This is set to '1' if the entity being encoded belongs to that category; otherwise it is set to '0'. Each category has features that distinguish the

entities making it up. Some of these features are specific to the category; whilst others apply equally well to other categories. For instance, in the person category we encode whether the child is referring to self or not, and whether the person being referred to is a caring person or not. For the animal category we encode whether the animal has a distinct sound or not, whether it has horns or not, and whether or not it has a furry coat.

Features that differentiate between the categories are encoded by a single digit position, with a '1' encoding the presence of the feature and a '0' encoding its absence. Category specific features are encoded by three digit positions. The first digit position is set to '1' if the feature is present and to '0' if that particular feature is not present. The second digit position is set to '1' if the feature is relevant to encoding the entity but is missing in that particular entity. The third digit is set to '1' if the feature is not relevant to encoding that particular perceptual entity category; otherwise it is set to '0'.

A 10 × 10 self-organising map trained on some of the perceptual entities from the Bloom 1973 corpus has been used to assess the ability of the proposed perceptual entity coding scheme to cluster together similar perceptual entities and to discriminate between different classes of perceptual entities.

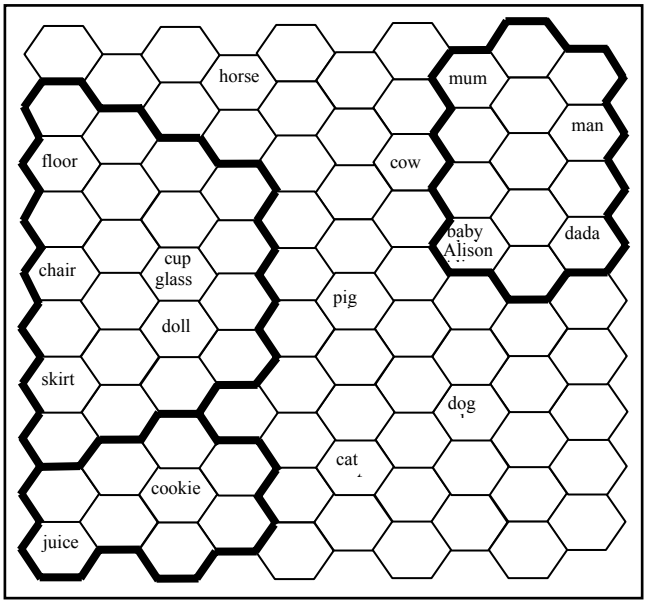

Fig. 4. Visual inspection of some perceptual entities using a 10 × 10 SOM.

As illustrated in Fig. 4, the coding scheme for perceptual entities successfully discriminates between the different classes of entities whilst bringing together similar entities. For instance the person entities namely *mum*, *man*, *baby*, *Alison* and *dada* are clus tered together, whilst the animal entities, namely *cat*, *pig*, *cow* and *horse* are clustered ogether. Also *juice* and *cookie*, which are food entities are separately clustered, whilst furniture and room entities, namely *floor* and *chair* are clustered together.

4.3 Coding Scheme for Communicative Intention

Bloom [7] suggests that at the single word utterance stage the child learns to represent the regularities in his/her environment in terms of the relations between the persons, objects and events in the world. For instance, objects are acted upon, can exist, cease to exist and recur; people do things and they appear and disappear and so on. She refers to these relations as *conceptual relations*. Bloom views language acquisition at the one-word stage as a child's linguistic effort to expresses his/her communicative intention regarding a perceptual entity by uttering a word which encodes the desired conceptual relationship.

A coding scheme that incorporates redundancy for each feature has been developed for the conceptual relations inferred in Alison's utterances (see Table 5). Incorporating redundancy enables the conceptual relation codes to differ by more than one digit position. This improves the neural network's ability to discriminate between the encoded conceptual relations. It also improves the coding scheme's robustness against noise since each non-zero digit is replicated.

Table 5. Scheme for encoding conceptual relations.

Code Block Name	Digit Name	Digit Number	Digit Meaning
Child's communicative intention	Comment	$D_1D_2D_3$	All feature digits set to '1' if feature present, else set to '0'
	Naming	$D_4D_5D_6$	
	Locative	$D_7D_8D_9$	
	Possessive	$D_{10}D_{11}D_{12}$	
	Request	$D_{13}D_{14}D_{15}$	
	Rejection	$D_{16}D_{17}D_{18}$	
Event description features	Recurrence	$D_{19}D_{20}$	All feature digits set to '1' if feature present, else set to '0'
	Existence	$D_{21}D_{22}$	
	Non-Occurrence/ Failure	$D_{23}D_{24}$	
	Disappearance	$D_{25}D_{26}$	
	Cessation	$D_{27}D_{28}$	
Physical State Features	Size	$D_{29}D_{30}D_{31}$	1st digit set to '1' if feature is subject of concern, else set to '0'. 2nd digit set to '1' if feature present, else set to '0'. 3rd digit set to 1 if feature negated else set to '0'.
	Cleanliness	$D_{32}D_{33}D_{34}$	
	Upness	$D_{35}D_{36}D_{37}$	

In this coding scheme, each feature relating to the child's inferred communicative intention is coded by three digits. If the feature is present, all the three digits are set to '1', and if the feature is absent all the three digits are set to '0'. Event description features are similarly encoded using two redundant digits. Physical state features have been encoded using three digits. The first digit is set to '1' if the feature is the subject of interest. The second digit is set to '1' if the feature is present, and the third digit is set to '1' if the feature is absent.

4.4 Coding Scheme for Word Utterances

Words are presumably organised into similarity neighbourhoods in the mental lexicon based on phonological similarity [53]. It is assumed that a similarity neighbourhood includes all the words differing from a given word by single phoneme substitution, deletion or addition. For instance, the similarity neighbourhood for *sit* includes words such as *sip*, *sat*, *hit*, *it*, and *spit*. It is therefore possible that children can analyse sound segments such as syllables and words in terms of their constituent phonemes.

Ladefoged [54] proposed a phonetic coding scheme whereby each phonetic symbol is encoded as a vector quantity of phonetic features consisting of the acoustic features that make up the sound, as well as the articulatory features derived from how the vocal tract, mouth, tongue and associated organs create the sound. In this scheme, consonants are distinguished by *manner of articulation* and *place of articulation*. Manner of articulation can be classified as one of these: *nasal*, *stop*, *fricative*, *approximant*, *and lateral*. Place of articulation can be one of: *bilabial*, *labio-dental*, *dental*, *alveolar*, *palatoalveolar*, *palatal*, *velar*, *and glottal*. Similarly, vowels are distinguished by *height* and *tongue position*. Height can be one of: *high*, *mid-high*, *mid*, *mid-low*, *and low*. Tongue position can be one of: *front*, *central*, and *back*. To this classification, Li and MacWhinney [55] added a third dimension, phoneme status, to explicitly specify whether a given phoneme is a vowel, a voiced consonant or a voiceless consonant.

We have adopted Ladefoged's phonetic coding schemegas modified by Li and MacWhinney and used it to encode Alison's one-word linguistic utterances. Each of the words uttered by Alison can be encoded by four ghonetic features or less, as discovered by Abidi and Ahmad [15]. Consequently, in our coding scheme each word is encoded as a concatenation of four phonetic vectors. Padding is used where the word has less than four phonetic features to keep the length of feature vectors the same. Table 6 illustrates how we encoded some of the words uttered by Alison.

Table 6. Phonetic coding of some child utterances from the Bloom 1973 corpus.

Child Word	Word Phone	Phonetic Numerical Coding											
		Phoneme 1			Phoneme 2			Phoneme 3			Phoneme 4		
		D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3
gone	/gɒn/	0.750	0.921	0.733	0.100	0.250	0.355	0.750	0.684	0.644	0	0	0
more	/mɔː/	0.750	0.450	0.644	0.100	0.250	0.270	0	0	0	0	0	0
there	/ðɜː/	0.750	0.606	0.822	0.100	0.175	0.270	0	0	0	0	0	0
cookie	/kʊki/	1.000	0.921	0.733	0.100	0.250	0.185	1.000	0.921	0.733	0.100	0.100	0.185
no	/nəʊ/	0.750	0.684	0.644	0.100	0.175	0.185	0.100	0.250	0.185	0	0	0

A 10 × 10 self-organising map (Fig. 5) trained on some of Alison's phonetically encoded utterances shows that the phonetic coding scheme adopted in this paper clusters together similarly sounding words whilst differentiating between differently sounding words. For instance, on the SOM differently sounding words like *truck*, *horse*, *uh oh* and *mary* are placed far apart whilst the words *big* and *pig* are placed next to each other, as are the words *this* and *juice*.

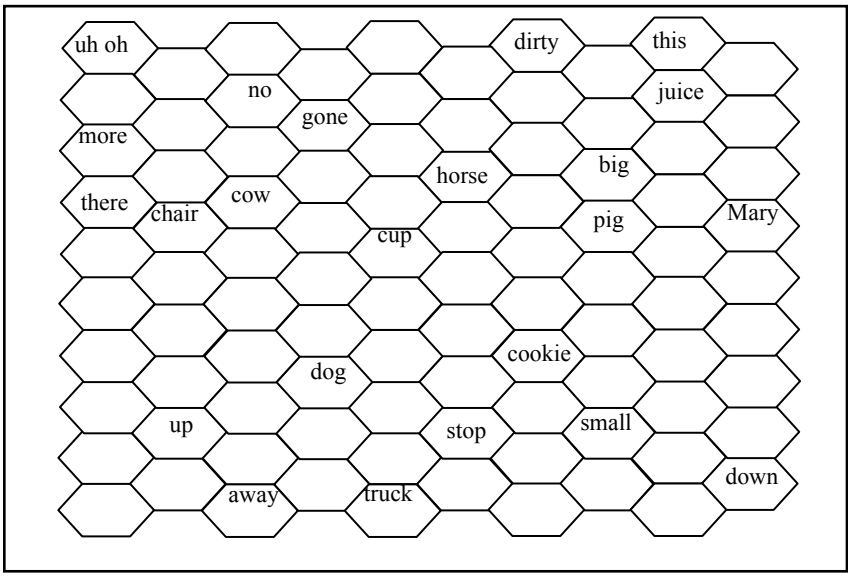

Fig. 5. 10 × 10 Phonology self-organising map.

However, words like *cup*, *cat*, and *cookie*, which all start with the phoneme /k/ are placed near each other although in practice they sound quite different. This is because the encoding scheme operates simply by concatenating phoneme values to come up with a word presentation. This approach assumes that the choice of pronunciation of each phone is independent of all other phones in the word, whereas, in practice, the pronunciation of a phone is affected by its neighbouring phones, other phones within the word and its position in the word, among other things [56].

Words are presumably organised into similarity neighbourhoods in the mental lexicon based on phonological similarity [22]. Ladefoged [23] proposed a phonetic coding scheme whereby each phonetic symbol is encoded as a vector quantity of phonetic features consisting of the acoustic features that make up the sound, as well as the articulatory features derived from how the vocal tract, mouth, tongue and associated organs create the sound. In this paper Li and MacWhinney's [24] adaptation of Ladefoged's phonetic coding scheme has been used.

5. SIMULATING ONE-WORD CHILD LANGUAGE ACQUISITION

From the Bloom 1973 corpus we created and encoded a training set consisting of 30 data triads comprising Alison's one-word utterances, perceptual entities corresponding to the utterances as well as the perceived conceptual relations.

A 10 × 10 modified counterpropagation network was trained on the dataset over 500 cycles. The counterpropagation network size was determined through experimentation to be large enough to sufficiently encode the one-word utterances together with their associated perceptual entities and conceptual relations. Following the submission of an input data triad, the network node which responded with the highest activation level was

deemed to hold the counterpropagation network's representation of that input triad. The associated modal elements held by that particular node can be retrieved by reading off the appropriate modal weights from the node, and using the nearest neighbour approach [57] to determine the exact modal element from the training dataset.

Every 10 epochs throughout the training period of 500 epochs, the training dataset was used to assess the network's ability to recall the correct one-word utterance given a bimodal input consisting of a perceptual entity and associated conceptual relation. As with the Plunkett, Sinha, Moller and Strandsby model [10], the network performance during training resembled children's vocabulary development during their second year. For instance, during the early stages of training, the network exhibited high error rates in recalling one-word utterances associated with bimodal combinations of perceptual entities and conceptual relations. However, as training progressed, the production of correct words suddenly increased until the network was able to generate a correct word for each of the bimodal inputs presented.

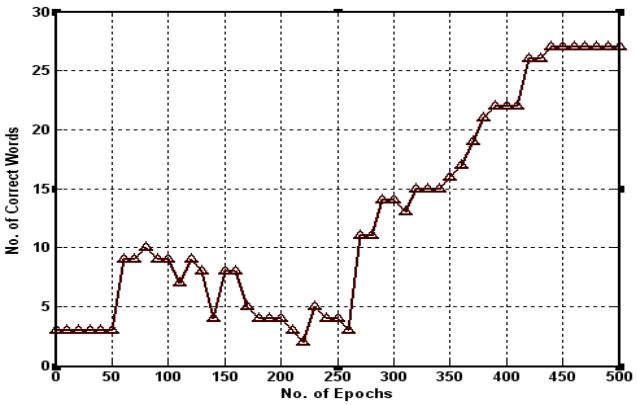

Fig. 6. Plot of correctly recalled words as a function of the number of training epochs.

Fig. 6 shows a learning trajectory for a network with an initial learning rate of value 0.2. Increasing the learning rate caused the network to learn the 30 one-word utterances at a faster rate, and decreasing the learning rate also resulted in the network taking longer to master the one-word utterances. However, for all the values of learning rate the network still goes through an initial period of high error rate, followed by a period of lower error rate, which in turn is followed by a period of high error rate and finally by a period in which the error rate decreases continuously until the training set is mastered.

The learning trajectory for the network as training progresses suggests that children initially master some one-word utterances earlier on during learning, and then as the learning phase continues they undergo a period when the generation of correct one-word utterances deteriorates before the onset of the "vocabulary spurt" which is characterized by a progressive decrease in the error rate. The nature of the learning trajectory exhibited by our model is therefore consistent with the "U-shaped" developmental curves typical of child developmental activities such as language acquisition [10, 58].

To date connectionist models of development have largely been based on supervised learning as opposed to unsupervised learning. This may be due, in part, to the over-

whelming success of the parallel distributed processing (PDP) approach [59, 60] to simulating key features of development such as "U-shaped" developmental curves. Consequently, we believe that the simulation of a "U-shaped" developmental trajectory by our model suggests that unsupervised neural networks may equally be used for modelling cognitive development. As we pointed out in Section 2 of this paper, the use of unsupervised learning in connectionist models of language would be very much in line with current opinion suggesting that language acquisition is essentially an unsupervised process.

We also assessed the ability of the counterpropagation network model to generalise to the correct one-word utterance following the input of a bimodal combination of perceptual entity and associated conceptual relation from a novel dataset of ten utterances independent of the training set. As shown in Table 7 in nine of the ten cases, the modified counterpropagation network generalised to the correct one-word utterance. This suggests that the network successfully generalises to produce appropriate one-word utterances even for novel situations.

Table 7. Utterances produced b the counterpropagation network model in response to novel input.

Conceptual relationship	Novel situation	Expected child utterance	Model's utterance
Comment object disappearance	(A) puts away horse	gone	gone
Comment object cleanliness	(A) indicating mum's Skirt is dirty	dirty	Dirty
Comment upness	(A) stating that horse on chair	up	dirty
Comment object recurrence	(A) Sees more than one pig	more	more
Request object disappearance	(A) refuses Cookie	no	no
Request upness	(A) holding her hand out to Mother to get down	down	down
Request object recurrence	(A) wants another cow	more	more
Reject object recurrence	(A) refuses no longer wants juice	no	no
Locate object	(A) pointing at cookie bag	there	there
Locate object	(A) Pointing at mother	there	there

Key: A – Alison; M – Mother.

6. SIMULATING THE ONE-WORD TO TWO-WORD TRANSITION

From the Bloom 1973 corpus we identified one-word and two-word utterances that seem to address the same communicative intention. However, we hasten to add that the situations a child wishes to speak about at the one-word and two-word stage are not exactly identical, as an analysis of the transcripts indicates. For a start, the transcripts seem to indicate that at the two-word stage the child has a greater ability to formulate and per-

ceive relationships between more concepts in his or her environment. In addition, it appears that the child at the two word stage interacts more with the objects in his or her environment than the child at the one-word stage. Thirdly, the transcripts indicate that the child's vocabulary has grown, and the two word utterances seem to indicate a deeper level of environmental awareness than can be ascribed to one-word utterances. Nevertheless, we identified fifteen pairs of utterances that broadly match each other (Table 8).

Table 8. Corresponding one-word and two-word utterances.

One-Word Utterance		Corresponding Two-Word Utterance		Word Corpus Frequency
Scenario	Alison's Utterance	Scenario	Alison's Utterance	
(A) takes cookie; reaching with other hand towards others in bag	cookie	(A) reaching for cookie box in bag. (A) takes out box of cookies	There cookie	2
(M) what is this? (A) pointing to chair	chair	(M0 pointing to chair. What is this?	That chair	5
(A) drinks juice; takes another cup	more	(M) pours herself juice. (A) picking up empty cup	More juice	2
(A) drinks juice, looks into empty cup, squashes cup; (M) where's the juice? (A) taking cup	gone	(M) pours juice; (A) drinks juice, looks into empty cup. (M) taking cup.	Gone juice	5
(A) sitting down	down	(A) still sitting on floor and looking at Mother who is still standing	sit down	1

Key: A – Alison; M – Mother.

In simulating the gated multi-net, we chose a network size of 8×8 for both the counterpropagation network and temporal Hypermap. This size is sufficient to encode the one-word utterances together with their associated concepts and conceptual relations on the counterpropagation network. It is also sufficient for encoding the two-word utterances on the temporal Hypermap. Each neuron on the temporal Hypermap had a tapped delay line of unitary length, as well as one threshold logic unit to compute the time varying context.

We simulated the gated multi-net for initial transition probabilities of 0.0001, 0.001, 0.01 and 0.1. For each probability value, the simulation was as follows: The counterpropagation network was initially trained for 20 cycles to ensure that all the single word utterances were encoded prior to the onset of the transition to the two-word stage. The two networks were then jointly trained for 30 cycles.

We collected transition data in each of the 30 cycles. First, ten exemplars were randomly selected from the fifteen element training data set. Then each exemplar was applied to the network, which responded by issuing out a one-word or two-word utterance. The number of two-word utterance responses for the ten exemplars was then recorded. This process was repeated twenty times before the network was allowed to undergo another cycle of training.

Figs. 8 and 9 show that as training progresses, the number of two-word utterances increases proportionally, before reaching a saturation value independent of the initial transition probability. The gated multi-net model therefore exhibits a gradual transition from one-word to two-word language utterance as seen in studies of child language acquisition. Prior to saturation, however, the rate of increase of two-word utterances, is dependent on the initial transition probability, with the higher the value of initial transition probability the larger the rate of increase of two-word utterances. Normal children also exhibit different rates of language development, with the rates between different children varying by a year or more [61]. Hence, by varying the initial transition probability value, we manage to simulate the variations in the rate of language development in normal children.

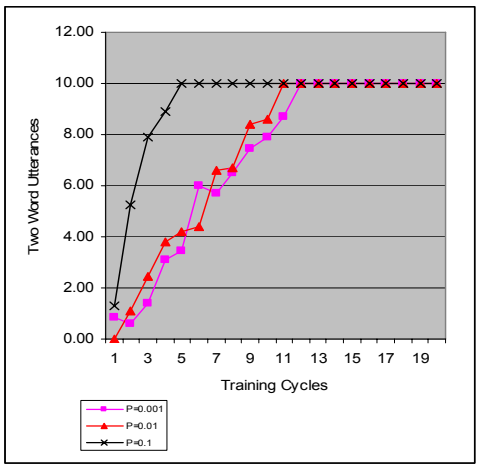

Fig. 8. Output two-word utterances plotted against number of training cycles for the child language data set dataset with the corpus frequency profile.

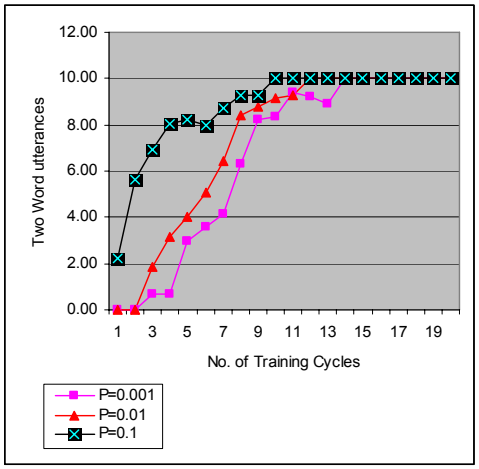

Fig. 9. Output two-word utterances plotted against number of training cycles for the dataset with no duplications.

Comparing Figs. 8 and 9, it is apparent that the gated multinet exhibits a steeper transition from one-word to two-word utterances when trained using the dataset with the corpus frequency profile than when it is trained using the uniform-frequency dataset. The environment to which a child is exposed is generally physically restricted owing to the fact that the child's mobility is limited. Hence, in the natural setting, the perceptual entities the child is likely to be familiar with tend to be those objects and people whom the child interacts with on a more or less daily basis. This set of entities is likely to be much smaller than the range of entities the child could be capable of speaking about. It there-

fore appears that the child's relative immobility tends to restrict the environment to which the child is exposed to, which in turn limits the perceptual entities and events the child is likely to talk about, hence the non-uniform frequency distribution of the child's utterances. In this regard, the model seems to lend support to Elman's suggestion that developmental restrictions on resources may constitute a necessary prerequisite for mastering certain complex domains [27].

7. CONCLUSION AND FUTURE WORK

As pointed out in the introduction to this paper, child language acquisition is complex, and it is very unlikely that the simple model reported in this paper tells the whole story about the child language acquisition process. However, this model has demonstrated that an unsupervised multimodal neural network framework can emulate to some degree some aspects of child language acquisition. As pointed out earlier, brain processes are now widely viewed as being multimodal processes that unfold in an unsupervised manner. In this regard, this model of child language acquisition suggests, albeit in a simplistic way, how the process might unfold in the child's brain in the first two years of life.

With regard to the transition from the one-word stage to the two-word stage, a time-dependent probabilistic gating mechanism has been used to gradually shift child language utterances from single words to two words, in line with observations made in child language studies [33]. It may be argued that such an approach is inappropriate since it gives the impression that these stages are implemented by different brain networks in the developing child. This would suggest that the portion of the brain involved in early child language becomes redundant later in life as cognitive functionality shifts from one part of the brain to another. A more likely scenario is that child language develops as the brain matures. In this case, therefore, the transition of child language from the one-word stage to the two-word stage would suggest the growing maturation and complexity of the brain networks involved in language processing. Such growth would depend on physiological developments within the brain as well as adaptation to the linguistic environment. Consequently, this model of child language stage transition is best viewed as a spatial representation of the development over time of the brain networks involved in language processing. A more realistic model would be to simulate the one-word to two-word transition process using neural network architectures capable of adapting their structure to accommodate the interstage structural and behavioural changes in the developmental data. A study is currently being carried out to realise such a model using neural constructivism [63] to come up with a single unsupervised neural network architecture that can model the development of child language acquisition from the one-word to the two-word stage.

Finally, the model of child language acquisition discussed in this paper takes into account the child's communicative intentions. According to Tomasello *et al.* [64], an action, which in our case is a speech utterance, is carried out in accordance with an intention, and the result of the ensuing action – *i.e.* success or failure, determines the person's subsequent reaction. Child language acquisition can therefore be viewed as a control-theoretic problem in which the response of the caregiver as well as the child's environment interact with the child's drives, emotions and desires, leading to the formation of goals,

which in turn lead to the formation of communicative intentions, which in turn lead to the formation of single-word utterances. These word utterances elicit a response from the caregiver, thereby closing the control-system loop. A neural multi-net approach to model child language acquisition as a control-theoretic problem is currently being investigated.

In conclusion, this model's greatest contribution is possibly recasting child language as an unsupervised multimodal process that can be implemented in a neural network framework. Nevertheless, whilst this may be a possible perspective in which to view the process of child language acquisition, the model still remains, at best, a very rudimentary view of the child language acquisition process.

REFERENCES

1. S. Pinker, "Formal models of language learning," *Cognition*, Vol. 7, 1979, pp. 217-283.
2. C. S. Taber and T. J. Timpone, *Computational Modeling*, Thousand Oaks, CA, 1996.
3. M. S. C. Thomas and W. J. B. van Heuven, "Computational models of bilingual comprehension," *Handbook of Bilingualism: Psycholinguistic Approaches*, J. F. Kroll and A. de Groot, eds., Oxford University Press, New York, 2005.
4. W. H. Sumby and I. Pollack, "Visual contribution to speech intelligibility in noise," *Journal of the Acoustical Society of America*, Vol. 26, 1954, pp. 212-215.
5. H. L. Pick and E. Saltzman, "Modes of perceiving and processing information," *Modes of Perceiving and Processing Information*, H. L. Pick, Jr. and E. Saltzman, eds., John Wiley, New York, 1978, pp. 1-20.
6. B. Stein and M. Meredith, *The Merging of the Senses*, MIT Press, MA, 1993.
7. L. Bloom, *One Word at a Time: The Use of Single-Word Utterances before Syntax*, The Hague, Mouton, 1973.
8. M. Small, *Cognitive Development*, Harcourt Brace Jovanovich Publishers, San Diego, 1990.
9. B. MacWhinney, "Models of the emergence of language," *Annual Review of Psychology*, Vol. 49, 1998, pp. 199-227.
10. K. Plunkett, C. Sinha, M. F. Muller, and O. Strandsby, "Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net," *Connection Science*, Vol. 4, 1992, pp. 293-312.
11. T. Kohonen, *Self-Organization and Associative Memory*, 3rd ed., Springer, Berlin, 1989.
12. P. Li, "Language acquisition in a self-organizing neural network model," *Connectionist Models of Development: Developmental Processes in Real and Artificial Neural Networks*, P. Quinlan, ed., Psychology Press, Hove and New York, 2003, pp. 115-149.
13. R. Miikkulainen, "A distributed feature map model of the lexicon," in *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ Lawrence Erlbaum, 1990, pp. 447-454.
14. R. Miikkulainen, "Dyslexic and category – Specific aphasic impairments in a self-organising feature map model of the lexicon," *Brain and Language*, Vol. 59 1997, pp. 334-366.
15. S. S. R. Abidi and K. Ahmad, "Conglomerate neural network architectures: The way

ahead for simulating early language development," *Journal of Information Science and Engineering*, Vol. 13, 1997, pp. 235-266.
16. P. Li, I. Farkas, and B. MacWhinney, "Early lexical development in a self-organizing neural network," *Neural Networks*, Vol. 17, 2004, pp. 1345-1362.
17. E. K. Warrington, "The selective impairment of semantic memory," *Quarterly Journal of Experimental Psychology*, Vol. 27, 1975, pp. 635-657.
18. R. A. McCarthy and E. K. Warrington, "Visual associative agnosia: A clinico-anatomical study of a single case," *Journal of Neurology*, *Neurosurgery*, *and Psychiatry*, Vol. 49, 1986, pp. 1233-1240.
19. R. A. McCarthy and E. K. Warrington, "Evidence for modality-specific meaning systems in the brain," *Nature*, Vol. 334, 1988, pp. 428-430.
20. T. Shallice, "Specialisation within the semantic system," *Cognitive Neuropsychology*, Vol. 5, 1988, pp. 133-142.
21. T. Shallice, "Multiple semantics: Whose confusions?" *Cognitive Neuropsychology*, Vol. 10, 1993, pp. 251-261.
22. E. K. Warrington and R. A. McCarthy, "Multiple meaning systems in the brain: A case for visual semantics," *Neuropsychologia*, Vol. 32, 1994, pp. 1465-1473.
23. D. O. Hebb, *The Organisation of Behavior: A Neuropsychological Theory*, Wiley, New York, 1949.
24. A Caramazza, A. Hillis, B. Rapp, and C. Romani, "The multiple semantics hypothesis: multiple confusions?" *Cognitive Neuropsychology*, Vol. 7, 1990, pp. 161-189.
25. M. A. L. Ralph, K. S. Graham, K. Patterson, and J. R. Hodges, "Is a picture worth a thousand words? Evidence from concept definitions by patients with semantic dementia," *Brain and Language*, Vol. 70, 1999, pp. 309-335.
26. R. Vandenberghe, C. Price, R. Wise, O. Josephs, and R. S. J. Frackowiak, "Functional anatomy of a common semantic system for words and pictures," *Nature*, Vol. 383, 1996, pp. 254-256.
27. P. Bright, H. Moss, and L. K. Tyler, "Unitary vs multiple semantics: PET studies of word and picture processing," *Brain and Language*, Vol. 89, 2004, pp. 417-432.
28. L. Bloom, *The Transition from Infancy to Language: Acquiring the Power of Expression*, Cambridge University Press, New York, 1993.
29. J. L. Elman, "Connectionist models of cognitive development: Where next?" *Trends in Cognitive Science*, Vol. 9, 2005, pp. 111-117.
30. A. Ninio and C. Snow, "Language acquisition through language use: The functional sources of children's early utterances," *Categories and Processes in Language Acquisition*, Y. Levy, I. Schlesinger, and M. D. S. Braine, eds., Erlbaum, Hillsdale, NJ, 1988, pp. 11-30.
31. L. R. Gleitman and E. L. Newport, "The invention of language by children: environmental and biological influences on the acquisition language," *Language: An Invitation to Cognitive Science*, L. R. Gleitman and M. Liberman, eds., MIT Press, Cambridge, MA, 1995, pp. 1-24.
32. R. Brown, *A First Language: the Early Stages*, George Allen and Unwin, London, 1973.
33. J. H. Flavell, "Stage-related properties of cognitive development," *Cognitive Psychology*, Vol. 2, 1971, pp. 421-453.
34. R. Hecht-Nielsen, "Counterpropagation networks," *Applied Optics*, Vol. 26, 1987,

pp. 4979-4984.
35. S. Grossberg, "Some networks that can learn, remember, and reproduce any number of complicated space time patterns," *International Journal of Mathematics and Mechanics*, Vol. 19, 1969, pp. 53-91.
36. A. Nyamapfene and K. Ahmad, "Unsupervised multimodal processing," in *Proceedings of the 25th IASTED International Multi-Conference on Artificial Intelligence and Applications*, 2007, pp. 14-19.
37. V. R. de Sa and D. H. Ballard, "Category learning through multimodality sensing," *Neural Computation*, Vol. 10, 1998, pp. 1097-1117.
38. P. S. Dale and L. Fenson, "Lexical development norms for young children," *Behavior Research Methods*, *Instruments*, *and Computers*, Vol. 28, 1996, pp. 125-127.
39. C. Fellbaum, *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA, 1998.
40. T. Kohonen, "The hypermap architecture," *Artificial Neural Networks*, T. Kohonen, K. Makisara, O. Simula, and J. Kangas, eds., Elsevier, Amsterdam, Netherlands, Vol. II, 1991, pp. 1357-1360.
41. A. Araújo and G. Barreto, "Context in temporal sequence processing: A self-organizing approach and its application to robotics," *IEEE Transactions on Neural Networks*, Vol. 13, 2002, pp. 45-57.
42. D. Wang and B. Yuwono, "Anticipation-based temporal pattern generation," *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, Vol. 25, 1995, pp. 615-628.
43. D. Wang and B. Yuwono, "Incremental learning of complex temporal patterns," *IEEE Transactions on Neural Networks*, Vol. 7, 1996, pp. 1465-1481.
44. D. Wang and A. Arbib, "Complex temporal sequence learning based on short-term memory," *Proceedings of the IEEE*, Vol. 78, 1990, pp. 1536-1543.
45. D. Wang and A. Arbib, "Timing and chunking in processing temporal order," *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, Vol. 23, 1993, pp. 993-1009.
46. M. C. Mozer, "Neural network architectures for temporal pattern processing," *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. S. Weigend and N. A. Gershenfeld, eds., Addison-Wesley Publishing, Redwood City, CA, 1993, pp. 243-264.
47. W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, Vol. 5, 1943, pp. 115-133.
48. S. Amari, "Learning patterns and pattern sequences by self-organizing nets of threshold elements," *IEEE Transactions on Computers*, Vol. C-21, 1972, pp. 1197-1206.
49. W. K. Estes, "An associative basis for coding and organisation in memory," *Coding Processes in Human Memory*, A. W. Melton and E. Martin, eds., Winston, Washington, DC, 1972, pp. 161-190.
50. D. E. Rumelhart and D. A. Norman, "Simulating a skilled typist: A study of skilled cognitive-motor performance," *Cognitive Science*, Vol. 6, 1982, pp. 1-36.
51. B. MacWhinney, *The CHILDES Project: Tools for Analyzing Talk*, 3rd ed., Lawrence Erlbaum Associates, Mahwah, NJ, 2000.
52. A. Ninio, "The relation of children's single word utterances to single word utterances in the input," *Journal of Child Language*, Vol. 19, 1992, pp. 87-110.
53. J. Luce and P. Luce, "Similarity neighbourhoods of words in young children's lexicons," *Journal of Child Language*, Vol. 17, 1990, pp. 205-215.

54. P. Ladefoged, *A Course in Phonetics*, 2nd ed., Harcourt Brace Jovanovich, New York, 1982.
55. P. Li and B. MacWhinney, "PatPho: A phonological pattern generator for neural network," *Behavior Research Methods*, *Instruments*, *and Computers*, Vol. 34, 2002, pp. 408-415.
56. E. Fosler-Lussier, "Contextual word and syllable pronunciation models," in *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, 1999, http://www.cse.ohio-state.edu/~fosler/papers/asru99.pdf.
57. S. Theodorius and K. Koutroumbas, *Pattern Recognition*, 2nd ed., Academic Press, Elsevier Science, USA, 2003.
58. T. T. Rogers, D. Rakison, and J. L. McClelland, "U-shaped curves in development: A PDP approach. Contribution to a special issue on U-shaped changes in behavior and their implications for cognitive development," *Journal of Cognition and Development*, Vol. 5, 2004, pp. 137-145.
59. D. E. Rumelhart, J. L. McClelland, and PDP Research Group, *Parallel Distributed Processing*, *Volume 1: Foundations*, MIT Press, Cambridge, MA, 1986.
60. D. E. Rumelhart, J. L. McClelland, and PDP Research Group, *Parallel Distributed Processing*, *Explorations in the Microstructure of Cognition*, *Vol II: Psychological and Biological Models*, MIT Press, Cambridge, MA, 1986.
61. S. Pinker, "Language acquisition," *An Invitation to Cognitive Science*, 2nd ed., L. R. Gleitman, M. Liberman, and D. N. Osherson, eds., MIT Press, Cambridge, MA, 1995, pp. 135-182.
62. J. L. Elman, "Learning and development in neural networks: The importance of starting small," *Cognition*, Vol. 48, 1993, pp. 71-99.
63. S. Quartz and T. Sejnowski, "The neural basis of cognitive development: a constructivist manifesto," *Behavioral and Brain Sciences*, Vol. 20, 1997, pp. 537-596.
64. M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, "Understanding and sharing intentions: The origins of cultural cognition," *Behavioral and Brain Sciences*, Vol. 28, 2005, pp. 675-691.

Abel Nyamapfene received his Ph.D. in Computing from the University of Surrey, Guildford, UK in 2006, and his M.Sc. in Communication Engineering and B.Sc. (Hon) in Electrical Engineering from the University of Zimbabwe in 1990 and 1997 respectively. He has been a Lecturer in Electronics and Computing in the School of Engineering, Mathematics and Physical Sciences, University of Exeter, UK, since 2006. He was with the Department of Electrical Engineering, University of Zimbabwe from 1998 to 2002. Prior to this he was a Research and Development Engineer with the Zimbabwe Post and Telecommunications Corporation from 1991 to 1997. His research interests include child language acquisition, neural networks, communication networks and reconfigurable computing.